Parametric Tracking with Spatial Extraction Across an Array of Cameras

A thesis submitted to Middlesex University
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Patrick SEBASTIAN

School of Science and Technology

Middlesex University

2016

Abstract

Video surveillance is a rapidly growing area that has been fuelled by an increase in the concerns of security and safety in both public and private areas. With heighten security concerns, the utilization of video surveillance systems spread over a large area is becoming the norm. Surveillance of a large area requires a number of cameras to be deployed, which presents problems for human operators. In the surveillance of a large area, the need to monitor numerous screens makes an operator less effective in monitoring, observing or tracking groups or targets of interest. In such situations, the application of computer systems can prove highly effective in assisting human operators.

The overall aim of this thesis was to investigate different methods for tracking a target across an array of cameras. This required a set of parameters to be identified that could be passed between cameras as the target moved in and out of the fields of view. Initial investigations focussed on identifying the most effective colour space to use. A normalized cross correlation method was used initially with a reference image to track the target of interest. A second method investigated the use of histogram similarity in tracking targets. In this instance a reference target's histogram or pixel distribution was used as a means for tracking. Finally a method was investigated that used the relationship between colour regions that make up a whole target. An experimental method was developed that used the information between colour regions such as the vector and colour difference as a means for tracking a target. This method was tested on a single camera configuration and multiple camera configuration and shown to be effective.

In addition to the experimental tracking method investigated, additional data can be extracted to estimate a spatial map of a target as the target of interest is tracked across an array of cameras.

For each method investigated the experimental results are presented in this thesis and it has been demonstrated that minimal data exchange can be used in order to track a target across an array of cameras. In addition to tracking a target, the spatial position of the target of interest could be estimated as it moves across the array.

Acknowledgements

I like to take this opportunity to express my deepest gratitude to my supervisor Professor Richard Comley for his enthusiasm, advice, encouragement and support, which he has shown for the work which I have performed under his supervision. Thanks are also due to Dr. Yap Vooi Voon for discussions and guidance we had while I was working on my thesis.

This work would not be possible without the assistance in terms of time and equipment from Universiti Teknologi PETRONAS, which I gratefully acknowledge.

I would like to thank several of my PhD colleagues for their encouragement and faith in me while I was working on my thesis.

I would like also to thank my wife, Sofiah, for her love, encouragement and steadfast support she has shown, which I could not have done without. Finally, a special thanks to my daughter, Keanna Haley, who is a welcome distraction and an inspiration to me.

In memory of Cappuccino 'Chino' Michael Sebastian

Whose arrival opened up our hearts and home

Through his calm and accepting nature.

Table of Contents

List of Figures

List of Tables

Abbreviations

AESOS	Automatic Evaluation System on Object Surveillance
ASDR	Average size detection rate
CLEAR	Classification of events, activities and relationships
EM	Expectation Maximization
EMM	Expectation Maximization Method
ETISEO	Evaluation du Traitement et de l'Interpretation de Sequences vidEO
FAR	False Alarm Rate
GPS	Global positional system
GT	Good tracking
GUI	Graphical User Interface
HDTV	High Definition TV
HSV	Hue, Saturation, Value
ID	Identification
LSBEM	Label and Size Based Evaluation Measure
LTDR	Label Tracking Detection Rate
n_1	Number of consecutive correct tracks
n_f	Number of frames that tracked target is within camera view field
N_{rg}	Number of frames
N_{tp}	Number of true positives
N_{fp}	Number of false positives
N_{tn}	Number of true negatives
N_{fn}	Number of false negatives
NCC	Normalized Cross Correlation
NTDR	Non-label Tracking Detection Rate
OBR	Object Detection Rate
OTE	Object Tracking Error
OTStd	Object Tracking Standard Deviation
OT	Overall tracking
PETS	Performance evaluation of tracking and surveillance
RGB	Red, Green, and Blue
S_{ab}	Score assigned for correctly tracked target

SDTV	Standard Definition TV
TC	Tracking consistency
TRDR	Tracker detection rate
TDR	Track detection rate
TF	Track Fragmentation
TSR	Track Success Rate
ViPER	Video Performance Evaluation Resource
YCrCb	Luminance (Y)/chrominance (Cr and Cb) colour space
YUV	Luminance (Y)/chrominance (U, V) colour space

Mathematical Notations

$\to$	approaches
Δ	the change in
$f(t)$	Function of continuous variable t
$F(w)$	Fourier transform
$\psi(t)$	wavelet function
$\langle , \rangle$	an inner product
$\int_{-\infty}^{\infty}$	integral with limits
$L^2(R)$	the vector space of all functions whose square are integrable
V^j	scaling space
$\phi(t)$	scaling function (mother wavelet)
$\in$	is an element of
$\langle$	is less than
$\rangle$	is greater than
$\lfloor a \rfloor$	largest number not greater than a
$\subset$	is a proper subset of
$h(n)$	set of coefficients
$\oplus$	direct sum
$\perp$	orthogonal
a_i	real-value expansion coefficients
$\sum$	the summation of
a_k^j	approximation coefficients
a_k	wavelet coefficients
Z	set of all integers
$\Leftrightarrow$	is logically equivalent to
W^j	detail space
$\{ \}$	a set

$\ell^2(R)$	the vector space of all functions whose square are summable
$\hat{\phi}(\omega)$	Fourier transform of $\phi(t)$
$e^{j\omega t}$	complex exponential
$H(\omega)$	discrete Fourier transform of
$\Rightarrow$	implies
$\downarrow 2$	downsampling by a factor of 2
$\uparrow 2$	upsampling by a factor of 2
$x_n[n]$	digital signal
H_1	denotes low-pass filter (analysis filter bank)
H_2	denotes high-pass filter (analysis filter bank)
G_1	denotes low-pass filter (synthesis filter bank)
G_2	denotes high-pass filter (synthesis filter bank)
$A_{2^j}^d f$	discrete approximation signal
$D_{2^j} f$	discrete detail signal
d_k^j	detail coefficients
∞	infinity
$\cong$	defined as

CHAPTER 1

Introduction to the Research

1.1. Background

Surveillance is a term which has its origins from the French word of *surveill* which means ‘to watch over’. Surveillance, for the purposes of this thesis, is defined as the action of monitoring the behaviour or activities of a person or a group of people for the various purposes of managing, directing or protecting the person or group of people being monitored or the surrounding people or general public. The main users of surveillance are typically governments, law enforcement and security organizations where surveillance is used to maintain control, recognize and monitor threats and also to prevent and investigate criminal activity. Other than the typical usage of surveillance to monitor individuals or groups of people, it is also used in areas such as wildlife for monitoring and determination of the population of species in a given area, or to monitor the progress of a disease in a community. As previously stated, the focus of this thesis is human surveillance although the results have much wider significance.

Surveillance types range from simple non-technological to technological methods such as human surveillance agents to postal interception to data monitoring in computer surveillance. Of the different type of surveillance, the method of interest in the current work is video surveillance. Video surveillance is the technique of utilizing video cameras to obtain and use video or visual imagery to observe areas of interest. The video cameras are often connected to a recording device, or a recording network. The video captured is usually watched live or through a recording by a person who’s job it is to monitor the video. The person watching the videos would typically be watching a large number of screens showing images from a large number of cameras that would be covering the area of interest.

With the advent of cheaper, high quality cameras, the usage of video surveillance systems is becoming common and it is expected to become more widespread as the population continues to grow. Utilization of video surveillance in a large scale environment such as public parks or shopping complexes mean that a large number of cameras would be used to monitor the areas of interest. The utilization of a large array of cameras means that the number of monitors required is also large. Observing a large number of screens for the purposes of tracking people and observing human behaviour is a difficult task for a human operator. Due to this difficulty or limitation of human operators, computer and camera technologies have been merged to provide the ability to track and monitor targets over a large array of cameras. The merging of these technologies in monitoring large areas of interest or people of interest has lead to the growth of the use of intelligence in surveillance systems. The introduction of intelligence in video surveillance systems gives the ability to determine if specific areas are being trespassed and for specific people to be tracked over an array of cameras in real time.

In most applications of video surveillance, people are the primary objects of interest, for example when monitoring a person walking through a secluded or security sensitive area or when monitoring a group of people in a public area. Monitoring the targets of interest over a large area requires that a large number of cameras and a number of screens for an operator to monitor the cameras. Apart from the act of just monitoring, tracking is the other act that a video surveillance system is capable of doing. Tracking is the act of actively identifying and maintaining a constant track or lock on a specific target. With different types of targets to track, methods of monitoring or tracking the selected target will vary. In addition to different types of targets to be tracked, the configuration of the video surveillance systems will also affect the tracking methods that may be developed such as single camera configurations to multiple camera configurations.

1.2. Applications

The use of video surveillance has risen in recent times due to fears of rising crime rates, of public disturbances and especially fears of terrorist acts. The typical application of video surveillance systems is in the monitoring and tracking of people in large public settings. The application of a large area system requires that the video surveillance system covers large areas and the number of cameras required will be dependent on the size of the area to be monitored. The utilization of cameras in large areas can be seen in cities such as London, New York and Chicago that have installed large numbers of cameras throughout the cities. An example of a large scale implementation of video surveillance systems is the Golden Shield Project in China. The video surveillance part of the project can be considered as one of the largest deployments of video surveillance systems considering that the objective is to monitor all or most of China which has the largest population in the world. All videos captured will be fed back to a central facility for storage and monitoring purposes.

1.3. Overview of Research

The major concern of this research is to identify the parameters to be used for tracking a target over an array of cameras. The act of tracking a target is a constant act of identifying a target based on characteristics of the target. In the case of the surveillance portion of the Golden Shield Project in China, the human face is used as a means for identifying people or targets of interest. This mimics human behaviour when monitoring a large number of people in a crowd where the portion that can be used for identification for the purpose of tracking would be the human face, which is the most common part of the body that is visible in a crowd. Apart from facial characteristics being used as a means of identifying targets, other characteristics that could be used range from features such as corners, edges, shapes, dimensions of the target, to colour of the target. In this instance, the target of interest in this research would be people moving through certain areas. When tracking or monitoring large areas of interest, monitored by a large number of cameras there is a need to be able to automatically track the same target as it moves across the view field of one camera to the next camera view field. The research work reported here is based on the application of a distributed video surveillance system that is applied to oil rigs or prisons or security sensitive areas where the number of people are not as large as in a public area such as a shopping mall or railway station. This simplifies the problem considerably but the suitability of the method developed to be scaled up to include operation in crowded environments will be borne in mind throughout this work.

The primary research question to be answered is what type of information can be extracted from a camera or video feed of a human moving target that may be used to consistently track a person of interest as they walk across the view field of a camera? Tracking of a target using a single camera is not the final goal of this research which is to be able to track a target over an array of cameras. Therefore, the information that is extracted and used for tracking a target should be able to be transmitted for use to track the same target over an array of cameras as the target person walks from the view field of one camera to another camera view field. A further consideration arises at this point and that is that the fields of view of adjacent cameras cannot be assumed to be overlapping. Hence, in addition to determining the type of information needed to track a target as it moves across a camera's view field, information also needs to be extracted from these cameras to determine the relative trajectory of the target in relation to the cameras and the relative distance of the target from the different cameras. The research question therefore needs to be refined to ask what type of information is needed to be used together to track a target while in view and to create spatial information to create a map of the tracked target when moving between view fields.

1.4. Organization of Thesis

The following chapters of this thesis are organised as follows.

Chapter 2 presents the literature survey and review of differences in the properties of images and colour spaces used in video surveillance. The differences in the different colour spaces are presented and displayed on its individual axes. The mathematical equation for each colour space is also shown to indicate how each colour space is mathematically linked to each other, especially the RGB colour space.

Chapter 3 provides a literature survey and review of the work done in the field of video surveillance and the effect of colour space on tracking of targets. This section discusses the different methods used in the tracking of targets as a target of interest moves across the view field of a camera. The methods used to track a target are based on a single parameter or a set of parameters to identify the signature of the target of interest. In addition to the different methods and parameters used in tracking a target, the colour space used in the different methods were also observed to determine the tracking performance.

Chapter 4 is the literature survey and review of the work done in measuring the performance of video tracking. The work looks at many different measures that have been proposed and used by different bodies to verify the usage of the metrics. In this section, the many metrics that were utilized were based on the ability or performance of the tracking to track specific targets as long as the targets are within the view field of the camera. This section also proposes some new metrics based on the findings or the lack of metrics for an improved set of metrics in reporting the performance of a tracking method.

Chapter 5 discusses a simple tracking method as a means of tracking a target which is the normalized cross correlation method. The normalized cross correlation tracking method was implemented in tracking a face. The tracking performance was observed over different colour spaces to determine the effect of colour space in tracking a face. Based on the results from tracking a face using normalized cross correlation, the next step in the investigation was to implement a whole person tracking using normalized cross correlation. The tracking of a whole person is an extension to the face tracking which involves a tracking of multiple coloured regions representing a whole person.

Chapter 6 looks at using a histogram of pixels from the tracked targets as a means of identifying and tracking the target as it moves across the camera view field. This section investigates different methods of obtaining the reference histogram and the statistical application of histogram similarity to determine the correct tracking of a target.

Chapter 7 explores an approach in tracking a target using the parameters of multiple blobs on a single camera. In this section the target is broken into different colour regions using the Expectation Maximization Method and from the segmentation into different colour regions sets of data points could be generated between colour regions. The data points extracted are then processed to generate a set of parameters using Gaussian Mixture Models. These parameters could then be used as a means of tracking the selected target. These sets of parameters are used for tracking a target on a single camera configuration. The parameters that are extracted could be used to track the same target across an array of three cameras.

Chapter 8 explores further the parametric tracking of a target across an array of cameras by having the target of interest being tracked in different scenarios that have another object of target within the camera view field. In addition to testing the capability of the parametric tracking, other blob parameters are used as a means for spatial estimation which in this instance was determining if the target of interest is nearer or further from each camera as it moves across the view field of the array of cameras.

Chapter 9 summarizes the main results of this thesis and the contributions made in the investigation of tracking a target across an array of cameras. In addition to the results obtained, this work also puts forward suggestions for the future direction or possible other research that can be developed from this work.

CHAPTER 2

Properties of Images

2.1. Introduction

The core of the video surveillance system is the video technology which has the ability to electronically capture, store, process and transmit videos. Videos are essentially a series of still images representing scenes captured by cameras. The series of still images allows for videos to be reconstructed after the still images are processed for the purposes of transmission or viewing. In looking at still images or processing still images to extract relevant information, it is important to understand the colour spaces used to visualize or display the images. Understanding the fundamentals of the colour space enables the extraction of relevant data from the targets of interest. Understanding of colour spaces is relevant with a number of different spaces available for use in video surveillance ranging from grayscale, RGB, YCbCr to HSV. The availability of different colour spaces provides different types or variations of data to be extracted from the different colour spaces.

2.2 Grayscale Colour Space

Grayscale images can be classified as intensity type images where the images in this format are images without colour. Grayscale images are images where each pixel value is a single sample which carries the intensity information only. These images are also known as achromatic images or one-colour images where the images are composed of pixels of varying shades of gray which start from black which has the weakest intensity to white which has the highest intensity. The data used in the representation of grayscale images is the measurement of the intensity or amount of light [Gonzalez, Woods et al., 2004]. The intensity of a pixel can be represented to be within a range starting from its minimum value which would represent black to its highest value which would represent white. This can be seen in the following Figure 2.1 which shows the continuous change of intensity from its lowest level to its highest level.

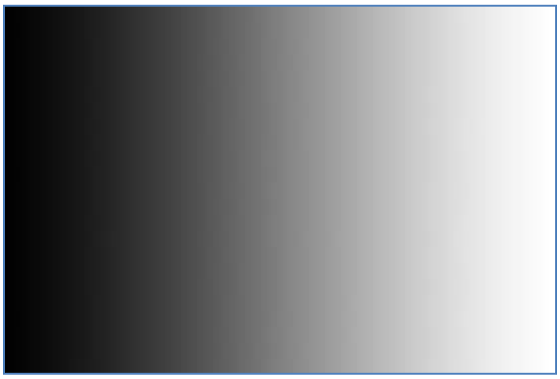


Figure 2.1: Continuous grayscale

The number of bits used in each pixel determines the number of levels of brightness [Umbaugh, 1998] or quantized levels in representing the varying levels of brightness. For example if 8 bits are used to represent each pixel, that would provide 256 levels of brightness or gray levels. An illustration of the number of levels in representing the varying levels of brightness in grayscale images can be seen in Figure 2.2. Figure 2.2 shows the intensity being divided or given 16 levels of brightness where the brightness would change from its lowest value to its highest level in just 16 steps.

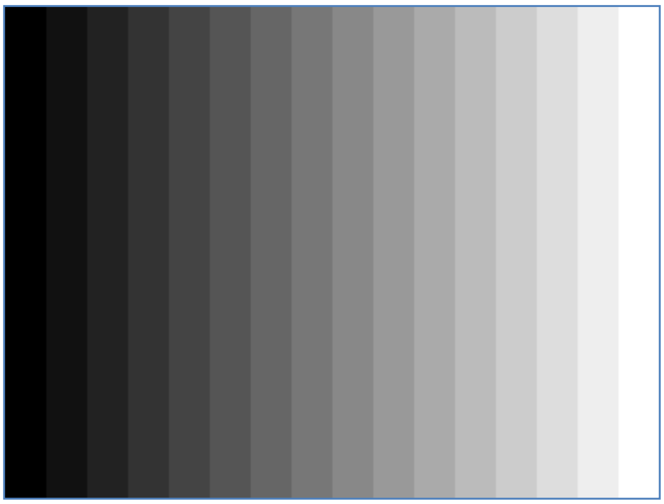


Figure 2.2: Quantised grayscale (16 level)

2.3 RGB Colour Space

After grayscale images, the next type of images used are colour images where colour information is embedded in each pixel. Colour images are made up of a set of multiple layers where each layer in a colour image represents a certain band in the visible light spectrum [Umbaugh, 1998]. In a typical colour image represented in red (R), green (G) and blue (B) layers, or RGB image, the information stored in this colour space would be the brightness in the individual spectral band. [Umbaugh, 1998]. The red, green and blue spectral colour bands represent the primary colours in the visible range of the electromagnetic spectrum [Gonzalez and Woods, 2002]. The RGB colour space was developed to correspond to the characteristics of the human eye which are sensitive to different wavelengths of the visible spectrum. In this instance the sensitivity or the ability to observe an image in colour is dependent on the number of receptors or cones that are sensitive to the red, green and blue light of the light spectrum, where approximately 65% of the cones are sensitive to red light, 33% are sensitive to green light and 2% are sensitive to blue [Gonzalez and Woods, 2002]. With the sensitivity of the eye to different wavelengths of the visible spectrum, colours are the different combinations of the primary colours of red, green and blue. In the RGB colour space, each colour appears in its primary spectral components of red, blue and green and this can be illustrated in Figure 2.3.

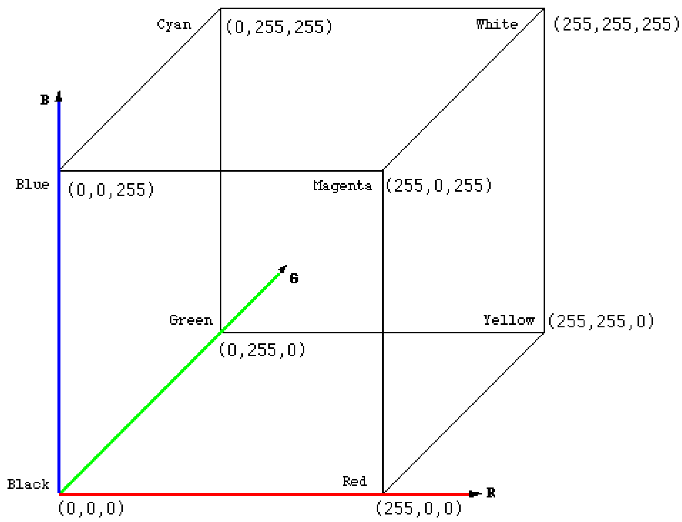


Figure 2.3: RGB Colour Space [Burdick, 1997, 315]

The RGB colour model illustrated in Figure 2.3 is based on the Cartesian coordinate system. In Figure 2.3, the different colours are indicated on the corners of the cube which are different combinations of the RGB values that generate cyan, magenta and yellow. Black is located at the origin of the cube and white is located at the furthest location from the origin of the cube. [Gonzalez and Woods, 2002]. Grayscale values are points where each point has equal RGB values, are located on the line that connects the black point to the white point in the RGB colour model. Figure 2.3 illustrates that each red, green and blue colour components have an 8-bit depth that indicates that the maximum value is 255 for each colour. Each colour pixel in the RGB colour space, where each pixel is said to have three values of (R,G,B), will therefore have a pixel depth of 24 bits which is the number of bits used to represent a colour in a particular colour space.

Due to limitations that were built into the RGB colour space where the brightness information is embedded into each colour layer, different colour spaces were also developed. [Kim, Cho et al., 2004, 785-788]. The varying levels of brightness in an image causes the RGB values in an image to shift [Chen, Li et al., 2005, 1766-1768] and introduces instability in the image [Zhao, Bu et al., 2002, 325-332]. The vulnerability of the RGB colour space to brightness makes this colour space easily affected on each layer of the RGB colour space. This indicates that each layer in the RGB colour space is correlated to each other [Kobayashi, Noma et al., 1999, 73-77; Zhao, Bu et al., 2002, 325-332]. The correlation of each layer in the RGB colour space shows that any effect or change on one of the layers also has a corresponding effect on the other layers.

In order to remove the effect of brightness on each colour layer, the RGB colour space is transformed into a different colour space that decouples the brightness effect from the colour information [Umbaugh, 1998]. The separation of the brightness information from the colour information embedded in the RGB colour space results in different colour spaces that have one layer of brightness and 2 layers of colour information [Umbaugh, 1998]. Other colour spaces typically used in the area of video tracking and video surveillance are YCbCr [Chen, Li et al., 2005, 1766-1768] and HSV [Stern and Efros, 2002, 236-241].

The adoption of the YCbCr and HSV colour spaces in image and video processing for the purpose of tracking was due to the properties of these colour spaces that have separated the brightness information from the colour information [Kim, Cho et al., 2004, 785-788]. In the YCbCr colour space, the information is separated into luminance in the Y layer and chrominance in the Cb and Cr layers. In the HSV colour space, the luminance information is placed in the V layer and the chromaticity information is placed in the H (hue) and S (saturation) layers [Gonzalez and Woods, 2002; Umbaugh, 1998]. The separation of the brightness information from the chrominance or chromaticity reduces the effect of uneven illumination in an image. The utilization of chrominance or chromaticity in the YCbCr and HSV colour space provides the opportunity for robust tracking. For robust tracking in the RGB colour space, the RGB colour space needs to be normalized to obtain chromaticity [Beetz, Radig et al., 2006, 39-42; Kawato and Ohya, 2000, 141-1418; Park, Seo et al., 2000, 133-136; Soriano, Martinkauppi et al., 2000, 839-842]. The chromaticity data obtained through the normalization of RGB values is still dependent on the data obtained from the RGB colour space that is easily affected from varying illumination. This differs for the YCbCr and HSV colour spaces that have different layers of information that has the luminance information separated from the colour information.

2.4 YCbCr Colour Space

The YCbCr colour space was defined by the International Radio Consultative Committee and is widely used in digital video and image compression schemes. The information stored in this image colour space is stored in the luminance (Y) and chrominance layers (CbCr) where Y stores the light intensity and Cb and Cr layers store the colour difference information. The Cb layer is the difference between the blue layer information and a reference value. The Cr layer is the difference between the red layer information and a reference value. The relationship between the RGB and YCbCr colour spaces can be seen in the following equations which are based on the ITU-R BT.601-7 [Union, 2011]. The equation used here was to derive the conversion between the RGB colour space to the YCbCr colour space to be used in Standard Definition TV (SDTV).

$$Y = 0.299R + 0.586G + 0.114B \tag{2.1}$$

$$Cb = -0.172R - 0.339G + 0.511B + 128 \tag{2.2}$$

$$Cr = 0.511R - 0.428G - 0.083B + 128 \tag{2.3}$$

The advent of High Definition TV (HDTV) prompted a change from ITU that proposed a change in ITU-R BT.709-5 [Union, 2002] concerning the equations related between the RGB and the YCbCr colour space.

$$Y = 0.213R + 0.715G + 0.072B \quad (2.4)$$

$$Cb = -0.117R - 0.394G + 0.511B + 128 \quad (2.5)$$

$$Cr = 0.511R - 0.464G - 0.047B + 128 \quad (2.6)$$

There are some differences in the equations relating the RGB colour space to the YCbCr colour spaces. These differences are due to the different coefficients applied to the equations. These coefficients are based on the sensitivity of the human vision to the RGB colour space [Poynton, 2012]. For the luminance or Y component, the coefficients related to the sensitivities of the human eye to each of the RGB colour space components. This means that the relative contribution of the red and blue colour components to luminance is lower than the green colour component. It can be seen that the green channel dominates luminance where the green channel comprises between 60 to 70 % of the luminance. Since the green channel dominates the luminance component, the colour components remaining to be used are the blue and red components. The brightness component is removed from the blue and red colour channels by subtracting the brightness and creating the colour difference or chroma components of Cb and Cr [Poynton, 2012].

The previous section describes the YCbCr colour space but it does not describe the use or requirement of the YCbCr colour space. The usage of RGB colour space for transmission or storage requires three times the capacity compared to a grayscale image. The human vision system has less spatial acuity or sensitivity for colour information compared to brightness. Due to the poor colour acuity of the human vision system, a colour image can be coded into a wideband monochrome component representing brightness or illumination and two narrowband components carrying colour information. In digital video, each colour channel has half the data rate or data capacity compared to the monochrome channel [Poynton, 2012]. Based on the properties of the colour information channel, which has the colour difference information, this enables the colour information to be subsampled and thus reduces the amount of data to be used in storing or transmission of digital video in YCbCr colour space.

The relationship between the YCbCr and RGB colour space can be illustrated in the following figure which shows the placement or the RGB cube within the YCbCr colour space. It shows that not all the possible values in the YCbCr colour space would be able to represent all possible RGB colours. Therefore, when converting from the YCbCr colour space to RGB colour space, there is a need to ensure that overflow or underflow conditions are taken care of in the RGB colour space.

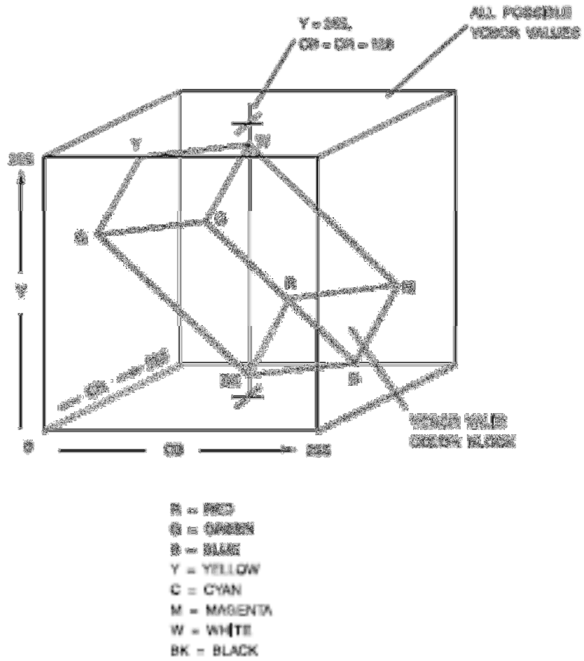


Figure 2.4: YCbCr Colour Space Model [Jack, 2007]

2.5 HSV Colour Space

Another colour space investigated was the HSV that has information stored in the hue (H), saturation (S) and value (V) layers respectively. The HSV colour space is a colour space that describes colour the way the human eye observes colour. The human eye describes a colour by the combination of its hue, saturation and brightness. Hue (H) is an attribute that describes the pureness of a colour, whereas saturation (S) describes the measure of the colour that is diluted by white light. Value (V) is brightness that is a measure of the intensity of the colour. The HSV colour space can be further explained by the HSV hexagonal cone in Figure 2.5.

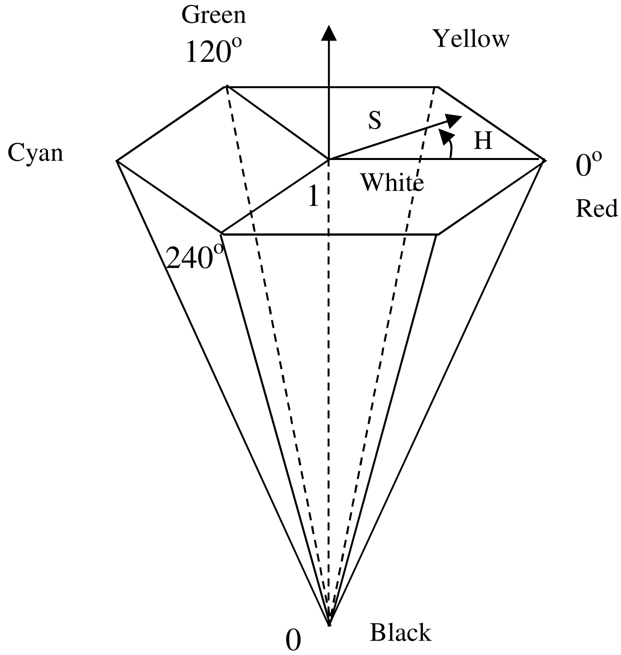


Figure 2.5: HSV Colour Space Model [Gonzalez, Woods et al., 2004]

The colour information in the HSV colour space is stored in the hue (H) and saturation (S) layers. The hue (H) parameter is expressed as an angle around the colour hexagon. The hue of a colour refers to the wavelength that closely matches the colours of the rainbow. Typically, at 0^o the hue is red, which is also used as the reference value, 120^o is green and 240^o is blue. Saturation (S) is the measure of the purity of the colour which is the distance from the V axis [Gonzalez, Woods et al., 2004]. The edges of the cone represent the colours at its purest where the saturation values are at unity. The concept of saturation can be further explained by looking at the colour of red that would be located at $H=0^o$ and $S=1$. By mixing white with the red, the red will be less intense and thus reduce its intensity. If more of the white colour is mixed with red colour, the red colour would become paler and the saturation decreases further [Castleman, 1996]. Saturation is then measured by the distance of the coordinate from the V axis. The value (V) parameter indicates the brightness levels or the grayscale values that range from black at the bottom of the axis to white at the top of the V axis. The relationship between the RGB and HSV colour space is given in the following equations:

$$H = \cos^{-1} \frac{\frac{1}{2}[(R-G)(R-B)]}{\sqrt{[(R-G)^2 + (R-B)(G-B)]}} \qquad (2.7)$$

$$S = \frac{\max - \min}{\max} \qquad (2.8)$$

$$V = \max \qquad (2.9)$$

$$\max = \max imum(R,G,B); \min = \min imum(R,G,B)$$

2.6. Summary

In conclusion, the differences between a number of colour spaces have been discussed. The different colour spaces that were discussed ranged from grayscale, RGB, YCbCr to HSV. Each colour space had different information stored for each pixel. The simplest colour space is the grayscale colour space which has intensity information captured for each pixel. The RGB colour space is an extension of the grayscale colour space where the RGB colour space displays colour pixels by a combination of intensity information on three different colour layers which were the Red, Green and Blue colour layers. The RGB colour space is also a colour space that is an additive type of colour space which adds or combines the different values from each colour layer to generate the final colour desired [Jack, 2007]. The other colour spaces investigated were the YCbCr and HSV colour spaces which differ from the RGB colour space in how the information is stored. The YCbCr and HSV colour spaces stored the intensity information on different layers compared to the RGB colour space. The RGB colour space had the intensity information embedded together with the colour information. The other colour spaces have the intensity information on a separate layer from the colour information. In the YCbCr colour space, the intensity or luminance information is stored in the Y layer and the colour information is stored in the Cb and Cr layers. The Cb and Cr layers are essentially the chroma or colour difference layers where blue and red channel colour information is subtracted by illumination. The HSV colour space is another colour space that has the brightness or illumination information on a separate layer compared to colour information. This colour space attempts to describe the colour the way the human eye describes colour such as hue, saturation and value. This chapter has presented the different types of colour spaces available with the different information stored within each colour space. The information stored in different colour spaces ranged from brightness or illumination, colour, colour difference and colour saturation. The utilization of the specific information from each colour space could assist in the segmentation [Lim, H.K. et al., 2000, 339-342; Sural, Qian et al., 2002, 589-592] or tracking of targets [Erdem, Sankur et al., 2003, 29-40] and even in background subtraction [Horprasert, Harwood et al., 1999; KaewTraKulPong and Bowden, 2001; Zivkovic and Heijden, 2006, 773-780]. The separation of colour information could provide an avenue for better tracking capabilities of coloured objects which can be affected by varying levels of illumination or brightness levels. The stability of the YCbCr and HSV colour spaces to variation in intensity make them the most suitable choices for video surveillance applications.

CHAPTER 3

Video Tracking Methodologies

3.1. Introduction

Video surveillance is part of an overall surveillance system which uses video cameras to monitor the behaviour and activities of a group of people or a single person. The monitoring is usually done for the purpose of influencing, managing or directing people or for the purpose of protecting people or places of interest. In a video surveillance system, an operator is available or required to monitor or observe the activity that is being captured by the video cameras. The monitoring or observing of a single camera is a relatively simple task for an operator but in a video surveillance system which employs multiple cameras and video screens, the workload on the operator would be high to the extent that an operator would find it hard to get the task of monitoring done. Apart from monitoring, the additional functionality in using video surveillance systems would be the ability to track specific targets in a system with a large number of cameras. Utilizing a human operator to track a target over a large number of cameras and monitors would be difficult. As such, the utilization of computers to automate or assist in the tracking of a target would be beneficial in tracking targets. In using computers to assist in tracking targets, tracking methodologies would have to be developed to track specific targets based on specific characteristics or ‘signature’ of the tracked target or targets.

3.2. Window Tracking

Among the simplest methods of tracking is window tracking. This technique utilizes a small reference image or rectangular region for tracking the reference image within an image frame. The windows tracking method can be tracked frame by frame by a correlation type matching method [Ali, Indupalli et al., 2006; Trucco and Plakas, 2006, 520-529]. The windows tracking method is the simplest method of tracking where it quantifies the similarity of two group of pixel values that are located in the reference image and the input image [Trucco and Plakas, 2006, 520-529].

This method determines the highest similarity index between the two groups of pixels to determine the correct tracking of a target. Typically the reference image is a smaller image from the input image where an attempt to locate the reference image within the input image would be done [Hii, Hann et al., 2006, 144-156; Lewis, 1995, 120-123]. Tracking by correlation is done by comparing the data from reference and input images which is based on the information stored on the different layers of the images [Sebastian and Vooi Voon, 2007]. Based on the colour format being used, different types of information is saved on the different layers. Apart from image data, the data can be transformed into other forms such as wavelets [Nillius and Eklundh, 2002]. The windows tracking method has certain issues such as image distortion between frames due to change in viewpoint, effects of changing patterns in an image and number of pixels used in tracking. The windows tracking method utilizes a tracking method that tracks all the pixels in the reference image which gives a large number of points to track. A smaller number of points to track is preferable compared to the large number of points required in window tracking [Trucco and Plakas, 2006, 520-529].

3.3. Feature Tracking

The issues faced by windows tracking have led to another technique of tracking which utilizes detected features in an image. Among the issues or problems faced in windows tracking were the reference image to be used for tracking and relative size of the reference image to be used to track the target of interest. These features can be further classified as local features and extended features [Trucco and Plakas, 2006, 520-529]. Features can be defined as detectable parts of an image which can be used to support a vision task that includes corners, lines, contours or specially defined regions. Feature tracking initially locates features in subsequent image frames and then matches the detected features accordingly between the frames. The manner of feature tracking is similar in method where the detected features are matched to each other. The difference of feature tracking compared to window tracking is that feature tracking only tracks certain identified features as compared to window tracking that tracks all pixel points available in the window.

Feature tracking can be further classified either as local features or extended features. Local features are image locations with specific properties such as edges, lines and corners [Trucco and Plakas, 2006, 520-529]. Local features offer invariant features for better tracking where uneven illumination can cause inaccurate tracking. Local features are not only detected on the target image to be tracked but also on the overall image. This gives a possibility of wrongly matching the local features [Trucco and Plakas, 2006, 520-529].

With the possibility of mismatching local features in an image, extended features were used to provide better or more robust tracking. Extended features use a larger part of the image compared to local features. Extended features cover basic shapes such as ellipses, rectangles, free form contours and image regions [Atsushi, Hirokazu et al., 2002, 2974 - 2981; Park and Aggarwal, 2002, 105-111; Trucco and Plakas, 2006, 520-529]. The advantage of extended features was its robustness to clutter compared to local features. Regions are defined as connected image parts with intensity or colour properties. Image contours are another feature used for tracking purposes where the contour information is used as the boundary to the whole object to be tracked [Trucco and Plakas, 2006, 520-529; Yang, Doherty et al., 2000, 95-101]. For robust tracking of an object it utilizes features available on the tracked object together with additional parameters obtained from tracking the extended features detected in an image. The additional parameters used in tracking range from velocity of features [Atsushi, Hirokazu et al., 2002, 2974 - 2981; Cai and Aggarwal, 1996, 68-72; Lu, Wang et al., 2004, 188 - 191], path of features [Bodor, Jackson et al., 2003; Cai and Aggarwal, 1996, 68-72; Niu, Long et al., 2004, 719-722], deformable shape of region [Haritaoglu, Harwood et al., 2000, 809-830; Niu and Abdel-Mottaleb, 2004, 546-556; Wang, Bebis et al., 2006], region colour [Matsuzawa and Kumazawa, 2000, 94-97; Park and Aggarwal, 2002, 105-111] and region sizes [Moritomo, Kiriyama et al., 2005, 3215-3218; Sato and Aggarwal, 2001, 87-94; Teknomo, Takeyama et al., 2001, 11-18]. Parameters such as velocity of features are used to track a target but it still requires that features used to be identified and used as reference points. In the utilization of the path of features, the method of tracking is by classifying the path of the tracked target which is further classified by the velocity of the target. Other tracking methods developed essentially used parameters that would have to be extracted from the tracked target. The information would have to be extracted from a blob of the tracked target.

3.4. Colour Based Tracking

In object tracking, the colour space utilized determines the features and parameters used in tracking an object. Grayscale tracking typically is done on a blob or a specific region that utilizes additional or accompanying parameters such as distance [Sato and Aggarwal, 2001, 87-94], velocity [Atsushi, Hirokazu et al., 2002, 2974 - 2981; Cai and Aggarwal, 1998, 356-362; Lu, Wang et al., 2004, 188 - 191], trajectory [Bodor, Jackson et al., 2003; Cai and Aggarwal, 1996, 68-72; Niu, Long et al., 2004, 719-722] and blob dimensions such as height [Cai and Aggarwal, 1999, 1241-1247; Sato and Aggarwal, 2001, 87-94]. The application of the accompanying parameters is used to maintain short term tracking of a target object. In addition to the combination of utilizing features and parameters, other methods of tracking are utilized for tracking purposes. The different measures used for tracking include properties of the tracked object which cover Manhattan Distance [Leo, D'Orazio et al., 2004, 913 - 916], Mahalanobis Distance [Cai and Aggarwal, 1996, 68-72; Matsuzawa and Kumazawa, 2000, 94-97], shape tracking [Haritaoglu, Harwood et al., 2000, 809-830; Niu and Abdel-Mottaleb, 2004, 546-556], histogram similarity index [Wahl, Hillenbrand et al., 2003, 474-481] and consistent labelling [Khan and Shah, 2003, 1355-1360]. Other methods of tracking utilize complex statistical and intelligent methods which were Bayesian Network [Diaz de Len and Sucar, 2002, 439 - 442; Kitani, Sato et al., 2005, 239-246] and neural network [Han, Sethi et al., 2004, 3065-3068].

The grayscale colour space allows for object tracking using the brightness information embedded in grayscale images. With colour images having more information embedded in the images, it would be expected that tracking using colour would be more robust, accurate and overcome problems associated with the grayscale colour space. One of the primary problems of the grayscale colour space was the brightness influence on grayscale images which affects tracking [Yang, Doherty et al., 2000, 95-101]. Brightness also plays a part in tracking effectiveness in a colour image where illumination was required to be uniform for effective tracking [Li, Chua et al., 2002, 309-314]. One method of minimizing the effect of brightness on colour image tracking would be to convert the typical RGB colour images to different colour spaces such as YCbCr [Chen, Li et al., 2005, 1766-1768; Chippendale, 2006, 487-492; Kim, Cho et al., 2004, 785-788] and HSV [Stern and Efros, 2002, 236-241] for robust tracking. The usage of different colour spaces other than RGB would be to overcome the correlation between layers of the RGB colour space [Kobayashi, Noma et al., 1999, 73-77]. The YCbCr and HSV colour spaces can be used for more accurate tracking due to these colour spaces having less noise [Zhao, Bu et al., 2002, 325-332] and the ability of using different layers without losing data [Chippendale, 2006, 487-492; Stern and Efros, 2002, 236-241] and tracking ability [Zhao, Bu et al., 2002, 325-332].

As in grayscale colour space tracking, colour tracking utilizes similar tracking methods such as shape tracking [Koschan, Kang et al., 2002, 126-131; Matsuzawa and Kumazawa, 2000, 94-97; Siebel and Maybank, 2002, 373-387] and blob [Hidaka, 2009, 1827 - 1830; Leo, D'Orazio et al., 2004, 913 - 916; Liang, Huang et al., 2007, III - 369 - III - 372; Sato and Aggarwal, 2001, 87-94] or region tracking [Bakowski and Jones, 1999, 794-798; Haritaoglu, Harwood et al., 2000, 809-830]. In addition to the existing tracking methods, colour information is also used in tracking [Bird, Masoud et al., 2005, 167-177; Bredereck, Jiang et al., 2012, 1 - 6; Moritomo, Kiriyama et al., 2005, 3215-3218; Park and Aggarwal, 2002, 105-111; Sharma, 2012, 161-168; Teknomo, Takeyama et al., 2001, 11-18]. For robust tracking in colour, the typical tracking method uses colour [Bird, Masoud et al., 2005, 167-177; Iocchi and Bollees, 2005, 872 - 875; Li, Chua et al., 2002, 309-314] and region properties such as colour of the head region, height of the head region, area, width, length, perimeter, shape and position [Chang and Gong, 2001, 19-26; Moritomo, Kiriyama et al., 2005, 3215-3218; Park and Aggarwal, 2002, 105-111; Teknomo, Takeyama et al., 2001, 11-18; Wang, Bebis et al., 2006; Yan and Forsyth, 2005, 370-377]. Although colour information adds robustness to object tracking it is also susceptible to brightness variations [Li, Chua et al., 2002, 309-314]. Colour information is used together with motion parameters such as velocity, position [Park and Aggarwal, 2002, 105-111] and motion vector [Ali, Indupalli et al., 2006; Dahmane and Meunier, 2005, 136-143; Wang, Bebis et al., 2006]. Apart from the tracking methods described, other methods of tracking utilize colour information as a means for determining the similarity between objects that are being tracked. The measure of similarity between a reference object and a tracked object can be measured by measures of Mahalanobis distance [Matsuzawa and Kumazawa, 2000, 94-97], Bhatacharrya distance [Wang and Liu, 2006, 10275-10279; Zivkovic and Krose, 2004, 798-803], colour distribution [Wang and Liu, 2006, 10275-10279] and histogram similarity index such as histogram intersection, Squared Euclidian Distance, Chi-Squared and Kullback-Leibler [Greenspan, Gordon et al., 2002, 970-973; Wahl, Hillenbrand et al., 2003, 474-481; Wang and Liu, 2006, 10275-10279; Zivkovic and Krose, 2004, 798-803]. Based on the different colour spaces, different tracking methods were developed based on the different information available on the colour spaces. Some of the tracking methods utilized in surveillance can be seen in the following Table 3.1 that lists the research completed by different authors on the tracking method used for tracking targets in video surveillance and the colour space used.

Table 3.1: List of Tracking Methods used against Colour Space

Authors	Colour Space	Tracking Method
Lewis [Lewis, 1995, 120-123]	Grayscale	Normalized cross correlation
Sebastian et al. [Sebastian and Vooi Voon, 2007]	Grayscale	Normalized cross correlation
Nillius et al. [Nillius and Eklundh, 2002]	Wavelet	Normalized cross correlation
McErlean [McErlean, 2006]	Wavelet	Phase correlation
Atsushin et al. [Atsushi, Hirokazu et al., 2002, 2974 - 2981]	Grayscale	Multiple cameras tracking on binary image of human blob. Tracking of target is based on ellipsoid shape and velocity of blob.
Lu et al. [Lu, Wang et al., 2004, 188 - 191]	Grayscale	Tracking is done from a top view and uses heads as a feature to be tracked. Tracking is based on the apparent speed of a tracked object.
Bodor et al. [Bodor, Jackson et al., 2003]	Grayscale	Tracking done based on path and velocity of tracked object.
Niu et al. [Niu and Abdel-Mottaleb, 2004, 546-556]	Grayscale	Tracking based on relative trajectories and position of each tracked object.
Cai et al. [Cai and Aggarwal, 1996, 68-72; Cai and Aggarwal, 1998, 356-362; Cai and Aggarwal, 1999, 1241-1247]	Grayscale	Tracking on multiple camera using velocity and position of tracked object for tracking. Using additional information of statistical measure for tracking (Mahalanobis Distance)
Sato et al. [Sato and Aggarwal, 2001, 87-94]	Grayscale	Tracking based on parameter of blob being tracked such as relative distance, velocity and size.
Han et al. [Han, Sethi et al., 2004, 3065-3068]	Grayscale	Tracking uses neural network to maintain tracking of target object.
Wahl et al. [Wahl, Hillenbrand et al., 2003, 474-481]	Grayscale	Tracking is based on similarity index calculated from image histogram.
Chang et al. [Chang and Gong, 2001, 19-26]	RGB	Multiple cameras track based on apparent height and colour of tracked target.
Iocchi et al. [Iocchi and Bollees, 2005, 872 - 875]	HSV	Tracking from a plan view using a set of colours to represent each target tracked.

Li et al. [Li, Chua et al., 2002, 309-314]	RGB	Uses colour invariance for tracking in a light varying condition
Siebel et al. [Siebel and Maybank, 2002, 373-387]	RGB	Multiple track algorithms used for robust tracking based on head, motion and region track.
Bird et al. [Bird, Masoud et al., 2005, 167-177]	RGB	Tracking based on short term biometrics of clothing colour.
Park et al. [Park and Aggarwal, 2002, 105-111]	HSV	Tracking done based on blob size, colour, location, perimeter, shape and orientation.
Matsuzawa et al. [Matsuzawa and Kumazawa, 2000, 94-97]	RGB	Tracking uses neural network for shape representation. Utilizes Mahalanobis Distance with colour information for tracking
Teknomo et al. [Teknomo, Takeyama et al., 2001, 11-18]	RGB	Tracking based on feature such as area, perimeter box width and colour.
Moritomo et al. [Moritomo, Kiriyama et al., 2005, 3215-3218]	RGB	Tracking based on area, width and height of blob, position and colour of blob.
Zivkovic et al. [Zivkovic and Krose, 2004, 798-803]	HSV	Uses Bhatacharrya Distance as similarity index for tracking purpose.
Ali et al. [Ali, Indupalli et al., 2006]	RGB	Uses multiple methods of tracking. Uses template matching for tracking. Uses blob features such as size, motion vector and centroid location Uses Euclidian and histogram comparison for matching.
Greenspan et al. [Greenspan, Gordon et al., 2002, 970-973]	RGB	Uses similarity index between distributions for tracking such as Gaussian Mixture distributions. Similarity index calculated based on intersection, Chi-Squared and Kullback-Leibler measures for tracking.
Wang et al. [Wang and Liu, 2006, 10275-10279]	RGB	Uses Bhattacharya Distance to measure similarity between colour regions.
Wang et al. [Wang, Bebis et al., 2006]	RGB	Shape information used as parameter for tracking
Yang et al. [Yang, Doherty et al., 2000, 95-101]	Grayscale ; Wavelet	Tracking is based on wavelets with parameters such as trajectory, size, grayscale distribution and texture
Amiri et al. [Amiri, Rabiee et al., 2003, 961-964]	Wavelet	Tracking is based on template matching for block matching.
Ellis [Ellis, 2002, 228-233]	RGB	Tracking based on centroid, height, width, chromaticity mean and trajectory
Chen et al. [Chen, Li et al., 2005, 1766-1768]	YCbCr RGB	Tracking based on colour detected against reference colour

Table 3.1 lists a number of papers in terms of the image type used in video surveillance against the tracking method used. Each tracking method developed is based on the information or image used in video surveillance. Colour information would seem to be a factor in tracking for video surveillance where there would be differences in the tracking performances [Sebastian, Vooi Voon et al., 2010, 298-312].

3.5. Conclusion

In video surveillance, the utilization of computers to assist in monitoring and tracking targets is becoming a necessity when a large number of cameras are deployed over a large area. In automating the process of tracking a target, the necessary information has to be extracted from the target of interest. The extracted data forms the means or the ability in tracking the specified targets. Multiple methods were available for tracking targets. In this instance the target of interest would be the tracking of a human being as the person walks across the view field of one camera to another camera. Multiple methods of tracking were developed which range from windows tracking, blob tracking, histogram tracking, feature tracking, velocity tracking and other parametric tracking methods. In addition to the different tracking methods investigated, it was also observed that colour space also has an effect on the methods developed. Based on Table 3.1, comparing tracking performances from the different methods had shown that different colour spaces had different tracking performances. The tracking methods investigated ranged from cross correlation, blob parameters and histogram similarity that were used in different colour spaces. Initial research had shown that different tracking methods were developed in different colour spaces and in each method developed, the tracking performances had shown improvement in different colour spaces which ranges from grayscale to RGB to YCbCr. Based on initial research and observations, it would be an area for further work and observations to be done on the effect of different colour spaces on the accuracy of tracking performance. This would be beneficial where the most suitable colour space could be chosen for the purposes of tracking a specific target which ranges from grayscale to RGB to YCbCr colour spaces.

Based on investigations done in this chapter, among the information used for the research is the utilization of the appropriate colour space for tracking a target. The colour space in question at this stage was colour space which has the colour information separated from illumination or brightness information. The colour space of interest in this instance would be the YCbCr and the HSV colour spaces. In addition to the specific colour space selected for the following investigation of the viability of tracking methodologies, normalized cross correlation, blob colour, histogram and blob parameter were the tracking methodologies that were selected as the methods that would be investigated as the tracking method to be used across an array of cameras.

CHAPTER 4

Video Tracking Performance Metrics

4.1. Introduction

In developing different video or target tracking methods, there is also a need to measure their relative performance in order to determine whether each tracking method developed is an improvement over a previous method using different tracking parameters. Tracking metrics are measures of the ability of an algorithm in processing signals ranging from still images, audio signals to video signals [Navas and Sasikumar, 2011, 50-56]. In the field of video surveillance, tracking of targets is part of the ability which requires that algorithms be developed to track targets which also require that the tracking performance be measured to determine the tracking ability of developed algorithms. Performance metrics are used as a measure to determine the correctness or accuracy of the tracking algorithms developed.

Different evaluation metrics have been developed to measure the performance of tracking algorithms [Bernardin and Stiefelhagen, 2008; Denman, Fookes et al., 2009, 541; Nascimento and Marques, 2006, 761; Nghiem, Bremond et al., 2007, 15; Nghiem, Bremond et al., 2007, 476-481; Wu, Sankaranarayana et al., 2010, 1443; Young and Ferryman, 2005, 317-324; Zhai, Shafique et al., 2006]. In developing performance metrics, the measuring of the tracked points have to be referenced against specific reference points which are also known as ground truth points. There are alternatives to using ground truth to evaluate the performance of a tracking algorithm and a number of these have been identified by Ellis [Ellis, 2002, 26-31]. The alternatives include a method that requires each target to carry a mobile global positioning (GPS) receiver to record 4D trajectory information that can be correlated directly with results of the video dataset. Another method cited takes advantage of the capability of automatically identifying targets in overlapping multi-camera views [Ali, Indupalli et al., 2006; Atsushi, Hirokazu et al., 2002, 2974 - 2981; Cai and Aggarwal, 1996, 68-72; Chang and Gong, 2001, 19-26; Ellis, 2002, 228-233] and a third method uses synthetic images sequences to assess the performance of the algorithm [Schlogl, Beleznai et al., 2005, 519-522].

In general, the objective of the development of each tracking method was to ensure the ability of tracking or maintaining an accurate track of selected objects or targets of interest. The performance of each algorithm is based on the features used for tracking the selected target. Tracking metrics are measures that indicate the performance or the ability of the tracking algorithms to successfully track the selected targets. In looking at the metric developed by different groups, there are similarities in terms of the measurement of particular performance metrics. Also not all metrics developed are able to completely determine the performance of a tracking algorithm and as such other supporting metrics that measure certain other characteristics are required [Denman, Fookes et al., 2009, 541], in order to give better reporting of the tracking ability of an algorithm. Apart from metrics being developed, the different metrics also propose that specific methodologies be employed.

4.2. Background

4.2.1 Ground Truth

The term ground truth is typically used in the field of aerial photography, satellite imaging, and remote sensing. The primary purpose of these activities is to gather data from a distance and the term, ground truth, refers to the data that is being collected. In a simpler form ground truth refers to "what is actually on the ground that needs to be correlated with the corresponding features in the scene (usually as depicted in a photo or image)" [Short, 2009]. A typical example is crop classification using high resolution satellite images by S. N. Omkar et al [Omkar, Senthilnath et al., 2008, 175-182]. Another example is the classification and error assessment of Landsat TM imagery [Schairer, 1998].

A Landsat image can contain various features of interest. These features could include, but not limited to, shapes, size, location, and colour. To establish the ground truth data these features of interest can be identified, and compared to the true land cover. Ground truth could also refer to the process of comparing a single pixel of an acquired image to what is there in reality. This process can verify the contents of the pixel on the image that is being examined.

Generally, ground truth is accomplished by carrying out surface observations and taking measurements of various features of an area of interest utilizing a digital image that has been obtained remotely. Obtaining ground truth data from, for example remote sensing, poses a significant problem. The problem is that the characteristics of the images changes, for example changing seasons, new roads, new buildings, etc. Therefore, ground truth data is subject to errors and uncertainty [Ellis, 2002, 26-31].

One of the early works that applied ground truth to video tracking surveillance is Ellis [Ellis, 2002, 26-31]. However, Ellis noted that applying ground truth to video tracking can present a different set of challenges which he classified into three categories. The first category is the ground truth location of a target based on a bounding box or a reference point on the object such as the top of the object or the centroid of the object. The second category is the accurate identification of the target boundary such as the target shape and target segmentation. The final category is the ground truth which is determined from the classification of the target such as correct identification of the target.

Extracting ground truth can be tedious and time consuming. This is especially true for the first two categories mentioned in the previous paragraph. Ground truth can be determined by going through a video sequence frame-by-frame manually or using a semi-automated tool to characterize the targets in the scene [Ellis, 2002, 26-31]. Manual analysis is prone to errors such as boredom, bias of the assessor, etc. Using a semi-automated tool to generate ground truth has been reported by Yang et al [Yang, Huang et al., 2006, 324-335] for chart image recognition and Wang et al [Wang, Phillips et al., 2001, 528-532] for the development of table detection algorithms. Doermann et al [Doermann and NMihalcik, 2000, 167-170] developed techniques for video analysis and metrics evaluation.

4.2.2 A Framework for Video Performance Evaluation

One of the early video performance evaluation works based on ground truth was proposed by Doermann and Mihalcik [Doermann and NMihalcik, 2000, 167-170]. In this work, Doermann and Mihalcik presented a framework for performance evaluation which uses both temporal and spatial aspects of detection. They outline a reconfigurable Video Performance Evaluation Resource (ViPER) which is intended to provide a ground truth format so that it is possible to represent both static and dynamic descriptors of a video sequence. A GUI was developed to record the necessary information of the video content. To facilitate the performance evaluation of a video sequence, it is necessary to consider whether two descriptions are close enough to satisfy a particular set of constraints, for example, temporal range, spatial locations, or other properties of the scene extracted by the system. The evaluation is based on a hierarchy of matching both time and space. In addition, the evaluation is also divided into detection and localization. Detection is based on the range of frames and localization constraints are used on both temporal range and attributes to determine the correctness of detection. In furtherance of the initial work proposed by Doermann and Mihalcik, various workshops and projects were undertaken that resulted in the development of various performance metrics such as PETS, CLEAR and ETISEO [Nghiem, Bremond et al., 2007, 476-481; Young and Ferryman, 2005, 317-324; Zhai, Shafique et al., 2006]. Another set of performance metrics based on the ground truth obtained from ViPER was proposed by Mariano, et al [Mariano, Min et al., 2002, 965-969] which is based on the overlapping of the bounding box between the tracked target and the ground truth. The best practices in determining the performance metrics and method of tracking can be best described by Thacker et al [Thacker, Clark et al., 2008, 305-334].

4.2.3 Requirements for Effective Performance Analysis

Ellis approach to performance evaluation takes into consideration how the algorithms deal with different physical conditions in the scene, for example, unrelated motion, weather conditions and lighting. The performances of algorithms are assessed using ground truth. The proposed approach compares the tracked data to the marked data, in order to determine whether data is a target position, 2D shape model, or classification of some description.

The aim of performance metrics is to explain the outcome of the algorithm measured against the ‘true’ values determined by the ground truth [Ellis, 2002, 26-31][5]. The metrics are used to quantify different types of errors over the complete dataset. Ellis proposed tracking metrics are derived from data that is placed within a 2x2 contingency table that classifies the correct and incorrect matches between the observed tracks and the ground truth tracks. This is illustrated in Table 4.1. The contingency table is able to display and categorize the data into true positives (N_{tp}), false positives (N_{fp}), true negatives (N_{tn}) and false negatives (N_{fn}) [Ellis, 2002, 26-31].

Table 4.1: 2x2 contingency table [Ellis, 2002, 26-31]

	Ground truth	
Observations	Positive	Negative
Positive	N_{tp}	N_{fp}
Negative	N_{fn}	N_{tn}

The classification of true positive indicates the number of observations confirmed by the ground truth and false positives indicate the number of observations that do not match the ground truth. Other observations such as true negatives are the number of observations that were correctly identified as not belonging to the ground truth and false negatives as observations that were incorrectly identified as belonging to the ground truth. From the observation in the contingency table, a number of values or indices can be derived.
Ellis results show that there is a variation of standard deviation estimated from a frame-by-frame difference of two datasets of video sequences. Ellis concluded that it is beneficial to adopt standards for the format of ground truth data structure, as ground data can be favourably biased towards the algorithm used.

In another work that is closely related to Ellis and Black et al's evaluation metric [Black, Ellis et al., 2003, 125-132] finds the correspondence between ground truths and tracked objects to compute positive and false positive matches to calculate Tracker Detection Rate (TRDR), False Alarm Rate (FAR), Track Detection Rate (TDR) and Track Fragmentation (TF). The purpose of TRDR, and FAR is to measure the object detection rate of the tracking algorithm. In addition, the rate at which individual objects are detected in relation to the ground truth is measured by the TDR. TF is a measure of the number of times a track label changes. All these metrics can be described by the following equations:

$$\text{Tracker Detection Rate (TRDR)} = \frac{\text{Total True Positives}}{\text{Total Number of Ground Truth Points}} \tag{4.1}$$

$$\text{False Alarm Rate (FAR)} = \frac{\text{Total False Positives}}{\text{Total True Positives} + \text{Total False Positives}} \tag{4.2}$$

$$\text{Track Detection Rate (TDR)} = \frac{\text{Number of True Positives for tracked objects}}{\text{Total Number of ground truth points for objects}} \quad (4.3)$$

$$\text{Track Fragmentation} = \text{Number of result tracks matched to ground truth tracks} \quad (4.4)$$

The application of the TRDR, FAR, TDR and TF metrics are based on the detection and location of the ground truth point. A true positive is defined as a ground truth that is located within the bounding box of an object that is detected and tracked by a tracking algorithm. A false negative is a ground truth that is not located within the bounding box of the tracked object. A false positive is an object that does not have a matching ground truth point. The metrics of TRDR and FAR are used to characterize the tracking performance of the object tracking algorithm when tracking multiple objects within the camera view field. The TDR metric is used to indicate the tracking completeness or tracking performance of a specific target. The TF metric indicates the number of times the track label changes where the lower the value of TF the better the performance of the tracking algorithm. The metrics developed from 'hit and miss' rate, contingency tables and the TRDR, FAR and TDR are metrics that can be classified as quantitative values. The difference between TRDR and TDR is that TRDR calculates the overall tracking ability for all tracked objects whereas the TDR metric calculates the tracking ability for a particular target that is being tracked.

Another set of metrics were described and developed by the ETISEO group [Nghiem, Bremond et al., 2007, 15; Nghiem, Bremond et al., 2007, 476-481]. The metrics developed ranged from metrics for object detection, object localization, tracking, classification and event recognition. One of the metrics from ETISEO that was similar to the TDR metric by Ellis was called the precision metric. This metric measures the detection rate of the object of interest [Inrria, 2006]. The equation for precision can be seen in the following equation:

$$\text{Precision} = \frac{\text{number of } GT}{\text{number of } OT} \quad (4.5)$$

The precision metric is a ratio of the good tracking action to the total detection actions. *GT* is defined as a match between the tracked object and the reference data. *OT* is defined as all detected objects which includes both correctly detected objects and wrongly detected objects [Inrria, 2006]. The metric of interest in this paper was the metrics of tracking consistency which are classified as tracking time metric, object ID persistence metric and object ID confusion metric. The main tracking metric is the tracking time metric [Nghiem, Bremond et al., 2007, 476-481]. This metric measures the percentage of time when a reference object is detected and tracked. The metric gives a global overview of the performance of the tracking algorithm. The tracking time metric can be seen in Equation (4.6):

$$T_{Tracked} = \frac{1}{NB_{RefData}} \sum_{RefData} \frac{card(RD \cap C)}{card(RD)} \qquad (4.6)$$

where *RD* is the time interval or area of interest of reference data and *C* is the time interval or area of interest of a candidate data or tracked target. The other complementary metrics to the tracking time metric are the object ID persistence and confusion metrics [Nghiem, Bremond et al., 2007, 476-481]. The persistence metric computes over time the number of tracked objects that are associated to one reference object. The confusion metric computes the number of reference object IDs per detected object. This metric detects the number of ID associated to a reference detected object where a high score on this metric indicates that each detected object is assigned multiple IDs which would indicate highly inaccurate matching between reference data and the tracked item. The equation for both the persistence and confusion metrics can be seen in Equation (4.7) and (4.8) [Inrria, 2006]. The equations of persistence and confusion indicate that the metrics calculate the persistence and confusion capabilities of tracking algorithms for an overall situation where all calculated based on an overall situation where all objects are tracked and not for individual tracked targets.

$$Persistence = \frac{1}{NB_{RefData}} \sum_{RefData} \frac{1}{NumObjectID_{RefData}} \qquad (4.7)$$

$$Confusion = \frac{1}{NB_{DetObjMatchRefData}} \sum_{DetObjMatchRefData} \frac{1}{NumRefID_{DetObj}} \qquad (4.8)$$

Other performance metrics are based on the position [Needham and Boyle, 2003, 278-289] and trajectory [Black, Ellis et al., 2003, 125-132] of the object being tracked. A trajectory of a target can be generally defined as a sequence of positions over a period of time. The metrics based on trajectory are typically obtained by comparing the tracked target trajectory against the trajectory of the ground truth. Object tracking error (OTE) is a metric that is derived from the distance difference between the tracked target against the ground truth [Black, Ellis et al., 2003, 125-132]. OTE is the measure of the mean distance between the target trajectory and ground truth trajectory:

$$Object\ \ Tracking\ \ Error\ (OTE) = \frac{1}{N_{rg}} \sum_{\exists i\, g(t_i) \wedge r(t_i)} \sqrt{(xg_i - xr_i)^2 + (yg_i - yr_i)^2} \tag{4.9}$$

where (xg_i , yg_i) are the coordinates for the ground truth and (xr_i, yr_i) are the coordinates for the detected target in that particular frame i. Typically a comparison of trajectories allows for the determination of similarity or difference between the trajectories [Needham and Boyle, 2003, 278-289]. The combination of metrics of TRDR, FAR and OTE were used in frame-based and object-based tracking in determining the tracks of tracking algorithms [Bashir and Porikli, 2006]. By comparing trajectories, different metrics can be formulated to compare the performance of tracking algorithms, e.g. Equations (4.9), (4.10) and (4.11):

$$d_i = |d_i| = \sqrt{(p_i - x_i)^2 + (q_i - y_i)^2} \tag{4.10}$$

$$m_1 = \mu(d_i) = \frac{1}{n}\sum_{i=1}^{n} d_i \tag{4.11}$$

where the coordinates or locations for the first trajectory are given as (x_i,y_i) and coordinates for the second trajectory are given as (p_i,q_i). The displacement (d_i) between points in a trajectory at any time is given in Equation (4.10). Based on the displacement between trajectory points, the performance metric of m_1 gives the average distance between trajectory positions. The distribution of the trajectory displacement indicates the tracker error spread. This particular metric is used for determining the average displacement or distance difference between the two trajectories that are being compared. The displacement being compared was the displacement or difference of location between the ground truth and the detected object in that image frame. This would mean that this metric measures the average distance or displacement of detected features compared to reference features in the image. This would indicate that the ideal reading for this metric would be no displacement at all between the detected feature and ground truth. A similar metric to the average displacement between trajectories is a metric based on spatially separated trajectories. This type of metric measures the constant spatial difference or displacement between the trajectory of a detected feature and the trajectory of the ground truth. In this instance, the ideal reading from this metric would be a constant displacement reading between the detected feature and the ground truth which would indicate that the detected feature had been consistently detected and had a constant displacement compared to the ground truth.

$$\hat{d} = \mu(d_i) = \frac{1}{n}\sum_{i=1}^{n} d_i \qquad (4.12)$$

4.2.4 Automatic Evaluation System on Object Surveillance

Schlogl, T., et al [Schlogl, Beleznai et al., 2005, 519-522] proposed an evaluation framework called Automatic Evaluation System on Object Surveillance (AESOS). AESOS comprises a set of error metrics and video reference data. The proposed framework demonstrated how motion segmentation and tracking can be used to determine the operational range of and entire surveillance system. As this paper is concerned with tracking in video surveillance, this section will focus only on the points salient to tracking evaluation. The proposed AESOS system generates scenes by selecting the number of objects, their trajectories and velocities. Simple sequences are used for the evaluation of the operational range.

Three spatial metrics were used to measure the performance. The metrics used include a Hit Rate (*H*), Miss Rate (*M*) and False attempt (*F*) [Schlogl, Beleznai et al., 2005, 519-522]. The hit rate is the number of successfully detected objects in a frame over the number of ground truth objects and these numbers are summed over all the frames in the video. In a similar manner, the miss rate is the number of undetected objects in a frame over the total number of ground truth points. Typically the miss rate is complementary to the hit rate. In addition to the hit and miss rate, another metric was proposed which was the false attempts. False attempts were classified as detections or tracks that did not correspond to any ground truth points or false tracking of objects.

4.2.5 Tracking Algorithms Using Object Labels

This particular work proposes a new performance evaluation called Label and Size Based Evaluation Measure (LSBEM) [Popoola and Amer, 2008, 733-736]. Basically, objects are assigned an unique label. This label and the size of the object are used for evaluating a tracking algorithm. LSBEM is calculated by using a metric, which compares the object unique label and size of each object to the unique label and size of the ground truth object. LSBEM metrics consist of Object Detection Rate (OBR), Average Size Detection Rate (ASDR), Label Tracking Detection Rate (LTDR), and Non-Label Tracking Detection Rate (NTDR).

The efficiency of the LSBEM is compared to the performance evaluation metrics proposed by Black el at [Black, Ellis et al., 2003, 125-132] which will be discussed in the next section. Six video sequences were used to evaluate the LSBEM. The results show that the LSBEM is capable of providing a better evaluation of the performance of tracking algorithms than the evaluation metrics mentioned in Black's work

4.3. Supporting Performance Metrics

Previous performance metrics such as TRDR, TDR, OTE and trajectory comparison were developed to assess the ability of a tracking algorithm in correctly identifying and tracking a target. The previously mentioned metrics only consider the performance in terms of correctly tracking a specific target. Additional or supporting metrics are required to determine the consistency or accuracy of a tracking algorithm. Among the metrics that were developed was the object tracking standard deviation (OTStd) [Nillius and Eklundh, 2002; Sebastian and Vooi Voon, 2007]. The OTStd metric is an extension of the OTE metric where it can be used together with OTE. OTStd can also be used in a tracking system that compares trajectories to determine the deviation between the tracked points and the ground truths [Trucco and Plakas, 2006, 520-529]. The OTStd metric provides a measure of the consistency of the tracked feature where the larger OTStd measure indicates that the tracking algorithm is less accurate. The OTStd measure is derived in the following Equation (4.13):

$$OTStd = \sqrt{\frac{1}{N_r} \sum_{\exists i\, g(t_i) \wedge r(t_i)} (dr_i - OTE)^2} \qquad (4.13)$$

The OTStd measure is calculated by obtaining the square root of the sum of squares of the subtraction of the track difference measurement (dr_i) against the OTE measure which is then divided by the number of frames (N_{rg}) that the ground truth is measured. This metric is typically used together with the OTE metric to measure the spread of the OTE measurement. In other words, the OTStd metric measures the accuracy or consistency of the OTE metric.

This metric is a measure of the accuracy or measure of deviation of the feature being tracked against the ground truth or reference location of the feature being tracked. The tracking metric of OTStd can be used together with the TDR to determine the performance of a tracking algorithm in different colour spaces [Nillius and Eklundh, 2002; Sebastian and Vooi Voon, 2007]. The result from the tracking experiment and evaluating the usage of the OTStd metric together with the TDR indicated that the colour spaces that have luminance and chrominance components have better tracking performance compared to RGB and grayscale colour spaces. The video used in the investigation of tracking metrics is a video of a face tracking algorithm. A snapshot of the tracking video can be seen in Figure 4.1. The snapshot in Figure 4.1 shows the reference image on the left side and the tracked image or video on the right side in Figure 4.1. The other videos used in determining the tracking results can be seen in Figure 4.4, Figure 4.5 and Figure 4.6 as snapshots for face_seq_2, face_seq_3 and face_seq_4 respectively. The tracking method used in this section is the normalized cross correlation method which is described in Section 5.2 which uses a reference image to find or locate a similar portion of the input image by means of normalized cross correlation. A reference image is used in determining or locating the reference feature in the tracked image. A dot in the tracked image indicates a similar point or feature from the reference image or reference point being tracked. A track is considered successful when the tracking dot remains within the region of the face that is being tracked which has a rectangle bounding box or area around the detected face. A successful or correct track can be seen in Figure 4.2 where the track dot remains within the face region in the tracked image or the tracked dot remains within the bounding box of the detected face region. An unsuccessful or incorrect track can be seen in Figure 4.3 that shows the track dot at a location outside of the face region in the tracking video. The initial tracking work done here was on face tracking and continued on to tracking a whole person. In addition to the tracking work done, different tracking metrics were evaluated on the viability of each metric to measure the performance of each tracking methodology.

Figure 4.1: Snapshot of tracking action on face_seq_1

Figure 4.2: Snapshot of correct tracking on face_seq_1

Figure 4.3: Snapshot of incorrect tracking on face_seq_1

Figure 4.4: Snapshot of tracking action on face_seq_2

Figure 4.5: Snapshot of tracking action on face_seq_3

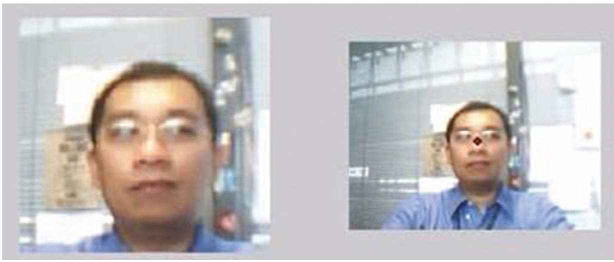


Figure 4.6: Snapshot of tracking action on face_seq_4

The videos used in the determination of tracking performance as illustrated in Figure 4.2 and Figure 4.3 have different targets tracked with different backgrounds and their tracking performance compared. The videos were used to determine the performance of the algorithm in tracking a face over different colour spaces. The comparison of tracking performances of facial tracking across different colour spaces starting from grayscale, RGB, HSV and YCbCr colour spaces can be seen in Figure 4.7 and Figure 4.8 where the tracking performances are compared in terms of TDR and OTStd. The TDR metric was used as it was relatively easy to determine the number of frames the tracking point or dot is within the facial region over the total number of frames where the face is present in the video. The OTStd is a measure of consistency of the tracking point against a reference point. In this instance for determining the performance of a tracking algorithm, a higher TDR value together with lower OTStd value would indicate better tracking performance from one colour space compared to another.

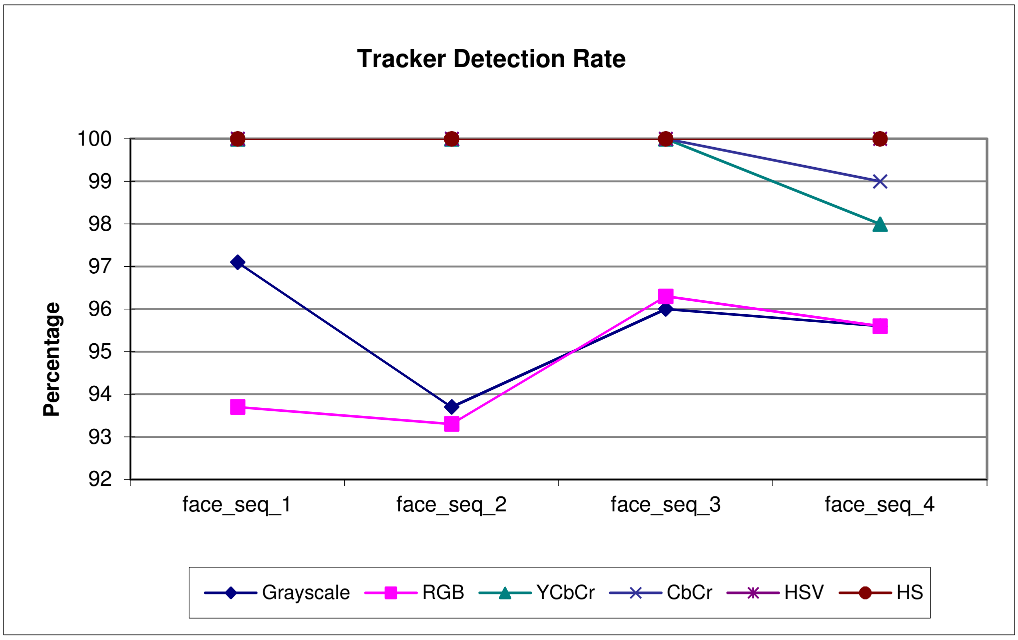


Figure 4.7: Track Detection Rate

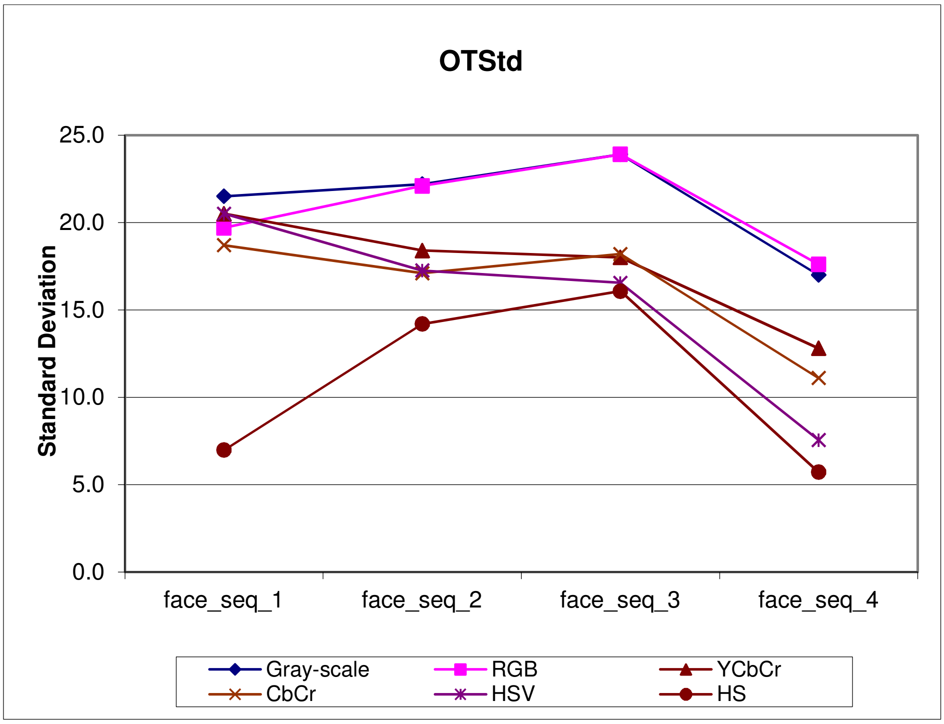


Figure 4.8: Object Tracking Standard Deviation

The results from the graphs in Figure 4.7 and Figure 4.8 indicate that for the different videos used, the colour spaces of HSV and YCbCr have consistently given higher TDR and lower OTStd results respectively compared to RGB and grayscale colour spaces. From observing the graphs, there are indications that the chrominance and chromaticity data from the YCbCr and HSV colour spaces can be used for tracking purposes without the luminance data. This conclusion can be obtained by observing the data from the OTStd metric that is not able to be determined from the TDR metric. Looking further in the tracking performance in Figure 4.7, Figure 4.8 and in Chapter 5 which indicates that the colour tracking performance in different colour spaces differed where the YCbCr and HSV colour spaces had better tracking performances. In the final selection of colour space, YCbCr was selected due to better tracking performance compared to HSV colour space. HSV was not selected due to additional performance in Chapter 6 that indicated that the HSV colour space could not be used for tracking a whole person target. This is because of the difference in the information stored in the different layers. The HSV has colour information stored in the H layer and purity of the colour stored in the S layer whereas the YCbCr colour space has colour difference information stored on both the Cb and Cr layers.

Another metric which can be used to characterize the performance of a tracking algorithm is tracking consistency (TC) [Sebastian and Vooi Voon, 2009, 318-322]. The TC metric is a metric that can complement the TDR metric as it measures the consistency or robustness of a tracking algorithm in maintaining the tracking action on a target object. The definition of consistency in this instance would be the act of consistently tracking the correct target without losing or missing the track. The tracking consistency score is determined by Equation (4.14).

$$Tracking\ Consistency = \frac{\sum_{b=1}^{n_2} \sum_{a=1}^{n_1} S_{ab}}{\left(\frac{n_f (n_f + 1)}{2} \right)} \tag{4.14}$$

The S_{ab} term in the TC equation is the score that has been assigned for correctly tracked targets. The determination of a correct track is based on the tracking algorithm that checks if the target being observed is the correct target and the tracked position is compared against the ground truth positions that are determined as stated in Section 4.2.1. The score can be seen in Table 4.2 which illustrates the method in assigning a tracking score [Sebastian and Vooi Voon, 2009, 318-322]. The data illustrated in Table 4.2 is a sample of a tracking action. The correct track column indicates a successful or correct track with a '1' and '0' otherwise. With the start of each correct track, S_{ab} is initialized with a score of 1 with the score incrementing by one for each correct successive track. The n_1 variable in the TC equation indicates the number of consecutive correct tracks in a string of consecutive tracks and the n_2 variable is the number of sets of consecutive tracks. The n_f variable in the TC equation is used to designate the number of frames that the tracked target is within the camera view or the number of ground truth points.

TABLE 4.2: SAMPLE OF TRACKING TABULATION

Frame	Correct Track	Score (S_{ab})
1	1	1
2	1	2
3	1	3
4	0	0
5	0	0
6	1	1
7	1	2
8	1	3
9	1	4
10	1	5

The TC metric was initially meant to complement the TDR metric where a relation between the TC and TDR scores can be seen in a comparison table based on simulated data shown in Table 4.3. The simulated tracking data in Table 4.3 is generated for different levels of TDR in a study to link the TDR to TC values. The relationship between the TC and TDR values indicate that the lower the TC value for a corresponding TDR value, the more fragmented the tracking would be. The number of fragmented tracks is indicated in the track fragmentation (TF) value. The TF is determined by observing the number of times that the track is 'lost' or not tracking its intended target [Black, Ellis et al., 2003, 125-132]. The TC values displayed in Table 4.3 utilizes simulated data generated from Table 4.2 that corresponds to the ratio of TDR to TF values. As listed in Table 4.3, a higher TF value has lower TC values for the corresponding TDR value. The values in Table 4.3 are the TC value for different levels of TDR and at different values of TF. As an example from Table 4.3 where at TDR level of 0.9 and a TF value of 1, the TC value would be at 0.81. The TC values at different TDR values show that the tracking performance deteriorates when the number of TF increases or the number of times that track is lost.

TABLE 4.3: TC RELATIONSHIP TO TDR

TDR	Track Fragmentation (TF)			
	TF =1	*TF = 2*	*TF = 3*	*TF = 4*
0.9	0.81	0.62	0.35	0.27
0.8	0.65	0.44	0.27	0.21
0.7	0.50	0.36	0.21	0.16
0.6	0.37	0.28	0.16	0.12
0.5	0.26	0.18	0.10	0.08

In determining TC in a tracking situation, the TDR vs. TC results can be seen in Table 4.4. The TC metric listed in Table 4.4 was generated from the same videos used to generate the tracking result illustrated in Figure 4.7 and Figure 4.8. The results listed in Table 4.4 show that for the various TDR values, the corresponding TC values demonstrate that the consistency of tracking did not correlate to the related TDR value. This indicates that the tracking was fragmented with the tracking action being lost or not maintained and then being reacquired throughout the overall tracking operation. Ideally the TDR and TC should be at 1.0 for corresponding video sequences and colour spaces. However, it can be seen for the grayscale colour space, despite the high TDR rates of above 0.9, the corresponding TC values were less than 0.5 which indicate that there were 2 or more times that the track action on the target of interest was lost.

TABLE 4.4: TDR VS. TC

	TDR				TC			
Color space	face_seq_1 (TDR)	face_seq_2 (TDR)	face_seq_3 (TDR)	face_seq_4 (TDR)	face_seq_1(TC)	face_seq_2 (TC)	face_seq_3 (TC)	face_seq_4 (TC)
Grayscale	0.97	0.94	0.96	0.96	0.26	0.48	0.23	0.48
RGB	0.94	0.93	0.96	0.96	0.25	0.48	0.24	0.48
YCbCr	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.49
CbCr	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.50
HSV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
HS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

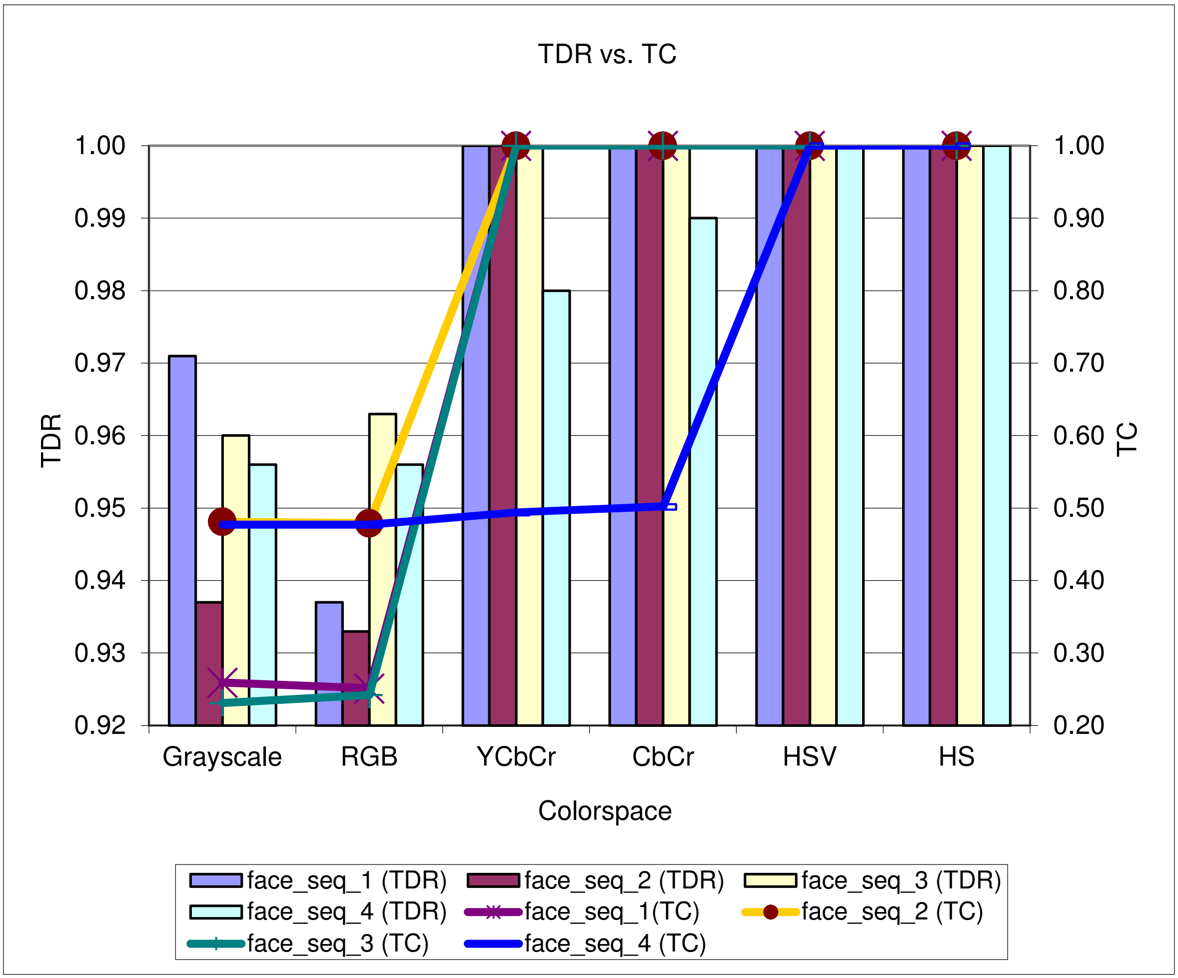


Figure 4.9: TDR vs. TC comparison

4.4 Comparison of Tracking Metrics

In introducing a new metric to measure the performance of a tracking algorithm, a comparison against the existing metric has to be done. The metric in question is the TC metric which is a measure of the consistency of the tracking action of an algorithm. The existing metric for determining the tracking performance of consistent tracking was listed as tracking success rate (TSR) [Zivkovic and Heijden, 2006, 773-780]. The equation for TSR is as listed in (4.15):

$$Tracking\ \ Success\ \ Rate\ (TSR) = \frac{Number\ \ of\ \ non-fragmented\ \ objects}{Total\ number\ \ of\ \ ground\ \ truth\ \ objects} \qquad (4.15)$$

TSR reports the performance of the tracking algorithm with respect to the track fragmentation or to the number of times the tracking is maintained or the consistency of the tracking action. Keeping in mind the type of information required to report TSR of a tracking, the TSR for the data listed in Table 4.4 can be seen as follows in Table 4.5. The definition of the number of non-fragmented objects can be defined as the number of consecutive or successive tracks of the tracking algorithm. Tracking fragmentation (TF) indicates the number of non-fragmented tracks.

The TSR metric, as indicated in Table 4.5, showed that the tracking action was above 90% for all different colour spaces which is similar to the TDR metric as seen in Table 4.4. Looking at the data in Table 4.5 there were no TF readings for some of the tracking videos in the different videos. This indicates that for the respective videos with no TF values that there were no breaks in the tracking action. If the TDR metric in Table 4.4 is compared to the TSR metric in Table 4.5, the TSR metric can be seen to be similar to the TDR. It would seem that the TSR metric does not reflect the ability of tracking a target consistently especially when taking into account tracking fragmentation. The TSR metric does not indicate or reflect the effect of fragmentation into the metric which the TC metric indicates. As indicated in Table 4.5, the higher the TF value, the TC metric would be lower as this would indicate that the tracking is more fragmented. This is in contrast to the TSR metric which does not indicate any difference in its measure with different TF values. Based on the data in Table 4.5, the TC metric provides a better indication of the tracking consistency compared to the TSR metric which is further proven by the accompanying TF metric.

In validating the performance metric, standardized data was obtained from ETISEO which is a project that does performance evaluation on video surveillance systems. This project was sponsored by the French government with multiple international partners such as York University Canada, University of Southern California, University of Reading, Universite Paris Dauhphine and University of Genoa. The standardized metrics were developed and the standard videos selected by the multiple partners in the ETISEO project [Kim, Cho et al., 2004, 785-788].

In applying the TC metric to a set of standard videos and results used by the ETISEO group, a comparison between the tracking time metric by ETISEO to the TC metric being proposed can be done. The data from the ETISEO video would also have results of the other metrics that have been proposed by ETISEO [Ali, Indupalli et al., 2006; Chen, Li et al., 2005, 1766-1768; Kim, Cho et al., 2004, 785-788] which can be used to compare to other metrics proposed by Black and Ellis [24]. The table of the compared data can be seen in Table 4.6 [Sebastian, Vooi Voon et al., 2011, 493-502].

Table 4.5: TSR vs TC vs TF

	TSR				**TC**				**TF**			
Colour space	*face_seq_1*	*face_seq_2*	*face_seq_3*	*face_seq_4*	*face_seq_1*	*face_seq_2*	*face_seq_3*	*face_seq_4*	*face_seq_1*	*face_seq_2*	*face_seq_3*	*face_seq_4*
Grayscale	0.96	0.93	0.95	0.95	0.26	0.48	0.23	0.48	5	2	5	2
RGB	0.92	0.93	0.95	0.95	0.25	0.48	0.24	0.48	5	2	5	2
YCbCr	1.00	1.00	1.00	0.97	1.00	1.00	1.00	0.49	0	0	0	2
CbCr	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.50	0	0	0	2
HSV	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	0	0	0
HS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	0	0	0

Table 4.6: ETISEO Metric compared to Other Metrics

Video	**ETISEO**				**Ellis**		**TC**
	Precision	Tracking Time	Persistence	Confusion	TDR	TF	
ETI-VS2-AP-11-C4	0.47	0.84	0.93	1.00	0.47	2	0.12
ETI-VS2-MO-1-C1	0.42	0.20	0.67	0.75	0.42	3	0.20
ETI-VS2-RD-10-C4	0.94	0.27	0.63	0.72	0.94	9	0.13
ETI-VS2-RD-6-C7	0.99	0.53	1.00	0.80	0.99	1	0.38
ETI-VS2-BE-19-C4	0.37	0.20	0.26	0.84	0.37	13	0.02
ETI-VS2-BE-19-C1	0.79	0.47	0.77	0.56	0.79	5	0.16

Table 4.6 lists the metrics determined from the different standard videos used by the ETISEO group which are namely the precision, tracking time, persistence and confusion metrics [Ali, Indupalli et al., 2006; Chen, Li et al., 2005, 1766-1768; Kim, Cho et al., 2004, 785-788]. The other metrics listed in comparison were the TDR and TF metrics proposed by Black and Ellis [Stern and Efros, 2002, 236-241; Zivkovic and Heijden, 2006, 773-

780]. The final metric of TC in Table 4.6 is the current proposed metric meant to determine the consistency of a tracking algorithm in tracking a specific target.

Looking at the data in Table 4.6, similar metric values can be seen in the precision metric from ETISEO and the TDR metric by Black and Ellis. The similarity in the values between the different metrics indicated that both metrics were measuring the same item. By looking at the tracking time, persistence and confusion metrics and comparing them to the precision and TDR metrics, would not indicate that there is no relationship between these metrics. The tracking time, persistence and confusion metrics are based on tracking multiple targets simultaneously. The precision and TDR values obtained for each video was based on tracking a specific target in this instance it was the tracking of a person walking through a camera view field. In this case the consistency metrics of tracking time, persistence and confusion would not be able to give a proper measure of tracking a single moving target such as a person walking within a camera view field. This can be seen in the situation where for video ETI-VS2-AP-11-C4 the precision metric has a value of 0.47 whereas the tracking time has a value of 0.84. Results such as this do not make sense where the tracking time metric was higher compared to the precision metric where it has higher consistency in tracking compared to correctly detecting the target.

Looking at Table 4.6 that compares tracking performance of standard videos from ETISEO based on performance metrics by ETISEO and Ellis, a relationship between the TC, TDR and TF metrics can be inferred. As listed in Table 4.6, the TC metric value is an approximation of the TDR metric divided by the TF value. The TC metric gives an approximation of the consistency of the tracking action.

When comparing the tracking consistency metric in Table 4.5, results had shown that the TC metric gave indication in terms of the measurement of the consistency of the tracking consistency compared to TSR. This would indicate that the TC metric would be the best metric for measuring tracking consistency. Looking at Table 4.6 there is no difference in the tracking performance as reported by both Precision and TDR which indicates that neither one of the tracking performance metrics would give better reporting capability.

4.5 Conclusion

This section examines metrics used to determine the performance of tracking algorithms. The measure of a tracking algorithm is its ability to determine the correctness and the accuracy of each tracked target. These measures are able to give an indication on the tracking ability of each tracking algorithm. The metrics available cover both quantitative and qualitative metrics where the usage of both classes of metrics, such as TDR and OTE, give a more accurate picture than a simple usage of one metric. Additional metrics are introduced to further determine the accuracy of a tracking algorithm in OTStd and the consistency of a tracking algorithm in TC. The TC metric is a newly introduced metric that measures the consistency of a tracking algorithm in maintaining the tracking action. The TC metric was used on standard videos used by the ETISEO group that had developed their own set of metrics to measure the performance of tracking algorithms. A comparison of the metrics has also shown that there are similarities between metrics developed by different people and organization such as the precision metric by ETISEO and the TDR metric by Black. It was also shown that the utilization of multiple metrics would be able to provide a better indication of the tracking ability of a tracking algorithm compared to a single tracking metric. In this chapter, the selected performance metric was the TDR metric which will be supported by the TC metric. The selection of TDR over OTE does not make any difference as both these performance metrics report the same item. The selection of TC as a secondary metric is a required step as there are no suitable metrics that could report or complement the TDR metric.

CHAPTER 5

Correlation Tracking

5.1. Introduction

Object tracking in video surveillance is the act of detecting a reference object within an input video or a series of images. The act of tracking a target is the main function in a video surveillance system where the object of interest would be identified either by an operator or a set of reference information and then be tracked while the object of interest is within the camera view field. Typical targets of interest in video surveillance systems would be people and is the area of study and research of this work. The tracking of people moving through a secluded or security sensitive area such as a prison courtyard, a secured document storage area, a public area such as a park or a shopping mall, is where the application of video surveillance is becoming prevalent. This is becoming true when areas to be monitored are spread over a large region and hence require a large number of cameras to observe and monitor the area of interest.

5.2. Correlation Tracking

Cross correlation is the correlation of two different signals and is the method of tracking that is investigated in this section. The application of correlation in tracking a target is one approach to feature detection [Lewis, 1995, 120-123]. One approach to feature detection using cross correlation is called template matching, where the template matching operation is based on the task of finding the best match between two different images [Gonzalez and Woods, 2002; Lewis, 1995, 120-123]. The utilization of template matching in tracking a target is based on the act of using a reference image, or pattern, that is compared to the input image to find the location of the reference image within the input image. The correlation operation between two images can be described in Equation (5.1):

$$c(u,v) = f(x,y)\,t(x-u,y-v) \qquad (5.1)$$

where *f(x,y)* is the feature or input image and t(x,y) is the sub-image or the reference image or the feature that is to be detected within the input image [Sebastian and Vooi Voon, 2007]. Equation (5.1) gives an indicator or a measure of the similarity between the input image and the reference feature or reference image. This method is considered in the tracking of a target as the reference image can be found in the input images.

Apart from cross correlation, normalized cross correlation is another application of cross correlation that can be used in tracking a target. Normalized Cross Correlation (NCC) is the operation where a template is shifted into different positions and at each position the intensities are multiplied and summed to produce a final normalized cross correlation matrix. The NCC equation matrix can be described as:

$$\gamma(u,v) = \frac{\sum_{x,y}\left[f(x,y) - \bar{f}_{u,v}\right]\left[t(x-u,-v) - t\right]}{\left\{\sum_{x,y}\left[f(x,y) - \bar{f}_{u,v}\right]^2 \sum_{x,y}\left[t(x-u,y-v) - \bar{t}\right]^2\right\}^{0.5}} \tag{5.2}$$

where *f(x,y)* is the intensity value of an image *f* at pixel *(x,y)*. Similarly, *t(x,y)* is the intensity value of the template image *t* at pixel *(x,y)*. The NCC is computed at every point *(u,v)* for *f* and *t*, which has been shifted over the original image *f(x,y)* by *u* steps in the x-direction and *v* steps in the y-direction [Hii, Hann et al., 2006, 144-156]. Lastly $\bar{t}$ is the mean of the mean value of the template *t* and $\bar{f}_{u,v}$ is the mean of *f(x,y)* in the region of the template shifted by *(u,v)* steps. All the NCC coefficients are stored in a correlation matrix defined by Equation (5.2). The NCC method was investigated due to the presence of the same target as the target moves across an array of cameras. The utilization of the reference image to detect and track a target in different backgrounds or scenes made the NCC method a starting point in tracking across an array of cameras. With that in mind the NCC method was investigated for the feasibility of being used as a means of detecting and tracking a target across an array of cameras.

5.3. Face Correlation Tracking

Correlation tracking was investigated using the face as the reference image in tracking a target. The tracking of a face requires that the reference image be obtained first before the correlation operation is done.

5.3.1. Face Detection

Skin colour provides an attribute that can be useful in tracking a facial target [Soriano, Martinkauppi et al., 2000, 839-842]. In addition to the information of skin colour, area of the face is another set of data that is provided [Kawato and Ohya, 2000, 141-1418]. Also by utilizing information of skin colour, a face detector can be developed by creating a skin colour mask. The development of a skin colour mask is able to filter out all other pixels in an image except the portions that fit within the skin colour range [Beetz, Radig et al., 2006, 39-42]. This indicates that the each pixel in an image can be classified either as skin or non-skin.

Skin colour has variations due to brightness, skin reflectance, emotional condition, sun tan, etc. [Park, Seo et al., 2000, 133-136]. In order to have robustness in determining skin colour, chrominance is separated from luminance in the original colour space [Soriano, Martinkauppi et al., 2000, 839-842]. This is to reduce the effect of varying levels of brightness or illumination in images used to identify skin [Beetz, Radig et al., 2006, 39-42]. It was also determined that different shades of skin colour have similar chromaticity [Soriano, Martinkauppi et al., 2000, 839-842]. Based on the chromaticity properties of skin, a mask or a segmentation step can be done to determine the portions or pixels that can be classified either as skin and non-skin [Beetz, Radig et al., 2006, 39-42; Park, Seo et al., 2000, 133-136]. The utilization of skin colour properties ensures that the correlation tracking is only done on the face. An example of skin segmentation to detect a facial image can be seen in the following Figure 5.1. The skin segmentation illustrated in Figure 5.1 was done in the YCbCr colour space. For skin segmentation the pixel values would have to fall within a specific range, in the Y colour layer the pixel range would be between 105 and 117, in the Cb colour layer the pixel value range would be between 110 and 113 and for the Cr colour layer the pixel value range would be higher than 128. Based on the skin segmentation done on the reference image, it can be observed that there is area that has been segmented as a skin region which can be classified as an error in segmentation.

Utilization of the reference image together with segmentation errors could potentially lead to incorrect tracking. However, the larger skin segmented area compared to the segmentation 'error' would give higher correlation values compared to the segmentation 'errors' and thus lead to correct tracking.

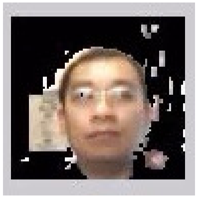


Figure 5.1: Skin segmented reference image.

5.3.2. Face Tracking Results

In order to implement the Normalized Cross Correlation tracking, the peak location of the correlation map needs to be located from the NCC correlation matrix as listed in Equation (5.2). The NCC operation generates a correlation matrix which has all the correlation results, from which the highest correlation point can be determined. The NCC tracking is based on the highest correlation point in the correlation matrix. This can be seen in the examples in Figure 5.2 and Figure 5.3. The sample images were captured from a standard web camera at a resolution of 320x240 pixels with the reference image having a size of 160x110 pixels. The image display, correlation display and correlation calculation was done in Matlab 2008a.

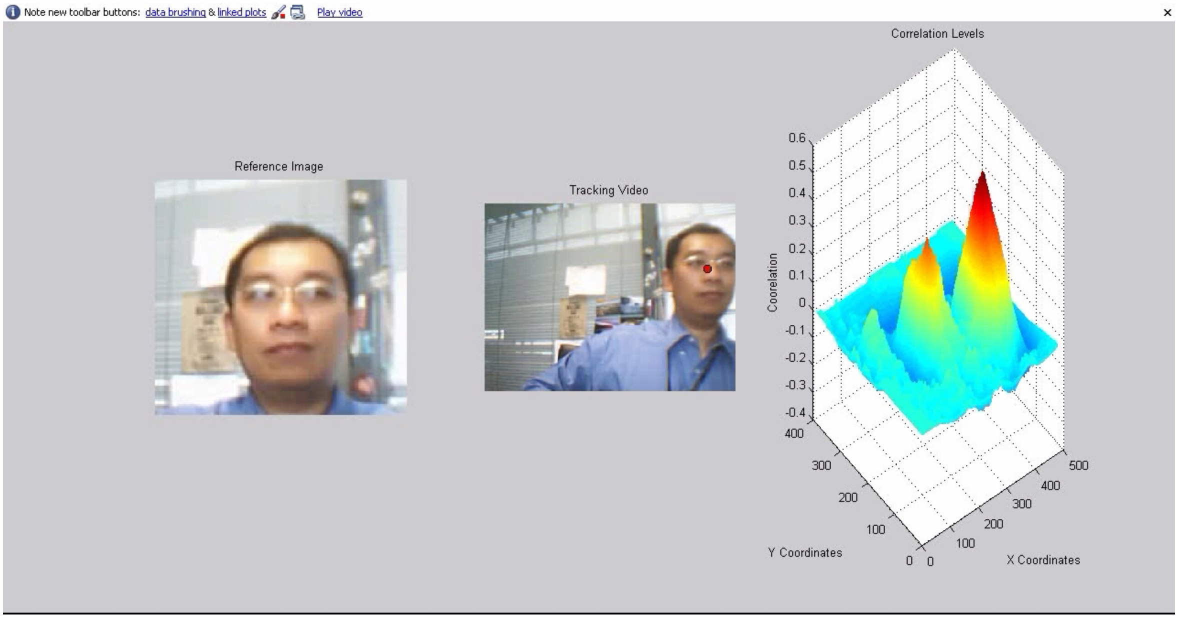


Figure 5.2: Face correlation tracking sample 1

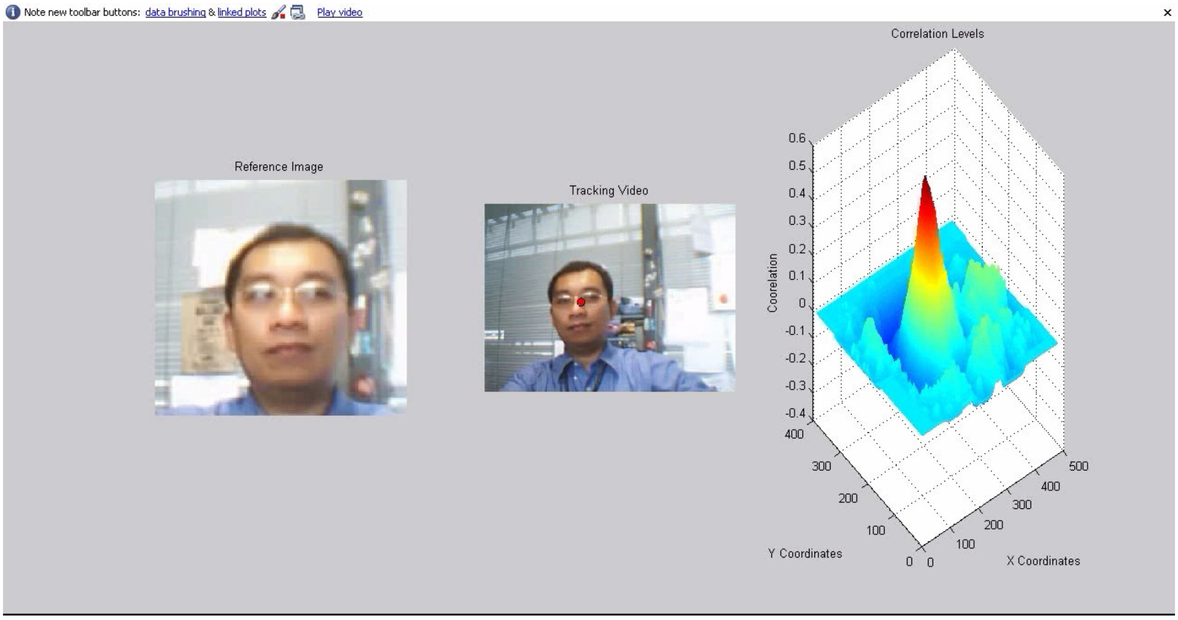


Figure 5.3: Face correlation tracking sample 2

In both Figure 5.2 and Figure 5.3, the correlation matrix is illustrated on the right side of the images. The highest point in the correlation matrix is designated by a tracking point in the tracking image which can be seen in Figure 5.2 and Figure 5.3. The peak location obtained from the NCC matrix will change as the tracked face changes position which is reflected in the target image where the tracking point is re-mapped. The NCC tracking operation is considered to be correct or successful when the tracking point is within a specific region. The region in this instance is a square region around the face.

The tracking performance of NCC face tracking is reported using the tracking performance metric of Tracker Detection Rate (TDR). The TDR metric reports the rate of correctly detecting the reference image. The results listed in Table 5.1 lists the TDR rate when using the NCC tracking method in different colour spaces [Sebastian and Vooi Voon, 2007; Sebastian, Vooi Voon et al., 2008, 2512-2516].

Colour space	face_seq _1	face_seq _2	face_seq _3	face_seq _4
Grayscale	0.97	0.94	0.96	0.96
RGB	0.94	0.93	0.96	0.96
YCbCr	1.00	1.00	1.00	0.98
CbCr	1.00	1.00	1.00	0.99
HSV	1.00	1.00	1.00	1.00
HS	1.00	1.00	1.00	1.00

Table 5.1: TDR rate using NCC tracking method in different colour spaces

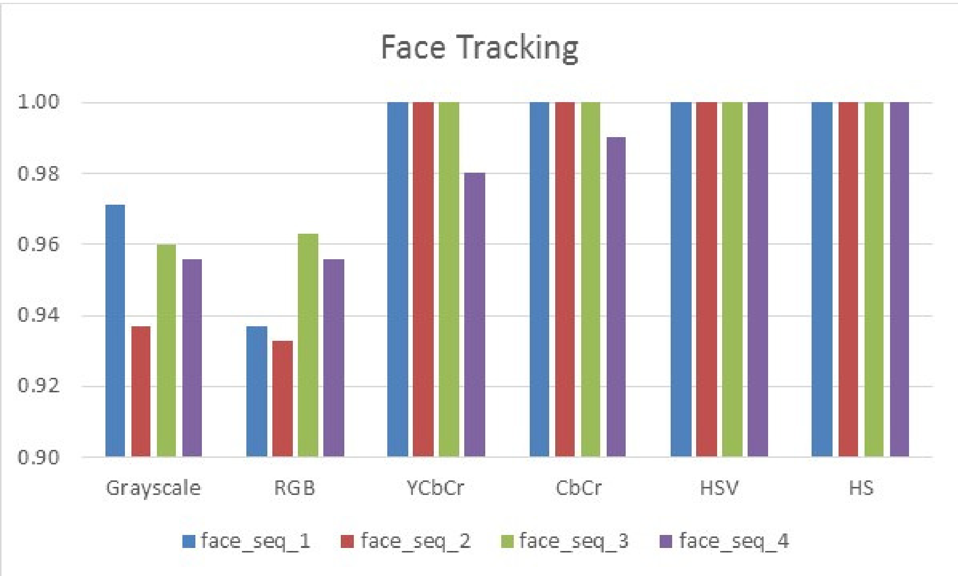


Figure 5.4: Face detection rate in different colour space

The results shown in Table 5.1 with the illustration of the data in Figure 5.4, show the face detection rate in different colour spaces. These results indicate that colour space has an effect on the NCC tracking capability in tracking and detecting a face. The results listed in Table 5.1 have results ranging from 0.93 to 1.0. Those values indicate the rate of correctly detecting the target of interest where 0.93 is an indication that 93% of total track action have resulted in correct track of target. In an ideal situation a TDR value of 1.0 is desired which indicates that the target is correctly tracked the whole time the target is visible in the camera view field. A correct track in face detection is when the highest correlation point is within the face region, which can be a rectangle area surrounding the face, and if the correlation point is outside the face region the track would be classified as an incorrect track. The reference images and sample snapshots of NCC tracking on the face can be seen in the following Figures 5.5 to 5.8.

Figure 5.5: face_seq_1 snapshot

Figure 5.6: face_seq_2 snapshot

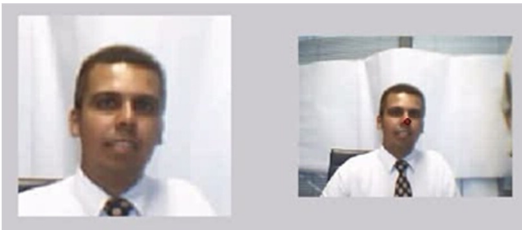


Figure 5.7: face_seq_3 snapshot

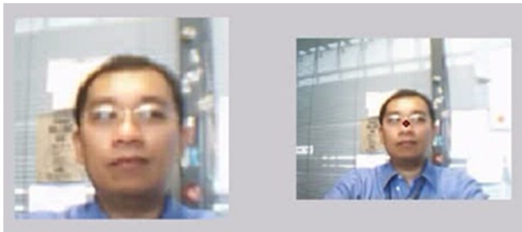


Figure 5.8: face_seq_4 snapshot

Input video	Time (s)	Number of Frames (frame count)	Resolution (pixel x pixel)	Frame Rate (fps)
face_seq_1	27.2	816	320 x 240	30
Face_seq_2	19.5	586	320 x 240	30
Face_seq_3	41.4	1243	320 x 240	30
face_seq_4	14.9	449	320 x 240	30

Table 5.2: Summary of Input Video Specifications for face tracking

The snapshots listed in Figure 5.5 to Figure 5.8 have the reference images on the left side of each figure. The source of the sample images shown in Figures 5.5 to 5.8 are taken from captured videos labelled as face_seq_1, face_seq_2, face_seq_3 and face_seq_4 respectively. A summary of the input video sources are listed in Table 5.2 that lists the elapsed time, number of frames, video resolution and frame rates of each video. These videos were captured through a web camera at a resolution of 320x240 pixels with no video compression techniques applied. Each captured video had different recorded lengths at 27.2, 19.5, 41.3 and 14.9 seconds respectively. These captured videos were used to examine NCC tracking on different skin colours. In terms of skin colour or tone, the primary observation from the data of NCC tracking from Table 5.1 indicates that tracking rates differ in different colour spaces. The tracking results of TDR will range from 0 to 1, with 1 as having 100% tracking rate and 0 as having 0 % tracking capability. From the results listed in Table 5.1, the colour spaces of YCbCr and HSV have higher tracking rates compared to RGB and grayscale colour spaces. The colour spaces of YCbCr and HSV have the colour information on different layers from the illumination information which is different from the grayscale and RGB colour spaces which have the colour information and illumination information embedded together.

Based on the data in Table 5.1, tracking performance on grayscale colour space is in the range from 0.94 to 0.97 and the RGB colour space has the TDR rate of 0.93 to 0.96. In both cases the results have similar tracking abilities as both colour spaces have illumination or brightness information embedded with colour information. The other colour spaces of YCbCr and HSV have better TDR results which range from 0.98 to 1.0. In addition to having higher tracking performance, the utilization of specific information from the YCbCr and HSV colour space was demonstrated in tracking a face. The utilization of specific layers such as the Cb and Cr layers in the YCbCr colour space and H and S layers in the HSV colour space was demonstrated as a comparable method of tracking.

5.4. Whole Correlation Tracking

As an extension of face tracking using Normalized Cross Correlation (NCC), the NCC tracking method was implemented for tracking a whole person. As such the reference image used would not be one specific region as was done on the face tracking operation and in this instance the whole person is used as the reference image. The rationale for extending or continuing the application of NCC from face tracking to whole person tracking was based on the initial results seen in Section 5.3. The tracking results had indicated that face tracking could be successfully done on specific colour space for better tracking results. Face tracking uses a single colour region as a means for tracking whereas whole person tracking is a tracking method that uses a whole person, which is made up of multiple colour regions, as a means for tracking using NCC. This section investigates the utilization of NCC together with a whole person or a multiple colour region as a means for tracking a target. Samples of the images used for whole person tracking can be seen in the following figures.

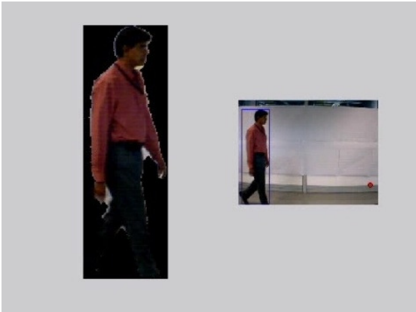


Figure 5.9: whole_seq1_1 snapshot

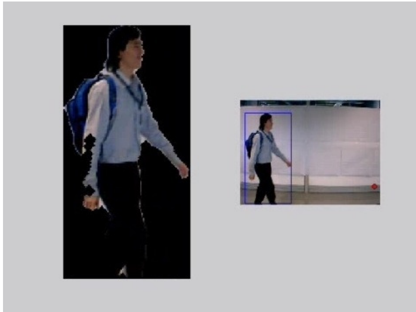


Figure 5.10: whole_seq1_2 snapshot

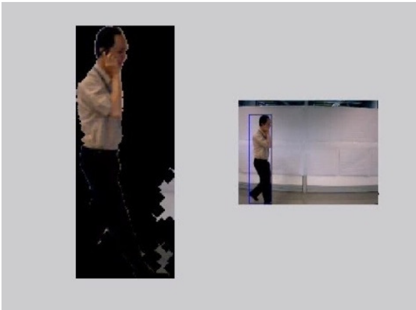


Figure 5.11: whole_seq1_3 snapshot

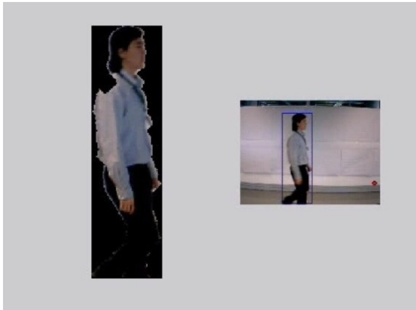


Figure 5.12: whole_seq1_4 snapshot

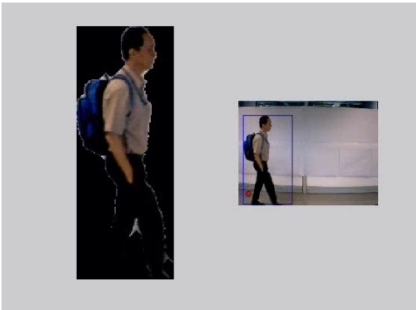


Figure 5.13: whole_seq1_5 snapshot

Input video	Time (s)	Number of Frames (frame count)	Resolution (pixel x pixel)	Frame Rate (fps)
whole_seq1_1	10.8	324	320 x 240	30
whole_seq1_2	7.8	234	320 x 240	30
whole_seq1_3	7.3	218	320 x 240	30
whole_seq1_4	8.5	256	320 x 240	30
whole_seq1_5	8.8	264	320 x 240	30

Table 5.3: Summary of Input Video Specifications for whole person tracking

In Figure 5.9 to Figure 5.13, the reference images used in the NCC operation can be found on the left side of the figures. The correlation operation on a whole person generates a correlation matrix similar to the illustration in Figure 5.2 and Figure 5.3. The highest point in the correlation matrix denotes the highest point of correlation or similarity between the reference image and the input image. If the correlation point is located within the boundaries or on top of the tracked input image, the NCC tracking is considered to be a successful track event and unsuccessful if the correlation point is not within the boundaries of the tracked target or outside of the tracked image. The results of the NCC tracking can be seen in the following table.

	Grayscale	RGB	YCbCr	CbCr	HSV	HS
whole_seq1_1	0.32	0.30	0.04	0.04	0.00	0.13
whole_seq1_2	0.18	0.18	0.05	0.05	0.03	0.08
whole_seq1_3	0.01	0.01	0.00	0.00	0.01	0.03
whole_seq1_4	0.25	0.22	0.05	0.05	0.05	0.12
whole_seq1_5	0.13	0.13	0.09	0.09	0.08	0.19

Table 5.4: NCC Whole Person Tracking in different colour spaces

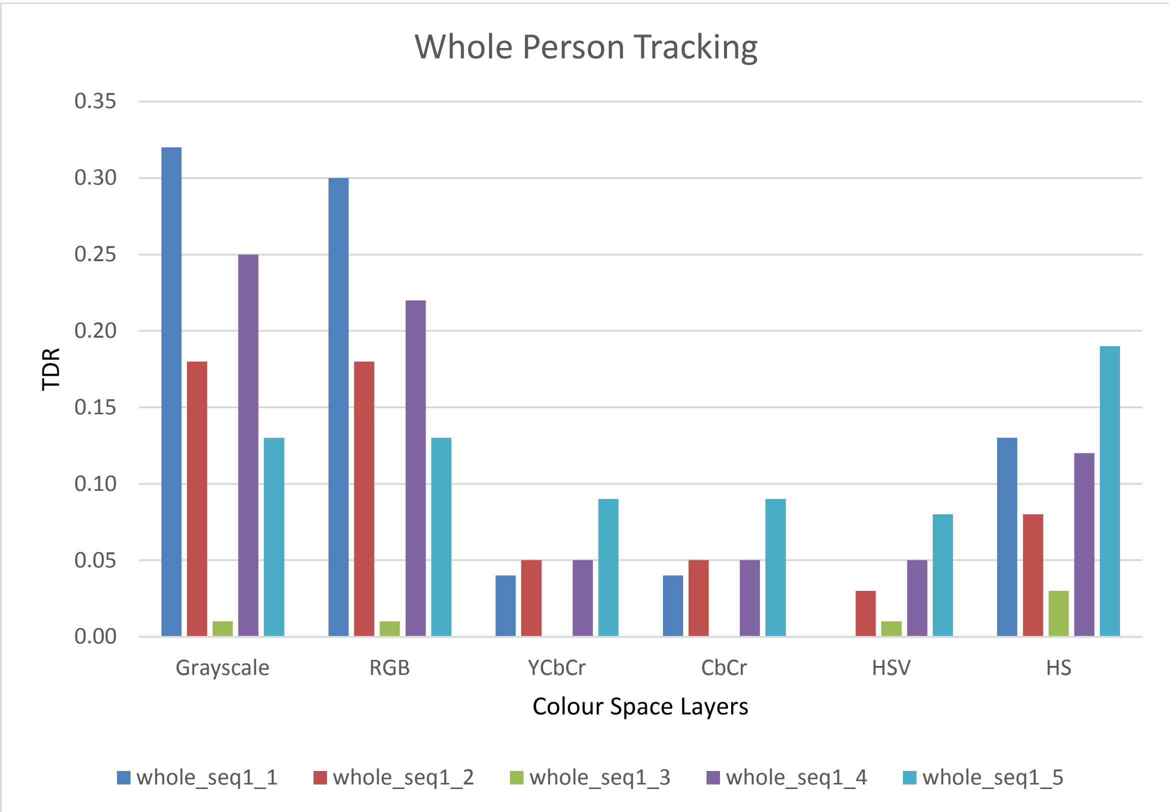


Figure 5.14: Whole person tracking performance in different colour spaces

The results shown in Table 5.4, together with the illustration in Figure 5.14, indicate that the NCC tracking on different colour spaces has different tracking performances. The tracking performance of a whole person also has different tracking performances in different colour spaces. The values shown in Table 5.4 shows the correct tracking rates of using NCC as the tracking method in different colour spaces. In tracking a whole person, the grayscale colour space had higher tracking rates compared to other colour spaces. The tracking in the colour spaces that have the illumination information separated from the colour information, such as YCbCr and HSV, is considerably lower compared to grayscale and RGB colour spaces. This tracking result is contrary to the results obtained from the face tracking using NCC. In tracking the target using NCC, the sample videos used were done in a controlled environment where the background and illumination in the video was uniform. The experiment setup for tracking a whole person tracking is as follows where a standard web camera is used with 320x240 pixel resolution and the background illumination used was based on artificial lighting. The background setup in this scenario had a background with a uniform colour or homogeneous background. All targets were moving at approximately the same speed and target distance from the camera was at the same distance for all different targets.

The whole person tracking was further examined with additional sets of videos where each video was setup differently such as uneven illumination, uneven background and noisy environment. The utilization of different setup or scenarios was to simulate typical or varying environments available in both indoor and outdoor environments in the real world. The premise or setup of the scenario of uneven illumination is that the artificial lighting only has one source from above which does not cover the whole camera view field and thus creates an effect where the illumination changes from one end of the view to the other end of the view. The second scenario is based on the setup where the background scene is non-uniform. The third scenario evaluated is a noisy environment scene that has items in both the foreground and background added with natural changing light in the background. Sample snapshots of the different scenarios can be seen in the following Figure 5.15.

a) Uneven Illumination

b) Uneven background

c) Noisy Environment

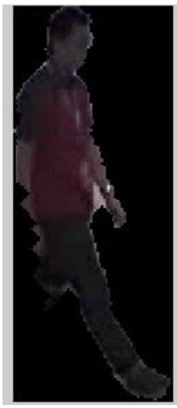
Figure 5.15: Sample Snapshot of Different Environment Setup

The tracking results of NCC in different scenarios of uneven illumination, uneven background and cluttered environment can be seen in respective Tables 5.3, 5.4 and 5.5. In applying the NCC tracking method across the different environment setup, a different set of video sequences were used with different targets to track. The different targets were used in order to track targets that range from multiple colour regions to single colour region targets that can be seen in the following Figure 5.16. Table 5.5 shows the summary settings of input videos for tracking in different background settings

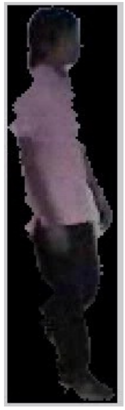
Input video	Uneven Illumination		Uneven Background		Noisy Environment	
	Time (s)	Number of Frames	Time (s)	Number of Frames	Time (s)	Number of Frames
whole_seq2_1	22.26	334	12.33	185	12.00	180
whole_seq2_2	13.33	200	10.66	160	13.53	203
whole_seq2_3	12.00	180	9.66	145	10.73	161
whole_seq2_4	13.66	205	12.33	185	13.06	196
whole_seq2_5	11.00	165	12.00	180	11.00	165
whole_seq2_6	11.66	175	11.66	175	10.00	150
whole_seq2_7	11.33	170	9.33	140	8.86	133
whole_seq2_8	12.00	180	10.00	150	8.40	126
whole_seq2_9	11.33	170	10.33	155	11.60	174

Table 5.5: Summary of second set of Input Video Specifications for whole person tracking

a) whole_seq2_1



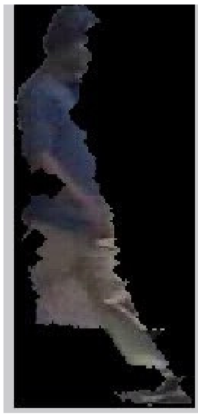
b) whole_seq2_2



c) whole_seq2_3

d) whole_seq2_4

e) whole_seq2_5



f) whole_seq2_6

g) whole_seq2_7

h) whole_seq2_8

i) whole_seq2_9

Figure 5.16: Sample Snapshots of Targets in different scenario setup

Input Video	Colour Space					
	Grayscale	RGB	HSV	HS	YCbCr	CbCr
whole_seq2_1	0.00	0.20	0.00	0.00	0.00	0.00
whole_seq2_2	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq2_3	0.00	0.00	0.92	0.93	0.12	0.00
whole_seq2_4	0.00	0.00	0.00	0.98	0.00	0.00
whole_seq2_5	0.29	0.00	0.00	0.00	0.00	0.00
whole_seq2_6	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_7	0.00	0.00	0.00	0.00	0.00	0.02
whole_seq2_8	0.00	0.40	0.00	0.40	0.00	0.00
whole_seq2_9	0.00	0.00	0.00	0.00	0.00	0.00

Table 5.6: NCC Whole Person tracking in uneven illumination

Input Video	Colour Space					
	Grayscale	RGB	HSV	HS	YCbCr	CbCr
whole_seq2_1	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_2	0.00	0.00	0.00	0.72	0.00	0.00
whole_seq2_3	0.00	0.00	0.36	0.48	0.00	0.00
whole_seq2_4	0.00	0.00	0.00	1.00	0.00	1.00
whole_seq2_5	0.00	0.06	0.00	0.00	0.15	0.00
whole_seq2_6	0.00	0.00	0.00	0.62	0.00	0.74
whole_seq2_7	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_8	1.00	0.40	0.00	0.00	0.00	1.00
whole_seq2_9	0.00	0.00	0.00	0.00	0.00	0.00

Table 5.7: NCC Whole Person tracking in uneven background

Input Video	Colour Space					
	Grayscale	RGB	HSV	HS	YCbCr	CbCr
whole_seq2_1	0.21	0.00	0.00	0.00	0.00	0.00
whole_seq2_2	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_3	0.00	0.00	0.12	0.24	0.69	0.18
whole_seq2_4	0.00	0.00	0.28	1.00	0.00	0.00
whole_seq2_5	0.45	0.00	0.00	0.00	0.00	0.00
whole_seq2_6	0.00	0.00	0.00	0.83	0.00	0.00
whole_seq2_7	0.15	0.07	0.00	0.00	0.00	0.00
whole_seq2_8	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_9	0.00	0.00	0.00	0.01	0.00	0.00

Table 5.8: NCC Whole Person tracking in a noisy environment

With the utilization videos with different setups such as uneven illumination, uneven background and cluttered environment, the results as seen in Tables 5.6, 5.7 and 5.8 did not show any colour space that had better or consistent tracking performance. The results only indicated the NCC tracking performance in different colour spaces and this is illustrated in the following charts of Figure 5.17, 5.18 and 5.19.

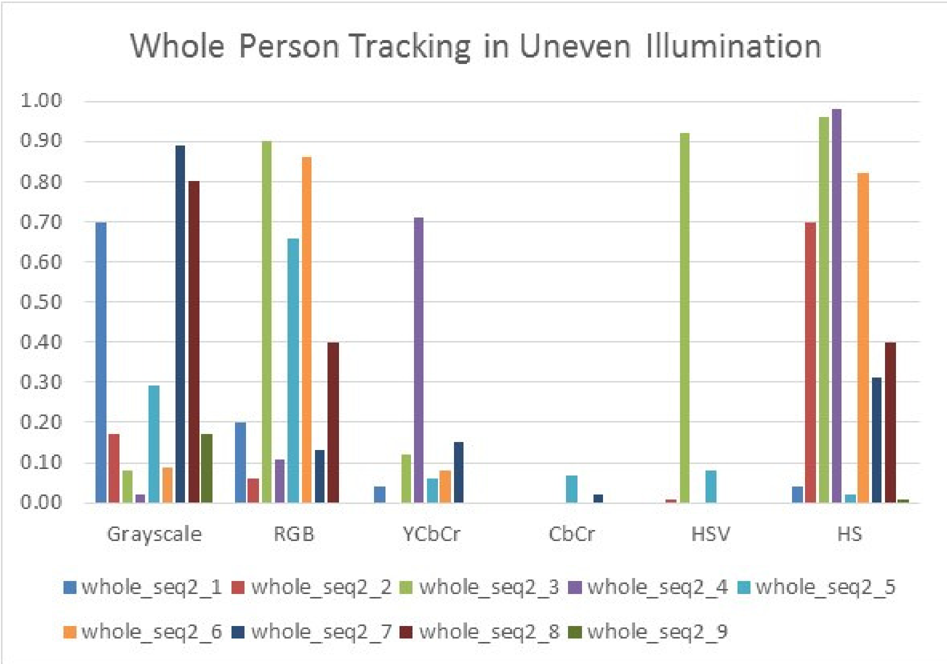


Figure 5.17: NCC Whole Person Tracking in Uneven Illumination

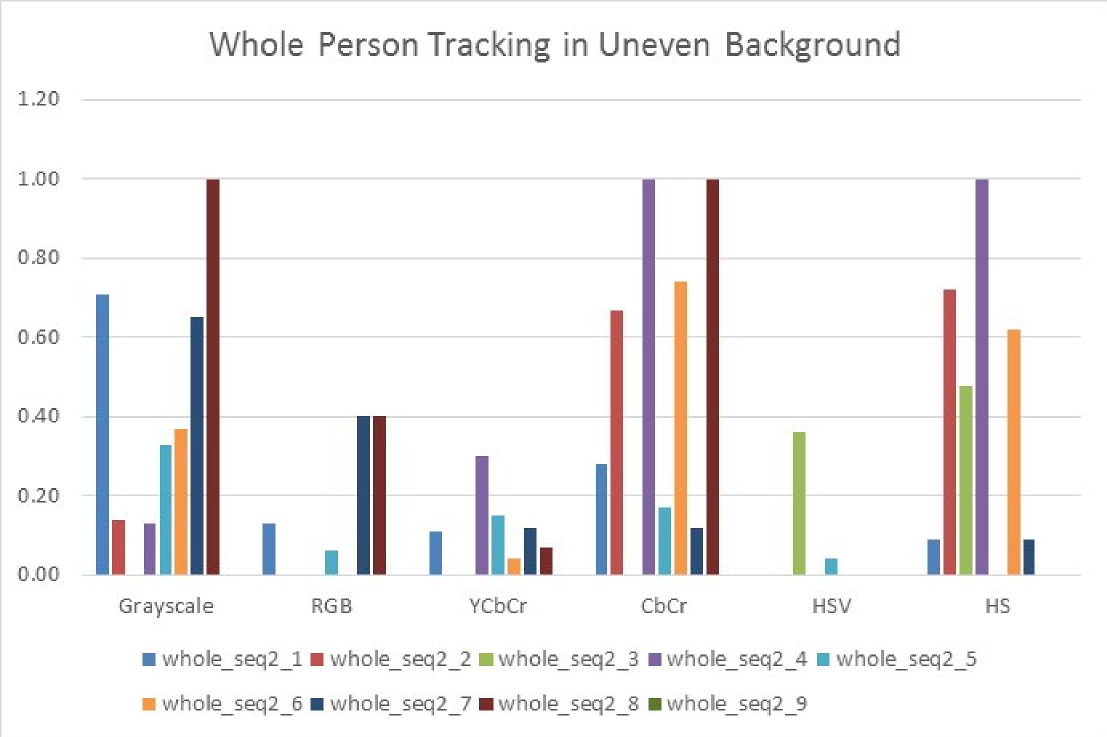


Figure 5.18: NCC Whole Person Tracking in Uneven Background

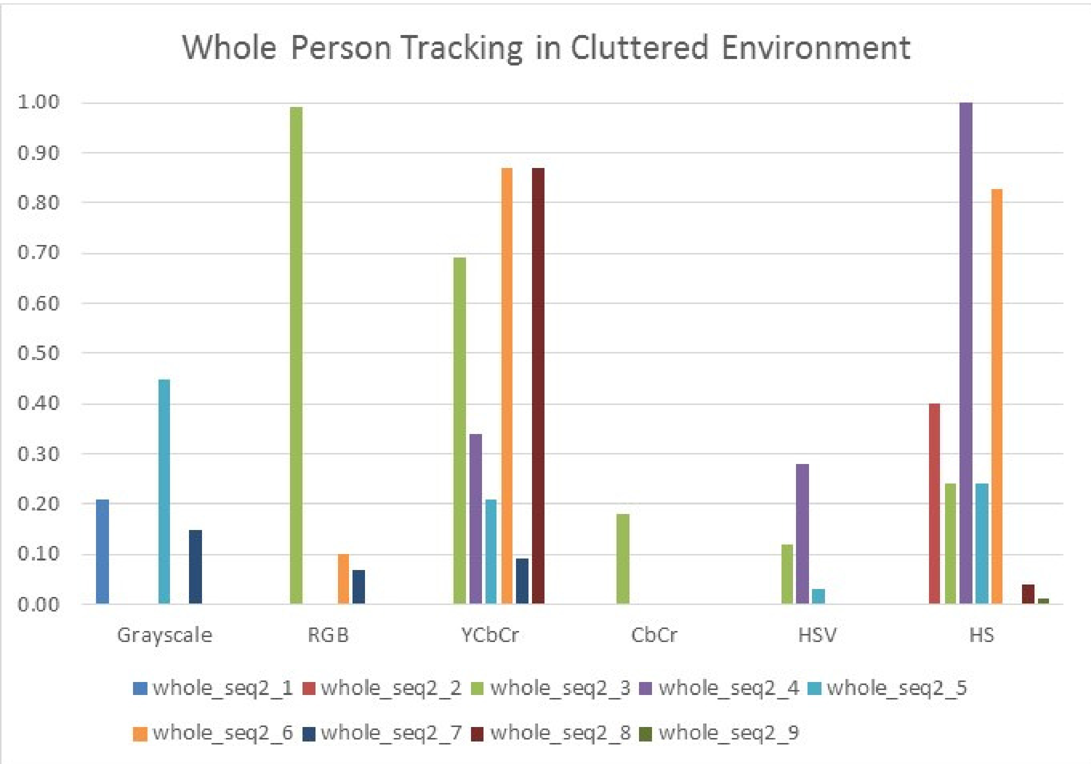


Figure 5.19: NCC Whole Person Tracking in a noisy Environment

The results obtained and illustrated in Figures 5.17 to Figure 5.19 have shown the different tracking levels in different colour spaces. For example, performance tracking levels in grayscale colour space was found to be different or inconsistent in different scenarios. While the tracking performance in grayscale colour space was inconsistent in each scenario, the tracking performance in grayscale colour space contradicted the tracking performance in other colour spaces. It was expected that grayscale colour space would have lower tracking performance compared to other colour spaces. Tracking results in Table 5.3, 5.4 and 5.5, only indicate the tracking performance of each video sequence using its respective image as reference for tracking or detecting the image of interest for tracking. In order to evaluate the feasibility of the NCC method for tracking a person, the tracking results for each video sequence will have to be compared against different reference images captured from the video sequences in each environment setup. The tracking performance can be seen in the following set of confusion matrices. A confusion matrix is a method used due to its discriminative capability and stability in representing data by comparing the corresponding input data to different reference parameters [Huang, Yang et al., 2012], where the table would list the tracking performance for each input video sequence against a reference image or parameter for each respective video sequence. The result of listing the tracking performance of all input videos compared to the reference parameter or image show the level of correct tracking against the level of incorrect tracking.

Input Video	Reference				
	whole _seq_1	whole _seq_2	whole _seq_3	whole _seq_4	whole _seq_5
whole_seq1_1	0.32	0.30	0.00	0.45	0.28
whole_seq1_2	0.20	0.18	0.00	0.23	0.23
whole_seq1_3	0.08	0.11	0.01	0.08	0.11
whole_seq1_4	0.34	0.53	0.00	0.25	0.34
whole_seq1_5	0.19	0.36	0.00	0.08	0.13

Table 5.9: Confusion Matrix for NCC Tracking Performance in Grayscale Colour Space

The matrix shown in Table 5.6 is the additional data of tracking performance comparison from Table 5.2, which uses the first video sequence which has been labelled as 'whole_seq1'. The matrix shown in Table 5.6 shows the tracking performance in grayscale colour space for each input run against the respective reference image. The expected result of the NCC tracking method should show that for the related input video run should have the highest tracking rate for the related reference image used. The tracking rates using the incorrect reference images should have lower tracking rates as compared to tracking using correct reference images. In this instance, the results shown in Table 5.6 indicate that the NCC method was not capable of tracking the correct target. When using reference data or reference image of 'whole_seq1_1', the highest tracking should occur when using 'whole_seq1_1' as the input video. However, in this instance video 'whole_seq1_4' had the highest tracking performance which indicates that NCC tracking could not correctly track its related target. The higher incorrect tracking rate is due to similar colour regions in the videos used which lead to incorrect tracking. The same type of result was also observed in the different colour spaces which indicated that the NCC was not a method that could be used in tracking a whole person. A summary of all confusion matrices for different colour spaces can be seen in Table 5.10. This table displays the tracking performances for related input videos when it has the highest correct tracking rates. When the tracking rates are listed as zero, that indicates that the video did not have the highest tracking rate compared to its reference image.

Input Video	Colour Space					
	Grayscale	RGB	HSV	HS	YCbCr	CbCr
whole_seq1_1	0.00	0.15	0.00	0.13	0.00	0.00
whole_seq1_2	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_3	0.01	0.00	0.01	0.03	0.00	0.00
whole_seq1_4	0.00	0.00	0.05	0.12	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.19	0.00	0.00

Table 5.10: Summary of Confusion Matrix in different Colour Spaces

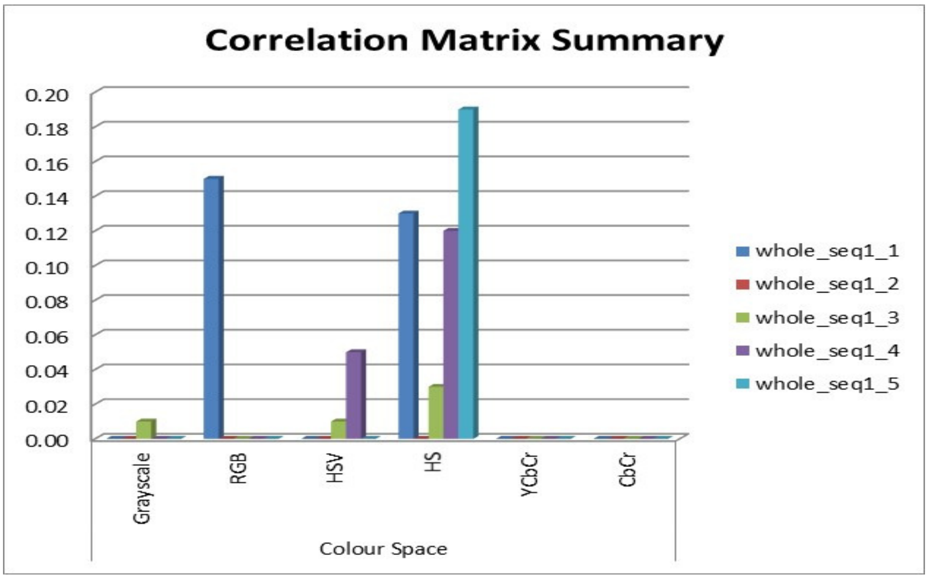


Figure 5.20: Confusion Matrix Summary in Different Colour Spaces

Table 5.10 and Figure 5.20, indicates that the NCC method for tracking a target did not have consistent tracking performance when compared across different colour spaces. The remaining detailed tracking data in different colour spaces can be seen in Appendix 1. In a similar manner, the NCC tracking method was tested further where the environment had three different scenarios such as uneven illumination, uneven background and cluttered environment. The NCC confusion matrix of tracking performances are summarized in the following Tables 5.11, 5.12 and 5.13. A detailed listing of the data for the second NCC tracking can be found in Appendix 2. Data in Tables 5.11, 5.12 and 5.13 are illustrated in the following Figures 5.19, 5.20 and 5.21.

Input Video	Colour Space					
	Grayscale	RGB	HSV	HS	YCbCr	CbCr
whole_seq2_1	0.00	0.20	0.00	0.00	0.00	0.00
whole_seq2_2	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq2_3	0.00	0.00	0.92	0.93	0.12	0.00
whole_seq2_4	0.00	0.00	0.00	0.98	0.00	0.00
whole_seq2_5	0.29	0.00	0.00	0.00	0.00	0.00
whole_seq2_6	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_7	0.00	0.00	0.00	0.00	0.00	0.02
whole_seq2_8	0.00	0.40	0.00	0.40	0.00	0.00
whole_seq2_9	0.00	0.00	0.00	0.00	0.00	0.00

Table 5.11: Summary of NCC Confusion Matrix of Tracking of whole person with uneven illumination

Input Video	Colour Space					
	Grayscale	RGB	HSV	HS	YCbCr	CbCr
whole_seq2_1	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_2	0.00	0.00	0.00	0.72	0.00	0.00
whole_seq2_3	0.00	0.00	0.36	0.48	0.00	0.00
whole_seq2_4	0.00	0.00	0.00	1.00	0.00	1.00
whole_seq2_5	0.00	0.06	0.00	0.00	0.15	0.00
whole_seq2_6	0.00	0.00	0.00	0.62	0.00	0.74
whole_seq2_7	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_8	1.00	0.40	0.00	0.00	0.00	1.00
whole_seq2_9	0.00	0.00	0.00	0.00	0.00	0.00

Table 5.12: Summary of NCC Confusion Matrix of Tracking of whole person with uneven background

Input Video	Colour Space					
	Grayscale	RGB	HSV	HS	YCbCr	CbCr
whole_seq2_1	0.21	0.00	0.00	0.00	0.00	0.00
whole_seq2_2	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_3	0.00	0.00	0.12	0.24	0.69	0.18
whole_seq2_4	0.00	0.00	0.28	1.00	0.00	0.00
whole_seq2_5	0.45	0.00	0.00	0.00	0.00	0.00
whole_seq2_6	0.00	0.00	0.00	0.83	0.00	0.00
whole_seq2_7	0.15	0.07	0.00	0.00	0.00	0.00
whole_seq2_8	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_9	0.00	0.00	0.00	0.01	0.00	0.00

Table 5.13: Summary of NCC Confusion Matrix of Tracking of whole person with noisy environment

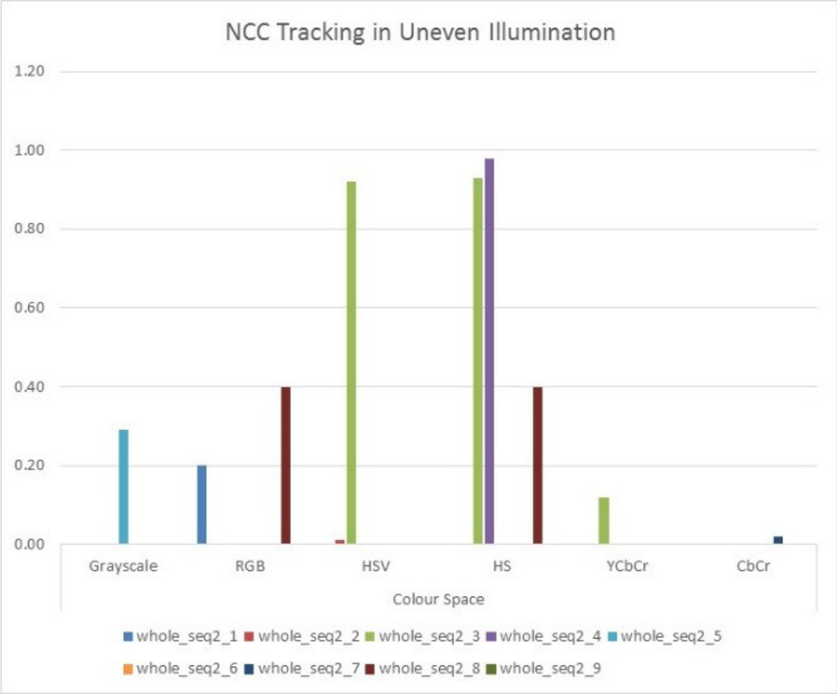


Figure 5.21: NCC tracking in Uneven Illumination

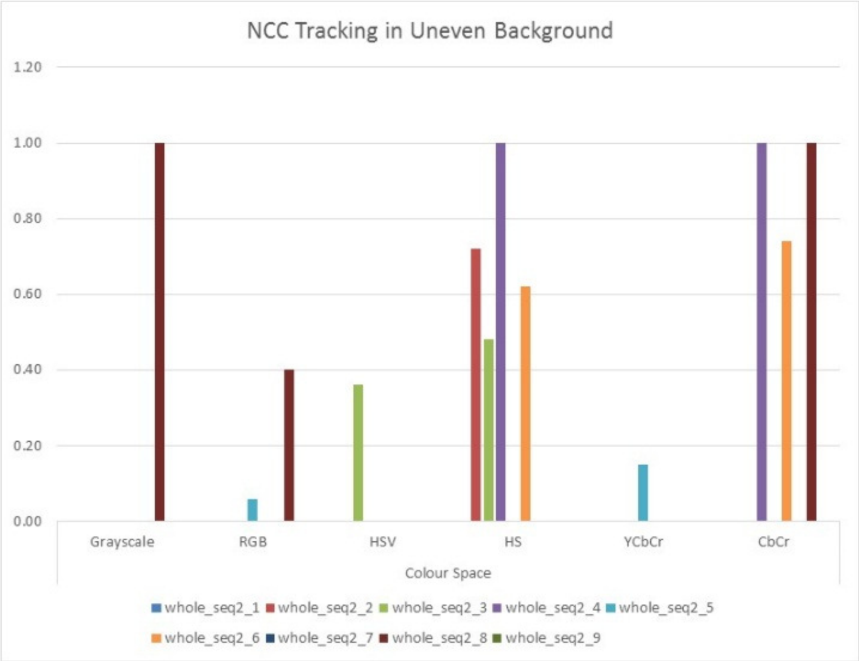


Figure 5.22: NCC tracking in Uneven Background

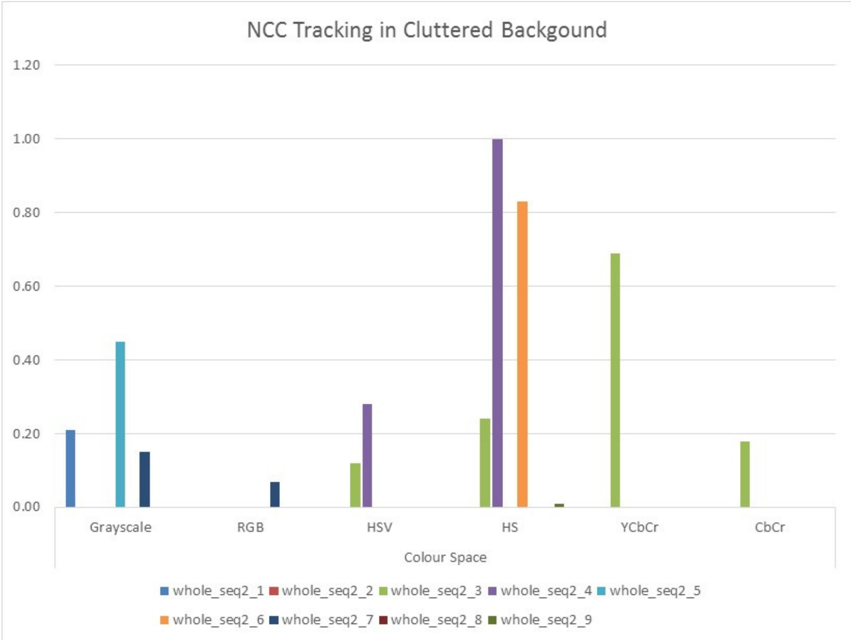


Figure 5.23: NCC tracking in Cluttered Environment

The tracking results illustrated have indicated that the NCC method was not capable of correctly tracking a target with multiple colour regions. This is due to the nature of the NCC method itself which finds the highest point of correlation between the input images against the reference image. In this instance where the whole person was used as the reference image, tracking results did not indicate any distinct performance difference when using either the correct or incorrect reference image. For tracking a whole person using NCC, the highest correlation points could be at different locations for the different colour regions used in representing a person. By looking at the results in Tables 5.8, 5.9 and 5.10, the data indicated that there was no discernible pattern to show that the NCC method used with a whole person reference image or a multiple colour region reference image was able to correctly track a target.

5.5 Discussion

In comparing the results obtained from the utilization of NCC as a method of tracking, the results have shown very different tracking capability for face tracking and whole person tracking. The results for face tracking have indicated from the different videos, that tracking and detecting rates differ in different colour spaces. From the different colour spaces, it indicates that face tracking using colour spaces that have illumination information separated from colour information have higher or better tracking ability. The colour spaces that have the illumination information separated from the colour information are the YCbCr and HSV colour spaces. The grayscale and RGB colour spaces have the illumination information embedded together with the colour information. As such, in the instance of face tracking, the tracking results indicate that YCbCr and HSV colour spaces have better tracking capability compared to grayscale and RGB colour spaces. In addition, the utilization of the colour information alone from the YCbCr and HSV colour spaces would be sufficient for tracking purposes.

In the other instance of NCC tracking, the utilization of the whole person as a reference for tracking a target gives contradicting results compared to tracking and detecting a face. The result from the whole person tracking indicates a different set of results in terms of the colour space to be used for tracking a person using NCC. For whole person tracking, based on the results the colour space with the best tracking result was the grayscale colour space. This result is the opposite of the result obtained from the face tracking and detection. In further tests, the whole person tracking gave further contradictory results that indicated that by including illumination information would further give incorrect tracking. Tracking using NCC would also have the instance where the highest correlation point to be not on the target being tracked but still would be classified as correct tracking due to the position of the point being within the region of interest.

The different tracking results obtained from the face and whole person tracking can be attributed to the information that is being used to track a target. The information used in the NCC tracking is based on the colour information of the reference image used. The higher rate of NCC face tracking and detection compared to the whole tracking is based on the availability of the colour information. On using the face as the reference image for tracking, the colour information is based on skin colour which has a narrow colour range [Beetz, Radig et al., 2006, 39-42; Kawato and Ohya, 2000, 141-1418; Park, Seo et al., 2000, 133-136; Soriano, Martinkauppi et al., 2000, 839-842]. In this instance for face tracking and detection, the colour range in the YCbCr colour space is as follows where the values on the Y layer are between 105 to 117, the values on the Cb layer are between 110 to 113 and values on the Cr layer are above 128. The values on each colour layer used in the detection of skin was used for different skin colour or skin tone. The range of values obtained on each colour layer was sufficient for people of different skin tone, that ranged from light or fair skin to tanned or light brown skin, to be segmented or detected within the camera view. For whole person tracking, the whole person target has multiple regions where each region has a different colour. The tracking and detection rates of the multiple colour regions indicate that the multiple region tracking has a lower tracking rate compared to a single colour region tracking similar to face tracking. Based on the difference of the tracking rate of a single region compared to multiple regions tracking, NCC tracking using a single region gives better tracking capability compared to multiple regions tracking. In spite of the contradicting results obtained in the NCC tracking of a whole person, the utilization of colour information, as seen in the NCC tracking of a face, did indicate that it would be a better method of tracking a target compared to illumination information alone where the results are listed in Table 5.1.

5.7. Summary

Initial tracking data using the NCC method had indicated that it had the ability to be able to track a specific target with the appropriate reference image. Face tracking was proven to have better tracking performance due to the reference image having one major colour region as compared to the whole person tracking which had multiple colour regions. In addition, the utilization of different colour spaces had different tracking performance which was evident in the face tracking evaluations. From the face tracking results, the colour space with the better tracking results were the spaces with the colour information separated from the illumination information. For the whole person tracking, the tracking results were the opposite of the face tracking which is due to the construct of a whole person target that is made up of multiple colour regions. The construct of a whole person with multiple colour regions is the opposite to a face construct which is made up of one primary colour region. Since a whole person is constructed or made up of multiple colour regions, the NCC method of detecting and tracking person would result in multiple correlation points. With the multiple correlation points, there is a higher incidence or possibility of obtaining the incorrect highest correlation point. The results from the whole person tracking indicate that using NCC for whole person tracking, or a multi-region target, indicate that colour information alone as a means to track a target does not give the same level of tracking performance as face tracking with one colour region. Given that the results from the utilization of NCC to determine correct tracking or detection of a target, especially a whole person or multiple colour region target, was inconsistent in the multiple scenarios, an alternative approach would have to be investigated. This approach would have to take into account the different colours available in a target with multiple colour regions. An approach that could take into account the various colours of a target would be a histogram comparison. The Histogram is a comparison method where the colour information of the whole target is used and not different sections or portions of a target for tracking a target. The NCC method used correlation to find the highest correlation point that appears above the tracked image as a correct track indicator. This method works with a single colour target which was tested in face tracking. The same method was tested on a whole person which had multiple colour regions which had then given contradicting and inconsistent tracking results compared to face tracking. Whole person tracking had used the same method of tracking a target but had also given high correlation and tracking points on similar portions of tracked targets such as the lower torso as an indication of a correct track. The results from the whole person tracking indicated that NCC tracking did not give consistent tracking of a target over varying scenario settings. As such the next step would be to utilize the colour information from a multiple region target to track the target which in this case a histogram comparison of colour information would be investigated for tracking a target.

CHAPTER 6

Histogram Tracking

6.1. Introduction

This chapter presents a method for tracking a target based on the histogram or distribution of the pixels of the whole target person of interest. The distribution of the pixels from each target should provide enough information for the purpose of tracking a target as the target moves across the view field of the camera. The pixel distribution of the target being tracked would be used to create a histogram that would be used as a reference. The method of tracking investigated calculates the similarity between histograms. The histograms in question are the histogram of the current image captured and the reference histogram. The technique of histogram comparison as a means of tracking a target can be seen in different applications that range from tracking objects [Wang and Liu, 2006, 10275-10279; Zivkovic and Krose, 2004, 798-803] to classifying shapes [Wahl, Hillenbrand et al., 2003, 474-481] to tracking people in video surveillance [Ali, Indupalli et al., 2006]. The histogram similarity calculations available range from histogram intersection, Euclidian distance, Chi Squared, Kullback-Leibler and Bhattacharyya distance [Wahl, Hillenbrand et al., 2003, 474-481; Zivkovic and Krose, 2004, 798-803]. The calculation of histogram similarity is dependent on the reference or sample histogram. A histogram similarity calculation should be able to provide a numerical value as the basis of how similar the current image or image frame being observed is compared to the reference image. Based on the previous work, histogram tracking was performed to track a specific region which could be a body portion [Wang and Liu, 2006, 10275-10279] or a colour specific region [Zivkovic and Krose, 2004, 798-803]. This method of tracking is different compared to the correlation tracking method mentioned in Chapter 5. In correlation tracking a correct track is dependent on the location of the highest correlation point to be within a specific target region whereas in histogram tracking a correct track is dependent on the similarity of the tracked target's histogram against the reference histogram.

6.2. Whole Person Histogram Tracking

Histogram tracking that was suitable for tracking a whole person as they walk through a camera view field was the primary interest for this research. The histogram similarity methods investigated were the histogram similarity, Euclidian distance, Chi-square test and the Kullback-Leibler divergence for which the equations are listed in order:

$$\cap(H_o, H_{o'}) = \sum_{i=1}^{d} \min\,(H_o(i), H_{o'}(i)) \qquad (6.1)$$

$$\varepsilon(H_o, H_{o'}) = \sum_{i=1}^{d} (H_o(i) - H_{o'}(i))^2 \qquad (6.2)$$

$$\chi_1^2(H_o, H_{o'}) = \sum_{i=1}^{d} \frac{(H_o(i) - H_{o'}(i))^2}{H_o(i)} \qquad (6.3)$$

$$\chi_2^2(H_o, H_{o'}) = \sum_{i=1}^{d} \frac{(H_o(i) - H_{o'}(i))^2}{H_o(i) + H_{o'}(i)} \qquad (6.4)$$

$$\kappa(H_o, H_{o'}) = \sum_{i=1}^{d} (H_o(i) - H_{o'}(i)) \ln \frac{H_{o'}(i)}{H_o(i)} \qquad (6.5)$$

In the similarity comparison equations $H_{o'}$ is the sampled histogram of the target that is being tracked and H_o is the reference histogram. d denotes the number of bins that a histogram is segmented into. As a target moves across the view field of a camera, the histogram of the target will change accordingly. This is due to the changing shape of the moving target which will affect the distribution of pixels. With the changing pixel distribution, the histogram will also change accordingly which will require that a reference histogram be derived from the input image histograms.

6.2.1. Experiment Method

In this instance of histogram tracking, the experiment was set up in a way where a sample or reference histogram is obtained from each video prior to the tracking evaluation being performed. The video capture was done with a standard web camera with a resolution of 320x240 pixels which was the same camera setup as used in Section 5.3.2. In this investigation, the videos used were labelled as ‘whole_seq1_1’ to ‘whole_seq1_5’ respectively which was seen earlier in Chapter 5 for the Normalized Cross Correlation Tracking experiment. Reference histograms for each video sequence are obtained from each video frame to obtain the distribution of the histogram similarity scores. The distribution of the histogram similarity scores is able to provide a set of parameters to determine if the parameters of the current input image is statistically similar or within a 95% confidence interval to the reference histogram. As such, the experiment methodology can be summarized in the following steps:

1. Obtain reference histogram from a sample set of input frames.
2. Calculate the histogram similarity against each frame in the video to determine the distribution of the similarity score.
3. Calculate the parameters from the distribution of similarity scores for determining the correct tracking of target.

   The parameters utilized in this experiment were the confidence interval of the median of the data which is given by the following equation:

   The lower 95% confidence limit is given by:

   $$\frac{n}{2} - \frac{1.96\sqrt{n}}{2} \quad ranked\,value$$

   The upper 95% confidence limit is given by:

   $$\frac{n}{2} + \frac{1.96\sqrt{n}}{2} \quad ranked\,value$$

   where n is the number of values observed. In this instance the confidence interval obtained is the confidence interval of the similarity score from the reference input frames.

4. An input image histogram is obtained and the similarity score is calculated against the reference histogram. The calculated similarity score is then compared to determine if it is within the upper and lower confidence interval. The comparison of the data to the interval values is to determine if the correct target is being tracked. If the correct target is tracked, the similarity score of the tracked target would appear within the confidence interval calculated as stated in Step 3. Whereas an incorrect tracking is a result where the tracked target similarity score would be outside of the reference confidence interval.

The experiments that were done were based on tracking a whole person using the histogram distribution of the pixels. The histogram tracking experiments were performed on different colour spaces. The primary reason for experimenting in different colour spaces was to determine the best colour space and layer to be used for histogram similarity tracking. Investigation into the best colour space to be used for tracking a target of interest was carried out due to the results that were obtained from facial tracking, which had indicated that colour spaces with separate colour information from brightness information have better tracking ability compared to grayscale colour space. The histogram similarity is calculated based on a single layer of any colour space. Three experiments were done based on the same method by which the reference histogram was obtained. The reference histogram is a set of frames where relevant information is extracted from these selected frames. In this section the relevant information extracted would be the pixel distribution or the histogram of the tracked target of interest. The reason for determining a reference histogram is due to the fact that as the target of interest moves across the camera view field the colour histogram on each colour layer changes as the target moves. The changing histogram is due to the changing shape of the target or person that changes the colour regions seen on the camera. The first method of obtaining the reference histogram is to obtain an average histogram from the selected input frames. The average histogram is formed by obtaining the average of each bin from each pixel distribution of the selected input frames that is stated in section 6.3 and 6.4. The second method to obtain a reference histogram from a composite image where all selected input frames are combined or overlaid into a single image to form a composite image and from there the pixel distribution or histogram is obtained. A composite image was selected to enhance or to make the histogram of the common area be prominent or consistent where it could be used for histogram comparison. The third method of histogram tracking was performed on segmented sections of a tracked person. In this instance the segmented sections were fixed according to each tracked target size by a rectangle to designate the regions of interest.

6.3. Average Histogram Tracking Results

Average histogram tracking is done based on utilizing the reference histogram which is obtained from the average of all input reference images. In this instance the average histogram is obtained from the following equation (6.6) [Ferman, Tekalp et al., 2002]:

$$Avg\ Hist_k(j) = \frac{1}{M}\sum_{i=b_k}^{e_k} H_i(j) \qquad (6.6)$$

where H_i denotes the histogram of the i th frame. b_k and e_k mark the start and end of a selected series of frames to be used as reference frames. The histogram similarity can then be calculated using the average histogram as the reference information to be compared against the input image histogram for similarity calculations. The similarity calculation generates a similarity score which can then be compared to a set of parameters that is derived from a distribution of similarity scores. The distribution of similarity scores is obtained from the calculation of the similarity score of the reference average histogram against the respective input videos. The tracking performance is displayed according to the different colour spaces used. The average histogram tracking results can be seen in the following tables for each of the histogram similarity scores. The results in the Table 6.1 indicate detailed tracking detection rates for input videos compared to the different reference histograms.

Reference Histogram	Input Video				
	whole _seq1 _1	whole _seq1 _2	whole _seq1 _3	whole _seq1 _4	whole _seq1 _5
whole_seq 1_1	0.00	0.00	0.00	0.03	0.00
whole_seq 1_2	0.00	0.07	0.00	0.00	0.00
whole_seq 1_3	0.00	0.07	0.00	0.00	0.00
whole_seq 1_4	0.00	0.10	0.00	0.07	0.00
whole_seq 1_5	0.00	0.00	0.00	0.00	0.00

(a) Histogram Intersection

Reference Histogram	Input Video				
	whole _seq1 _1	whole _seq1 _2	whole _seq1 _3	whole _seq1 _4	whole _seq1 _5
whole_seq 1_1	0.00	0.00	0.04	0.00	0.00
whole_seq 1_2	0.00	0.37	0.15	0.13	0.00
whole_seq 1_3	0.00	0.22	0.04	0.03	0.00
whole_seq 1_4	0.00	0.15	0.00	0.07	0.02
whole_seq 1_5	0.00	0.00	0.00	0.10	0.00

(b) Euclidian Intersection

Reference Histogram	Input Video				
	whole _seq1 _1	whole _seq1 _2	whole _seq1 _3	whole _seq1 _4	whole _seq1 _5
whole_seq 1_1	0.00	0.00	0.00	0.00	0.00
whole_seq 1_2	0.00	0.24	0.00	0.00	0.00
whole_seq 1_3	0.00	0.05	0.00	0.00	0.00
whole_seq 1_4	0.00	0.05	0.00	0.08	0.00
whole_seq 1_5	0.00	0.00	0.00	0.00	0.00

(c) Chi Square

ference Histogram	Input Video				
	whole _seq1 _1	whole _seq1 _2	whole _seq1 _3	whole _seq1 _4	whole _seq1 _5
whole_seq 1_1	0.00	0.00	0.00	0.00	0.00
whole_seq 1_2	0.00	0.24	0.00	0.00	0.00
whole_seq 1_3	0.00	0.07	0.00	0.00	0.00
whole_seq 1_4	0.00	0.05	0.00	0.08	0.00
whole_seq 1_5	0.00	0.00	0.00	0.00	0.00

(d) Chi Square 2

Reference Histogram	Input Video				
	whole _seq1 _1	whole _seq1 _2	whole _seq1 _3	whole _seq1 _4	whole _seq1 _5
whole_seq 1_1	0.00	0.00	0.00	0.00	0.00
whole_seq 1_2	0.00	0.34	0.00	0.00	0.00
whole_seq 1_3	0.00	0.05	0.00	0.00	0.00
whole_seq 1_4	0.00	0.07	0.00	0.07	0.00
whole_seq 1_5	0.00	0.00	0.00	0.00	0.00

(e) Kullback-Leibler

Table 6.1: Histogram Tracking Results in Grayscale Colour Space

The results have been compiled into a confusion matrix which is able to list and compare the data between the reference data against the input data. The table is able to show the correct tracking rate against the incorrect tracking rate. Table 6.1 shows the tracking detection rate in the grayscale colour space. The remaining detailed tracking detection rates can be seen in Appendix 3. Tables 6.1 (a) to (e) indicate the tracking performance using different histogram similarity methods ranging from histogram intersection to Kullback-Leibler methods. The highlighted sections in the table indicate the tracking performance of the input video against its own reference histogram. The tracking performances in the highlighted sections are supposed to have higher tracking detection rates as compared to other input videos. For each column in Table 6.1 the tracking rate is recorded for each input video compared to the respective reference histogram. Tracking results expected to be seen are that the highlighted sections to have higher tracking rates compared to the other rates in the same column. Ideally the highlighted sections would have much higher tracking rates compared to the other sections in each column and if other tracking data have higher rates compared to the highlighted section would indicate that the tracking parameters used were giving incorrect tracking results. Tracking in the HSV and YCbCr colour spaces also indicate that tracking using the Kullback-Leibler histogram similarity scores is able to correctly track targets as compared to the other histogram similarity methods. The data in Table 6.1 can be summarized in Table 6.2 which lists the input video tracking against the different histogram similarity methods. Where an input video has a zero rate, it indicates for that particular video that it did not have the highest tracking performance or in other words, it had higher incorrect tracking compared to correct tracking.

Input Video	Histogram Similarity Comparison				
	Histrogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.37	0.24	0.24	0.34
whole_seq1_3	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.07	0.00	0.08	0.08	0.07
whole_seq1_5	0.00	0.00	0.00	0.00	0.00

Table 6.2: Summary of Histogram Tracking Results in Grayscale Colour Space

The tracking performance in Table 6.2 indicates that there is no consistency in the tracking performance in any one of the histogram similarity methods utilized. The data indicates that even using its own average reference histogram to track the target in its own video, the tracking algorithm was not able to track in the grayscale colour space. This indicates that the utilization of a grayscale reference histogram is not capable or viable in tracking a target. In comparison to other colour spaces, the summarized tracking performances can be seen in Tables 6.3 for the HSV colour space and Table 6.4 for the YCbCr colour space.

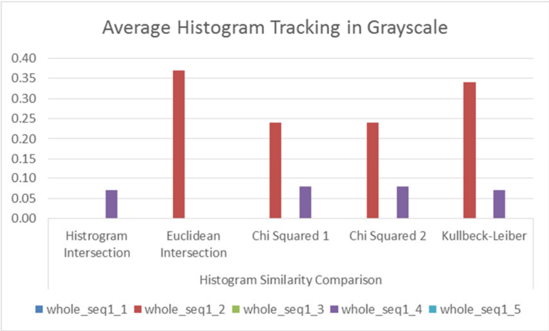
Input Video	Histogram Similarity Comparison				
	Histrogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.29	0.27	0.33	0.23	0.35
whole_seq1_2	0.34	0.27	0.32	0.34	0.34
whole_seq1_3	0.24	0.19	0.23	0.22	0.22
whole_seq1_4	0.33	0.13	0.28	0.00	0.23
whole_seq1_5	0.22	0.00	0.13	0.20	0.11

Table 6.3: Summary of Histogram Tracking Results in H layer of HSV Colour Space

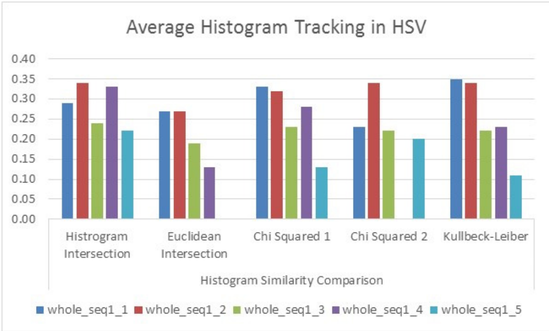
Input Video	Histogram Similarity Comparison				
	Histrogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leilber
whole_seq1_1	0.29	0.19	0.31	0.25	0.27
whole_seq1_2	0.24	0.00	0.29	0.29	0.29
whole_seq1_3	0.20	0.00	0.15	0.16	0.12
whole_seq1_4	0.15	0.20	0.28	0.33	0.35
whole_seq1_5	0.20	0.00	0.19	0.28	0.19

Table 6.4: Summary of Histogram Tracking Results in Cb layer of YCbCr Colour Space

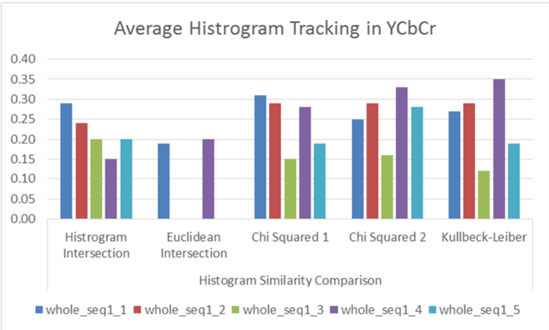
Looking at the tracking performance results in Table 6.2, 6.3 and 6.4, the results indicate that average histogram tracking in the HSV and YCbCr colour spaces was capable of tracking the target of interest compared to the grayscale colour space by having better or higher tracking rates. The histogram tracking done in the YCbCr and HSV spaces was on a single layer of colour information which indicated that tracking could be done using a single layer alone. The tracking performance using the average histogram as a means of tracking on different colour layers of different colour spaces are illustrated in Figure 6.1.



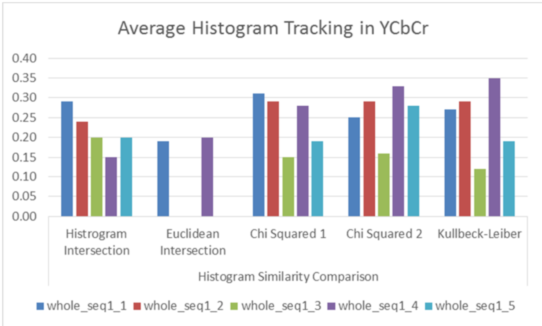
(a) Average Histogram Tracking in Grayscale



(b) Average Histogram Tracking in H layer of HSV



(c) Average Histogram Tracking in Cb layer in YCbCr



(d) Average Histogram Tracking in Cr layer in YCbCr

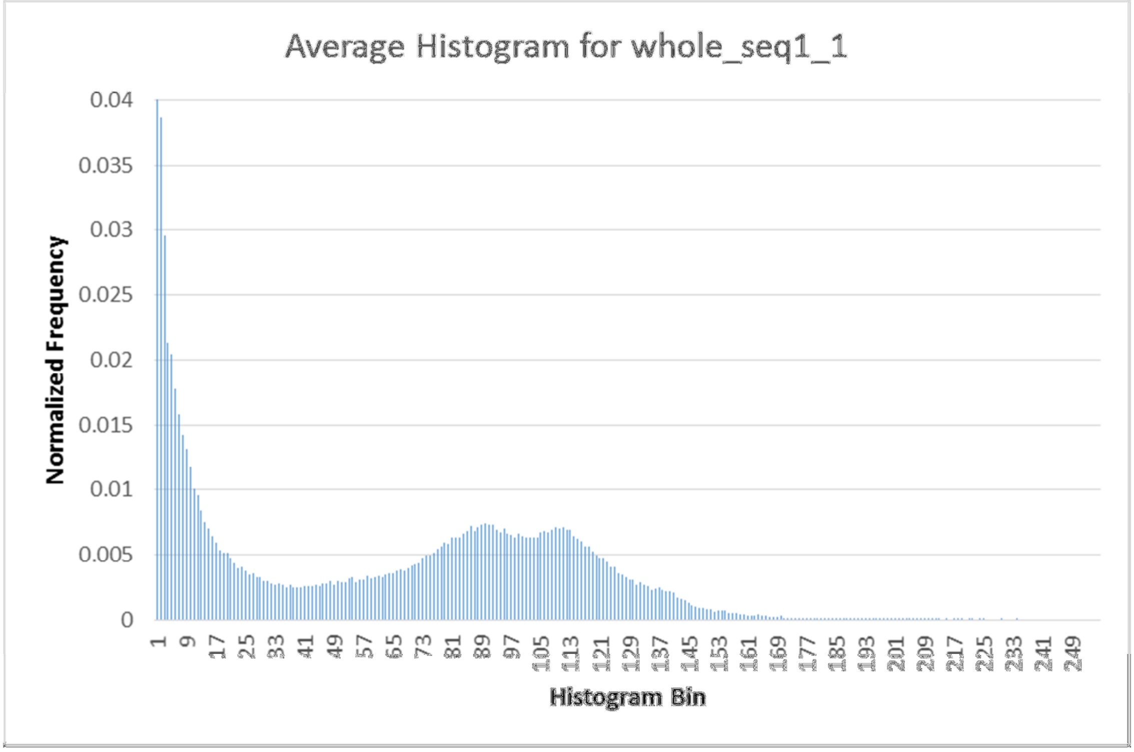
Figure 6.1: Average Histogram Tracking

6.4. Alpha-Trimmed Histogram Tracking Results

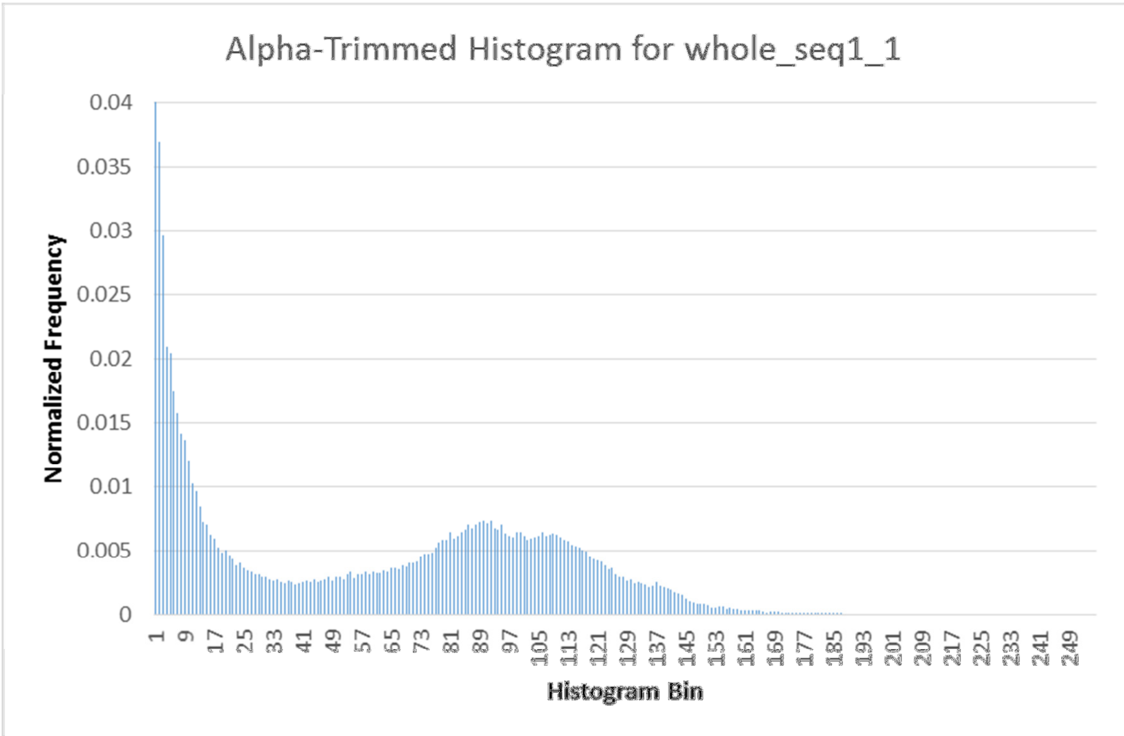
This section of work is an extension of the work done on using average histograms for tracking purposes. The method of using the Alpha-trimmed histogram tracking is based on a method of obtaining the reference histogram that is based on the following equation (6.7) [Ferman, Tekalp et al., 2002]. This method is considered as an extension of histogram tracking where the histogram used is trimmed to remove a number of bins to remove outlier data. The goal of the histogram trimming is to have a more robust histogram by removing the outlier bins and thus utilize the main component or primary colour histogram information for tracking. This step was taken for further evaluation of the colour histogram for tracking for which some initial tracking results in Section 6.3 had shown histogram comparison as a method for tracking targets in the HSV and YCbCr colour space.

$$\alpha Trim\ Hist_k(j,\alpha) = \frac{1}{M - 2 \bullet [\alpha M]} \sum_{m=[\alpha M]+1}^{M-[\alpha M]} h_j(m) \tag{6.7}$$

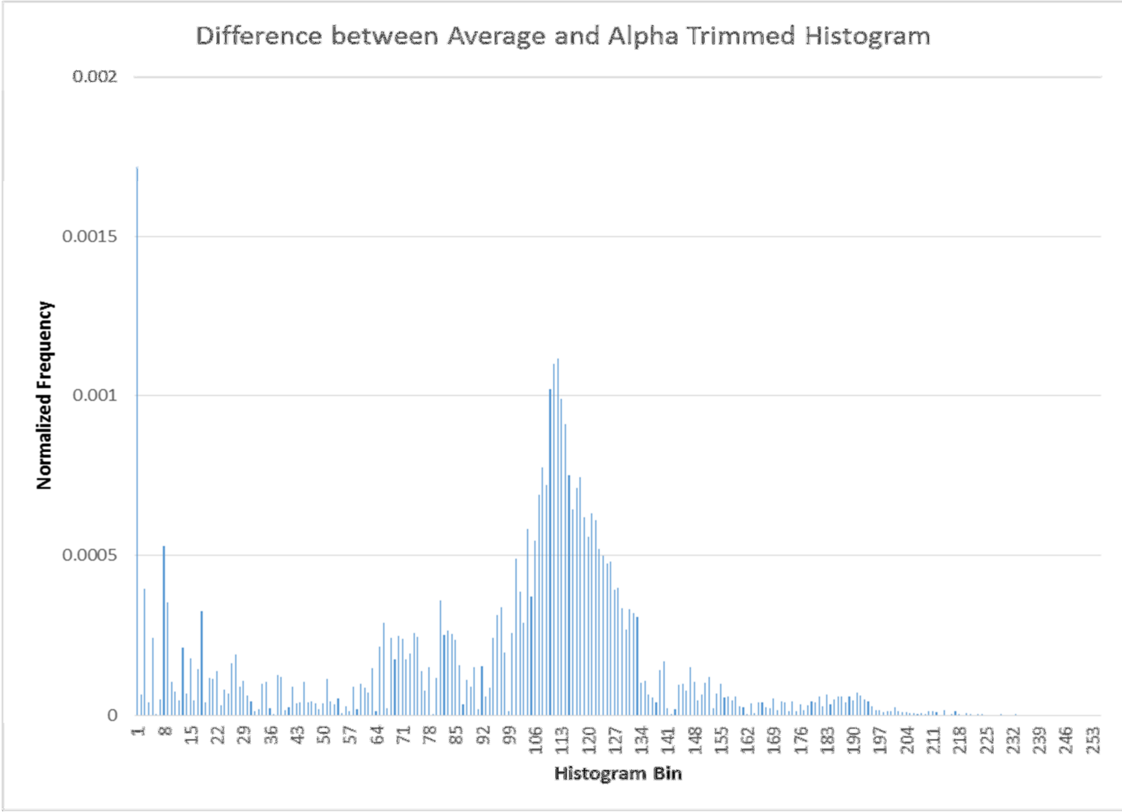
where the *[αM]* denotes the largest integer not greater than αM and h_j is the sorted array of histogram value for the *j*-th bin. The α parameter is known as the trimming parameter which is within the range of 0 to 0.5. This parameter controls the number of parameters to be excluded from the average computation. The trimming parameter reduces the contribution of aberrant or outlier points to the calculated mean by discarding an equal number of samples at each end of the sample distribution. When the trimming parameter, α =0, the resulting histogram is the same as the sampled mean where there is no data being discarded. And when the trimming parameter, α =0.5, the resulting histogram developed is based on the median values obtained after the values are trimmed. The other point to be taken into account is that the histogram obtained either by the average histogram or by the alpha-trimmed histogram is dependent on the sample frames or group of frames (GoFs). The use of the GoFs will reduce or eliminate the unwanted effect of luminance and chrominance variations. In addition to reducing the variation effect on luminance and chrominance, the true colour contents or colour information in the GoFs are enhanced or accentuated. It was stated that unless there was large luminance and chrominance variations observed within the GoFs, the average histogram provides enough colour information in the histogram to correctly represent a specific target. A sample of an alpha-trimmed histogram can be seen in the following Figure 6.2. This figure shows the difference between an average histogram in Figure 6.2(a) as compared to Alpha-trimmed histogram in Figure 6.2(b).



a) Average Histogram



b) Alpha-Trimmed Histogram



c) Difference between average and alpha trimmed Histogram

Figure 6.2: Average Histogram vs Alpha-Trimmed Histogram Comparison

Figure 6.2 (a) and (b) show a sample of the average and alpha-trimmed reference histograms obtained from the 'whole_seq1_1' video. Figure 6.2 (c) shows the difference between the average histogram compared to the alpha trimmed histogram, where the alpha-trimmed histogram has reduced points or frequency for each bin as outlier points are removed from the overall data. The reduction of the normalized frequency for each bin in the alpha-trimmed histogram makes it different compared to the average histogram due to the removal of outlier values for each bin. With the alpha-trimmed histogram as the reference histogram a sample of the tracking performance can be seen in the following table in the grayscale colour space.

Reference Histogram	Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.23	0.21	0.04	0.19	0.06
whole_seq1_2	0.24	0.17	0.27	0.15	0.15
whole_seq1_3	0.06	0.06	0.03	0.05	0.04
whole_seq1_4	0.03	0.03	0.02	0.02	0.01
whole_seq1_5	0.02	0.02	0.02	0.01	0.01

(a) Histogram Intersection

Reference Histogram	Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.13	0.21	0.21	0.02	0.21
whole_seq1_2	0.07	0.32	0.32	0.12	0.32
whole_seq1_3	0.16	0.19	0.15	0.04	0.14
whole_seq1_4	0.03	0.05	0.08	0.10	0.05
whole_seq1_5	0.22	0.28	0.19	0.00	0.17

(b) Euclidian Intersection

Reference Histogram	Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.21	0.21	0.00	0.08	0.71
whole_seq1_2	0.22	0.24	0.07	0.20	0.93
whole_seq1_3	0.18	0.20	0.20	0.32	0.85
whole_seq1_4	0.13	0.12	0.10	0.12	0.73
whole_seq1_5	0.06	0.07	0.07	0.13	0.52

(c) Chi Square

Reference Histogram	Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.04	0.31	0.06	0.19	0.10
whole_seq1_2	0.20	0.32	0.07	0.34	0.22
whole_seq1_3	0.01	0.18	0.16	0.22	0.27
whole_seq1_4	0.00	0.03	0.07	0.02	0.20
whole_seq1_5	0.07	0.11	0.17	0.20	0.30

(d) Chi Square 2

Reference Histogram	Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.15	0.19	0.00	0.13	0.04
whole_seq1_2	0.22	0.24	0.12	0.22	0.22
whole_seq1_3	0.18	0.18	0.31	0.15	0.12
whole_seq1_4	0.12	0.15	0.10	0.12	0.12
whole_seq1_5	0.04	0.07	0.04	0.17	0.24

(e) Kullback-Leibler

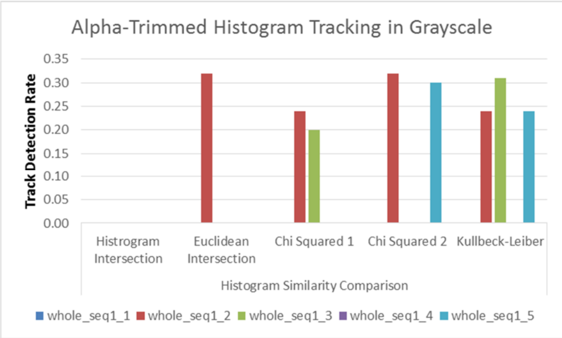
Table 6.5: Alpha-Trimmed Histogram Tracking Results using Histogram intersection in Grayscale Colour Space

The results from Table 6.5 show that the use of the alpha-trimmed histogram did not provide better tracking performance compared to average histogram tracking performance which is listed in Table 6.1. A summary of the tracking data in Table 6.5 is seen in Table 6.6 which summarizes the confusion matrices into a summary table for the grayscale colour space against each histogram similarity method. Data in each column is a reflection on the ability of the histogram similarity method to correctly track a target and a zero result indicates that it was not able to correctly track the target of interest. Figure 6.3 illustrates the summarized tracking performances across all colour spaces from data listed in Appendix 3.

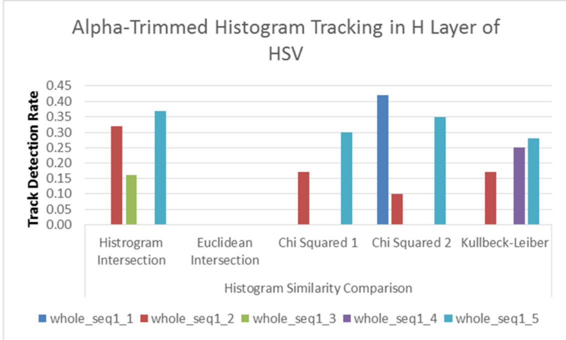
Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.32	0.24	0.32	0.24
whole_seq1_3	0.00	0.00	0.20	0.00	0.31
whole_seq1_4	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.30	0.24

Table 6.6: Summary of Alpha-Trimmed Histogram Tracking Results in Grayscale Colour Space

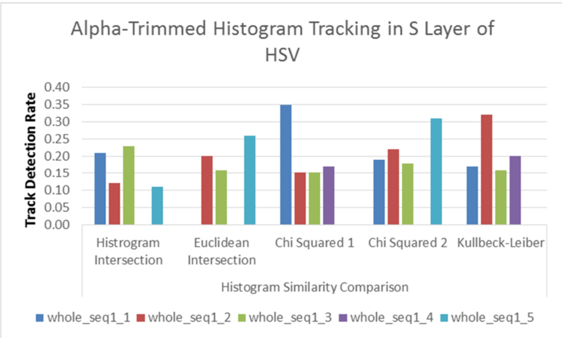
The remaining results and summary of results in other colour spaces can be seen in Appendix 4. The tracking results in the other colour spaces using the alpha-trimmed histogram have shown that the alpha-trimmed histogram used as reference does not provide better tracking performance. In fact, the tracking performance is worse and has higher incorrect tracking rates. It was initially thought that the removal of 'outlier' values would give better tracking performance. Based on the tracking performance, the utilization of 'outlier' values in generating the reference histogram is still essential. The 'outlier' values are part of the colour information of the image. The term 'outlier' here indicates the number of values that are excluded in the determination of the alpha-trimmed histogram which is determined by Equation 6.7. The colour information would change as the target moves across the camera view field where the shape of the target changes such as the swinging of arms and the leg movements would make the pixel colour distribution change. Based on tracking results the alpha-trimmed process of removing 'outlier' or data points for each histogram bin reduces the effectiveness of the alpha-trimmed histogram as a reference histogram for tracking.



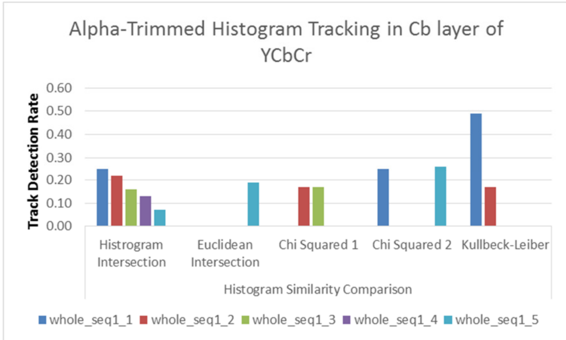
a) Alpha-Trimmed Histogram Tracking in Grayscale



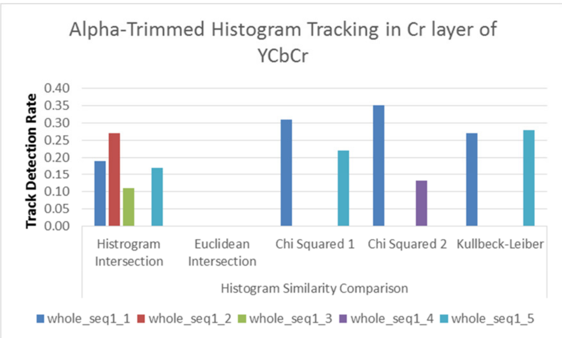
b) Alpha-Trimmed Histogram Tracking in H Layer



c) Alpha-Trimmed Histogram Tracking in S Layer



d) Alpha-Trimmed Histogram Tracking in Cb Layer



e) Alpha-Trimmed Histogram Tracking in Cr Layer

Figure 6.3: Alpha-Trimmed Histogram Tracking

Looking at the illustration of the results across different colour spaces, there is no consistent tracking performance across them. The results from the different histogram comparison methods also did not show any consistent tracking performance. This lack of consistent tracking performance across the different colour spaces and histogram comparison methods are dependent on the reference histogram which in this instance the method to obtain the reference is based on alpha-trimmed histogram that had been tailored for histogram intersection comparison. From the results obtained, the utilization of an alpha-trimmed histogram as a method of tracking was found to be lacking in consistency in tracking a target across different colour spaces as there was no one colour space layer that had consistent tracking performance.

6.5. Composite Image Histogram Tracking Results

A composite image is created to obtain the histogram that would be used as a reference in the calculation of histogram similarity. The image used is a composite or combination of all the input image frames which should, hypothetically, be able to provide an appropriate reference histogram because the image is able to encompass the overall histogram possibilities as the target of interest moves. Composite images are formed by the overlaying of 50 to 100 input frames where all images are aligned and the colour information are averaged. The following images give the different composite images and the related histogram. The primary hypothesis was that the composite image would generate a histogram where the main or common area would be generated or used as the main area for histogram comparison.

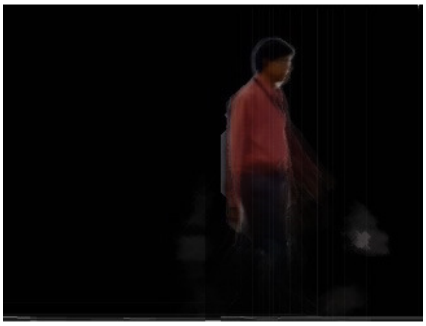
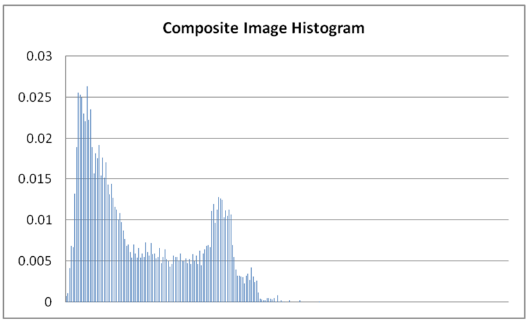



Figure 6.4: Composite of whole_seq1_1

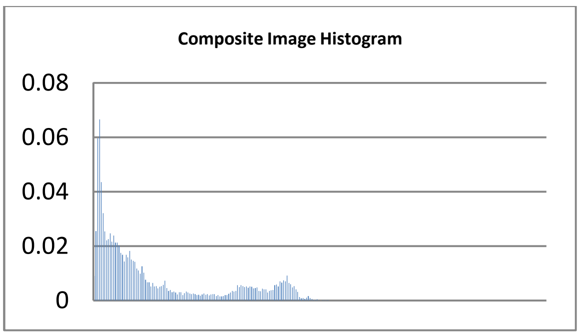



Figure 6.5: Composite of whole_seq1_2

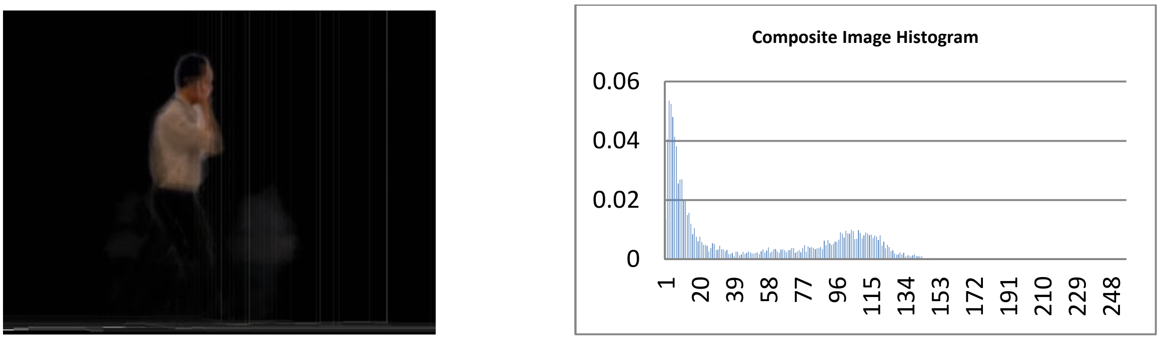


Figure 6.6: Composite of whole_seq1_3

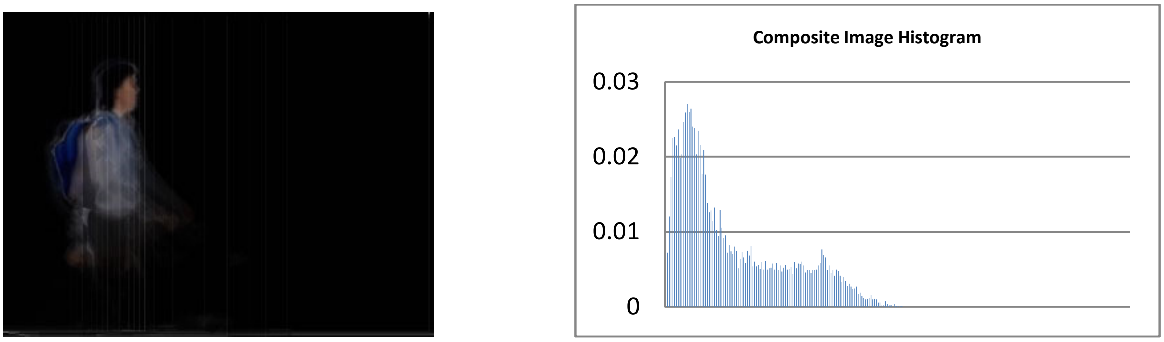


Figure 6.7: Composite of whole_seq1_4

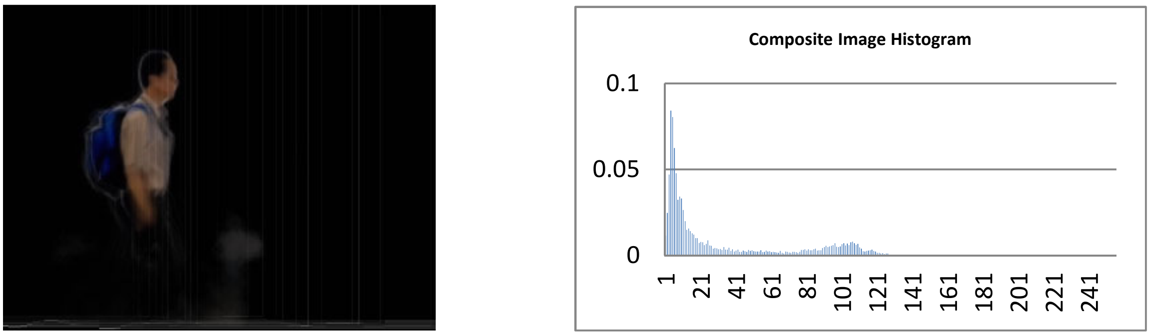
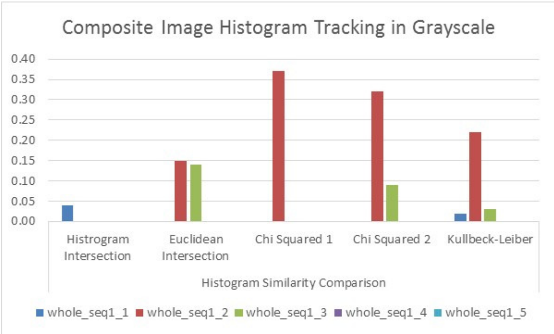
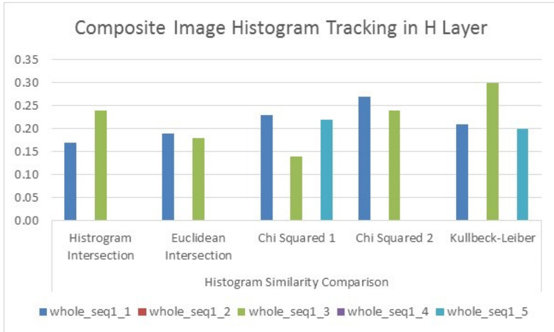


Figure 6.8: Composite of whole_seq1_5

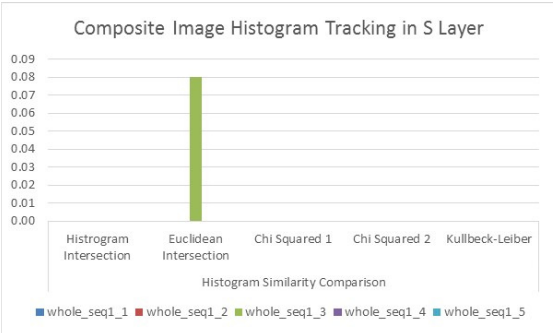
From the composite images and related reference histogram, the tracking performance can be seen in the following figure that summarizes and illustrates the tracking performances across the different colour spaces and histogram similarity methods. The detailed data collected can also be seen in Appendix 5.



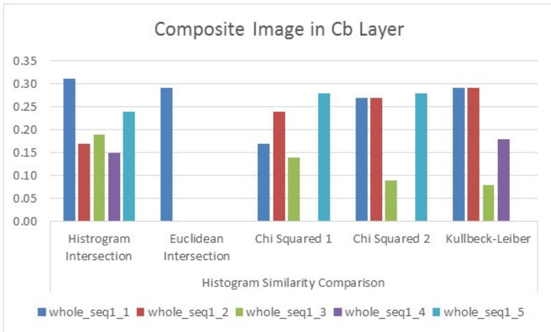
(a) Histogram Intersection



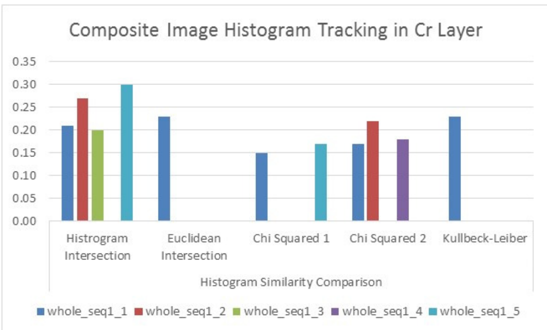
(b) Euclidian Intersection



(c) Chi Square



(d) Chi Square 2



(e) Kullback-Leibler

Figure 6.9: Composite Image Histogram Tracking Results

The tracking data illustrated in Figure 6.9 can also be seen in detailed in Appendix 5 that shows the detailed data and summarized data. The results as listed indicate that the tracking performance using the composite image histogram did not display significantly higher tracking performance compared to average histogram tracking and alpha-trimmed histogram tracking methods. This would indicate that the composite image histogram tracking method did not have the desired tracking performance. Tracking performance was expected to be better with the image being overlaid and making the common areas prominent. Tracking results from the overlaid image and the related histogram used as reference indicated that tracking had not improved as compared to using an average histogram. The utilization of a composite image histogram as reference did not take into account that the input image histogram has a different histogram with the input image having a constantly changing histogram that is affected by the movement of limbs as the target moves across the camera view field. The composite image histogram had emphasized the common areas and had minimized the moving limbs. The difference in the emphasis of the reference histogram as compared to the input image histogram is the cause for the lower tracking performance compared to the utilization average histogram as reference histogram.

6.6. Segmented Image Histogram Tracking Results

Segmented image histogram tracking is a method where the tracked target image is segmented into different sections such as head, torso and lower limbs. The histogram of each section is then obtained for histogram comparison score calculations. A correct track or identification of a selected target is based on the correct track of any one of the portions that have similar histograms which mean that if any portion where the histograms match the target being observed is considered a correct track. The hypothesis that is being tested here is that tracking of targets using the segmented sections of the target should be able to get higher tracking capability compared to the tracking a whole target. A sample of tracking performance of the segmented tracking method is as follows:

Reference Histogram	Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.65	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.71	0.32	0.57	0.07
whole_seq1_3	0.02	0.07	0.57	0.03	0.72
whole_seq1_4	0.00	0.56	0.05	0.57	0.65
whole_seq1_5	0.00	0.17	0.15	0.20	0.63

(b) Histogram Intersection

Reference Histogram	Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.69	0.00	0.00	0.00	0.00
whole_seq1_2	0.06	0.73	0.24	0.30	0.11
whole_seq1_3	0.00	0.22	0.53	0.18	0.20
whole_seq1_4	0.17	0.37	0.10	0.62	0.35
whole_seq1_5	0.00	0.29	0.49	0.37	0.56

(b) Euclidian Intersection

Reference Histogram	Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.73	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.66	0.09	0.25	0.02
whole_seq1_3	0.00	0.10	0.49	0.00	0.39
whole_seq1_4	0.00	0.20	0.00	0.53	0.37
whole_seq1_5	0.00	0.00	0.27	0.07	0.59

(c) Chi Square

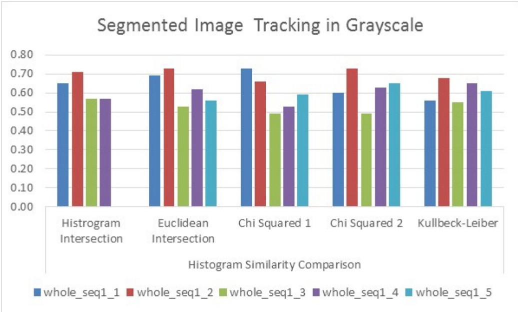
Reference Histogram	Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.60	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.73	0.23	0.42	0.00
whole_seq1_3	0.00	0.02	0.49	0.00	0.13
whole_seq1_4	0.00	0.27	0.00	0.63	0.52
whole_seq1_5	0.00	0.02	0.20	0.10	0.65

(d) Chi Square 2

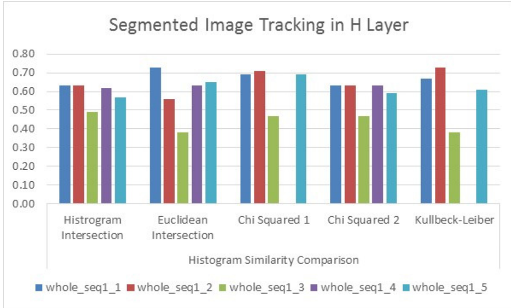
Reference Histogram	Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.56	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.68	0.15	0.42	0.00
whole_seq1_3	0.00	0.05	0.55	0.00	0.13
whole_seq1_4	0.00	0.24	0.00	0.65	0.35
whole_seq1_5	0.00	0.00	0.31	0.03	0.61

(e) Kullback-Leibler

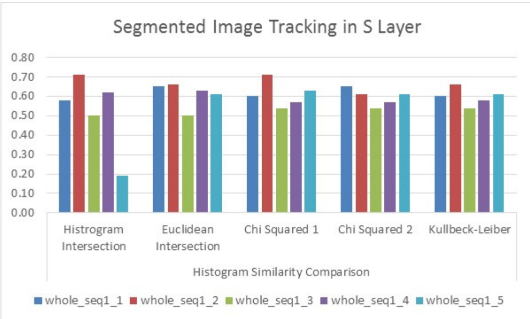
Table 6.7: Segmented Image Histogram Tracking Results in Grayscale Colour Space



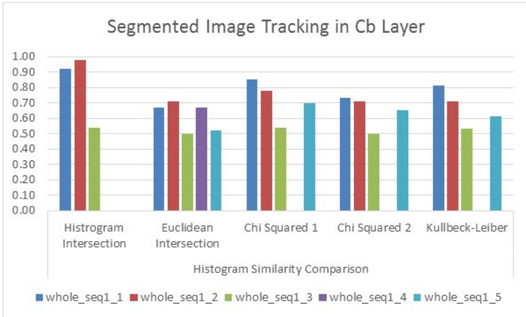
(a) Histogram Intersection



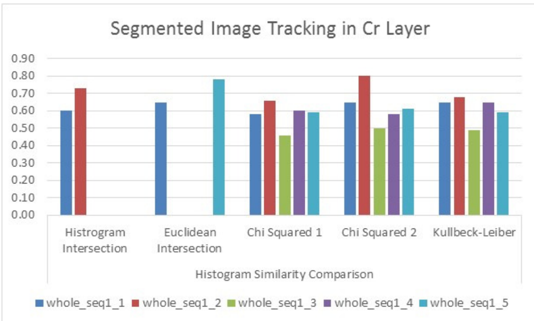
(b) Euclidian Intersection



(c) Chi Square



(d) Chi Square 2



(e) Kullback-Leibler

Figure 6.10: Segmented Image Histogram Tracking Results

The segmentation of the target of interest was done based on the percentage or manual measurement of the target portions. The target is segmented or divided into three areas of interest which are labelled as head, torso and lower limbs. In the video sequences used, namely the sequences labelled as 'whole_seq1', the percentages of regions of interest (ROI) are at 15 to 20 % for head, 30 to 35% for torso and 45 to 50 % for the lower limbs. The RIOs have rectangles to denote areas where the histogram has been compared against reference histograms. Based on the results obtained in the previous sections of utilizing different reference histograms, the reference histogram for the different regions are obtained using the average histogram method. The results in Table 6.7 and illustration in Figure 6.10 showed on average 10 % and higher correct tracking rates for segmented image histogram tracking as compared whole image histogram tracking. Segmented image histogram tracking allowed for higher correct tracking rates but also produced higher incorrect tracking rates which are indicated as zero tracking rate in the summarized data in Appendix 6. The higher incorrect tracking rates are due to the tracking of similar sections on different videos such as the head and the lower limbs. The remaining detailed tracking results in other colour spaces can be seen in Appendix 6. The summary of tracking performances for segmented tracking are also listed in Appendix 6.

6.7. Summary

In looking at the different methods of histogram similarity tracking, ranging from average histogram, alpha-trimmed histogram, composite image histogram and segmented image histogram as the reference histogram, there was no definitive colour space that had consistently higher tracking performance. By looking at the tracking performance of histogram comparison across the different colour spaces, it was observed that the HSV and YCbCr colour spaces had better correct tracking performance compared to grayscale colour space. Depending on the reference histogram used and the histogram similarity method employed the tracking performance varies according to the different colour spaces. Looking at the data listed in Appendix 3 which uses an average histogram as the reference histogram and the different histogram similarity measures, where the results indicate that the correct tracking rates were higher in the YCbCr colour space compared to the HSV and grayscale colour spaces. In comparison with the data in Appendix 4 which uses an alpha-trimmed histogram as the reference histogram, the tracking performance had more incorrect tracking rates as compared to tracking using the average histogram. The results listed in Appendix 3 and 4 are based on the histogram of the whole person as the target to be tracked. The results obtained indicate that the histogram similarity had a maximum of 40% tracking capability. Based on the relatively low rate of correct tracking using either an average histogram or an alpha-trimmed histogram as the reference histogram, another method was investigated where a composite image is created out of a group of reference frames or images to obtain the reference histogram. The full tracking results from using the composite image as the source of the reference histogram can be seen in Appendix 5. The results indicated that the tracking rates that were obtained were not significantly higher compared to using the average or alpha-trimmed histogram. The results from the composite image histogram tracking showed that the YCbCr colour space had better tracking rates compared to the HSV and grayscale colour spaces. This indicates that the tracking in the YCbCr colour space should be utilized in tracking targets as compared to the other colour spaces. Based on the relatively low correct tracking rates obtained from the whole person histogram tracking method, another approach was investigated where the whole person is segmented into different regions such as the head torso and limbs, where the histogram of each of the segmented regions are used as a means of tracking the correct target. Based on that idea of tracking a target using the histogram of the segmented portions of the target, the results can be seen in Appendix 6. The results show that the correct tracking rates were significantly higher compared to the other histogram similarity tracking methods. The tracking results also indicate that the incorrect tracking rates also increased due to the different sections being used in tracking a target. Based on the results, the colour space with better tracking performance was the YCbCr colour space.

Looking at the results obtained in Appendix 3 to 6, the results showed that the histogram similarity method had some capability in tracking a target. The low tracking rates in the whole person tracking was due to the moving target which would have its histogram change as it moved across the camera view field. As the target moves and the histogram changes, the histogram that has the closest similarity score to the reference histogram and deemed to be a correct track would appear as the target moves across the camera view field. The tracked target histogram will change as the target shape changes as it moves across the camera view field. Due to the results and findings of the changing histogram, it was decided to investigate the possibility of segmenting the whole image into different areas to further reduce the overall changes that occur to the histogram as it moves across the camera view field. The whole person image is segmented into three segments which were the head, torso and limbs. The result from the histogram tracking of the segmented regions showed that higher correct tracking rates were observed but at the same time higher incorrect tracking rates were also observed.

Based on the overall results of using histogram comparison as a method to track a target, an important step is the determination of the reference histogram. From the different methods of obtaining the reference histogram, the tracking rates of a whole person can be seen to be in the range of 20 to 30%. This shows that the input histogram is similar to the reference histogram as the target moves across the camera view field together with the changes in its histogram. With the histogram of the target of interest changing due to movement, the utilization of histogram comparison as a tracking method would be inaccurate. From observations of the histogram as the target of interest moves across the camera view field, the general shape of the histogram is the same which indicates that the colour distribution can be represented as a collection or group of normally distributed colour pixels with different mean values. The colour pixel distribution is the next step to be investigated where the target of interest could be segmented into different colour regions and the related information between colour regions could be used for tracking.

CHAPTER 7

Parametric Tracking of Multiple Segmented Regions

7.1. Introduction

This chapter describes a tracking method based on parameters but not based on the histogram characteristics of the target of interest. The target of interest in this instance is a person or persons of interest that move through a camera view field. The tracking method investigated here is the utilization of parameters extracted from different sections of the image. The sections are derived from segmenting the target of interest into different colour regions where a single target of interest would then be represented by a group of colour regions. The colour regions are areas of clothing and the general build of the person that is being tracked. From the collection or grouping of blobs that make up a single person, the connection or the information between the blobs generates the parameters that can be used for tracking purposes. The parameters of interest investigated in a multiple blob or regions were the parameters of vector between regions and the mean values of the regions. The vector in this instance has the property of distance and direction between the two regions of interest.

7.2. Colour Segmentation

Colour segmentation is a step which will separate a single image into different areas based on the colour properties. The YCbCr colour space was used in the colour segmentation process, which has shown better tracking performance compared to RGB and grayscale colour spaces [Kim, Cho et al., 2004, 785-788; Sebastian and Vooi Voon, 2007; Sebastian, Vooi Voon et al., 2008, 2512-2516; Sebastian, Vooi Voon et al., 2010, 298-312]. The YCbCr colour space is typically used in digital image and video transmission and storage where the luminance component and chrominance components can be sub-sampled and compressed for efficiency purposes. The colour information used for segmentation is located in the chrominance layers of Cb and Cr.

The segmentation to be done is based on the distribution of the pixels where each peak in the pixel distribution designates a colour region. The colour regions are then segmented using the Expectation Maximization (EM) method [Shao, Gao et al., 2014, 1-6; Tatiraju and Mehta, 2008; Wang, Sang et al., 2014, 4098 - 4103; Zhu, 2009, 1-5; Zivkovic and Krose, 2004, 798-803]. The EM algorithm is a data clustering algorithm that clusters data into groups of similar objects [Abbas, 2008; Shao, Gao et al., 2014, 1-6; Wang, Sang et al., 2014, 4098 - 4103]. The need for clustering can be seen in the pixel distribution of the target being tracked where a sample pixel distribution can be seen in Figure 7.2 and Figure 7.3 respectively. Based on the overall pixel distribution, multiple peaks are observed in the pixel distribution that indicate distinct colour regions. Apart from segmenting colour regions, EM algorithms can be implemented using segmented wavelets of an image [Dutta and Sarma, 2014, 466-770]. The EM algorithm is also implemented in segmenting features of interest in medical applications such as lung nodules [Yiming and Guirong, 2014, 20 - 23]. Different methods are available for segmenting any image into clusters or groups using methods such as k-means clustering and EM algorithm [Abbas, 2008; Adebisi, Olusayo et al., 2012, 62-71; Alldrin, Smith et al., 2003; Tatiraju and Mehta, 2008; Yunji, Bin et al., 2014, 4831 - 4835; Zhu, 2009, 1-5]. The clustering methods employed by k-mean, also known as Lloyd’s algorithm, clusters data according to the distance of the pixel from a central point or centroid. The pixels are clustered by determining the minimized distance or error around each centroid. The EM algorithm models the cluster around a probability density function (PDF) where the pixels are determined by the probability of the data point given the PDF of each cluster. This is where the k-mean methodology differs from EM in that each object is assigned to one cluster alone and the model or parameter derived is only the centroid. This differs from the EM method which is a probabilistic clustering method that deals with the uncertainty of cluster assignment for each object or data point and the parameters that can be generated are means and variance to model data distribution. The EM algorithm used in this research is the Gaussian Mixture Models (GMMs) where the overall distribution is modelled by multivariate or multiple Gaussian distributions which can be seen in Figures 7.2 and 7.3 for the overall pixel distribution and Figure 7.4 and 7.5 for the GMMs that were obtained. The selection of EM over k-means as the method to cluster data is due to the probabilistic method of EM which attempts to classify data points according to the probability of a point being centred around a central value. This is in comparison to k-means that determines a data point to be a member of a cluster based on a fixed distance from the central value. Based on the GMM, any overall pixel distribution is constructed by a set of individual Gaussian distributions where each distribution would have its own mean and standard deviation. Based on those observations, GMM provides an approach that would determine the best set of parameters of normal distributions to model the overall colour distribution. Segmentation based on colour properties allows the image to be separated into different colour regions which can denote the head, face, torso, lower limbs, skin and hands of the person being tracked [Dargazany and Nicolescu, 2012, 646-651; Ju, Wang et al., 2014, 1736 - 1742; Park and Aggarwal, 2002, 105-111; Zhu and Zhu, 2014, 101 - 104]. An example of the segmentation can be seen in the following figures. Figure 7.1 shows the input image which is then followed by Figure 7.2 and Figure 7.3 which are the pixel distributions on the Cb and Cr layers for the tracked target.

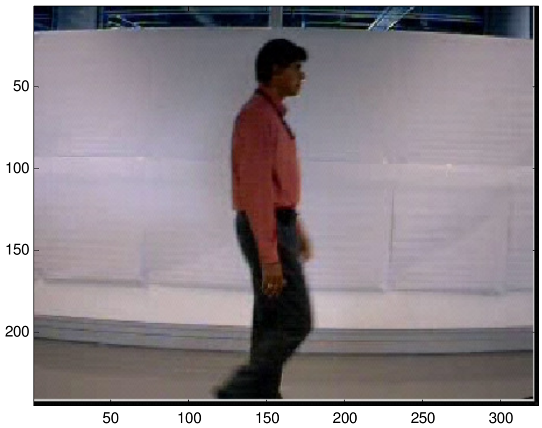


Figure 7.1: Input Image

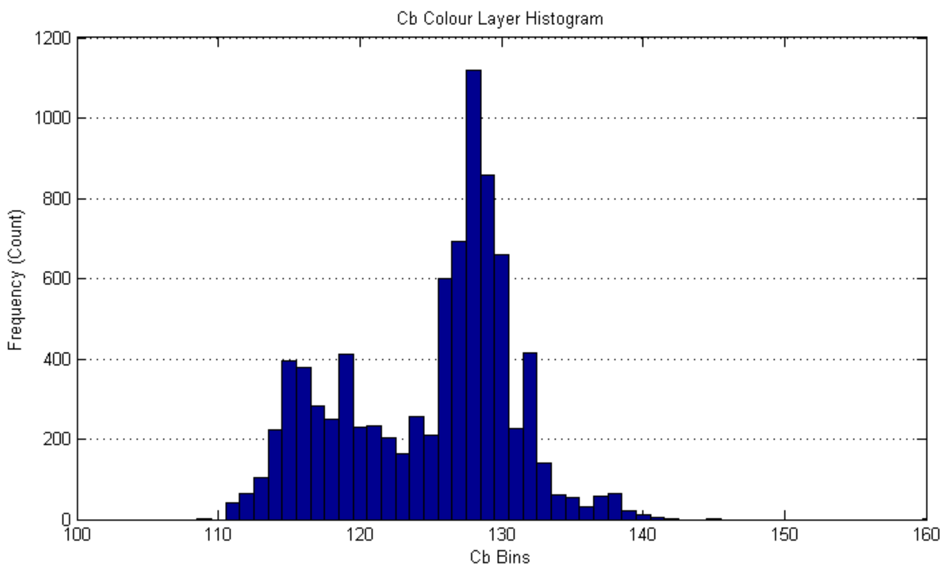


Figure 7.2: Cb Layer

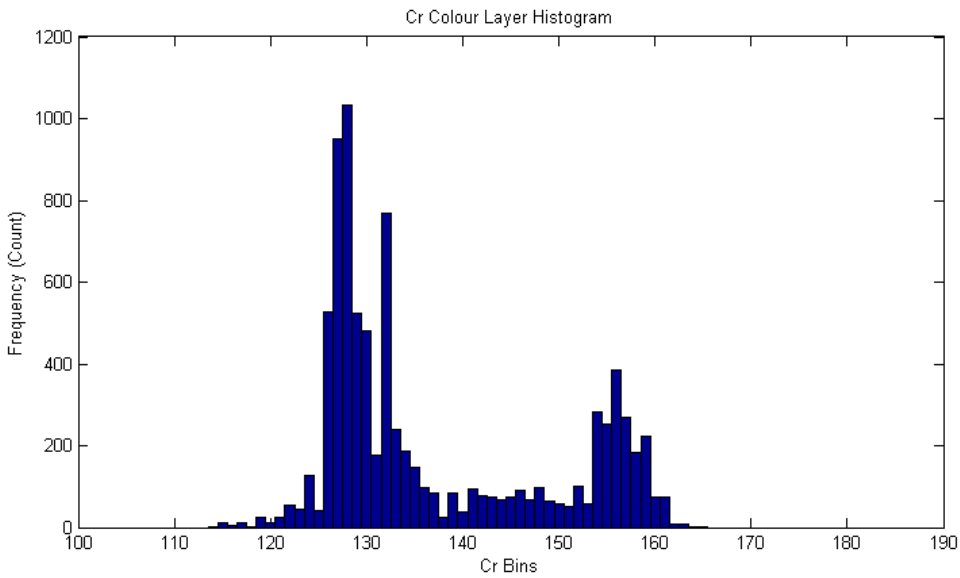


Figure 7.3: Cr Layer

It can be observed from the pixel histogram distribution on each chrominance layer that the distributions had different peaks. This indicates that for the different peaks it has different mean values and standard deviation. By observing the overall histogram, multiple normal distributions can be extracted. The multiple normal distributions can be combined to reform the overall histogram distribution. Each extracted normal distribution will have its own mean and standard deviation which would represent a specific colour. By using the EM method for extracting normal distributions, the extracted normal distributions for layers Cb and Cr can be seen in the following Figures 7.4 and 7.5. The extraction of the distributions were done using Matlab.

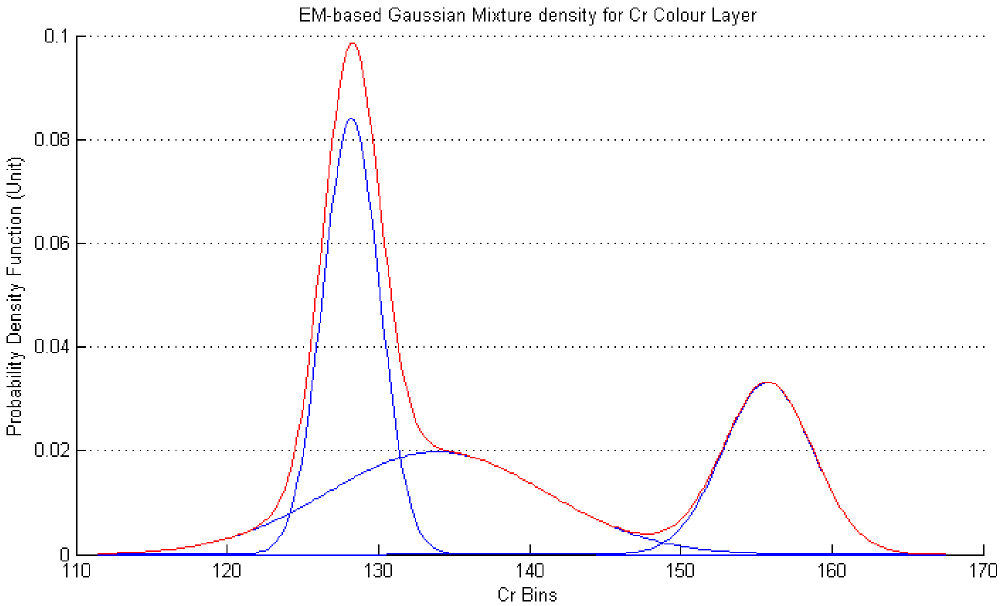


Figure 7.4: EM Extraction on Cr Layer

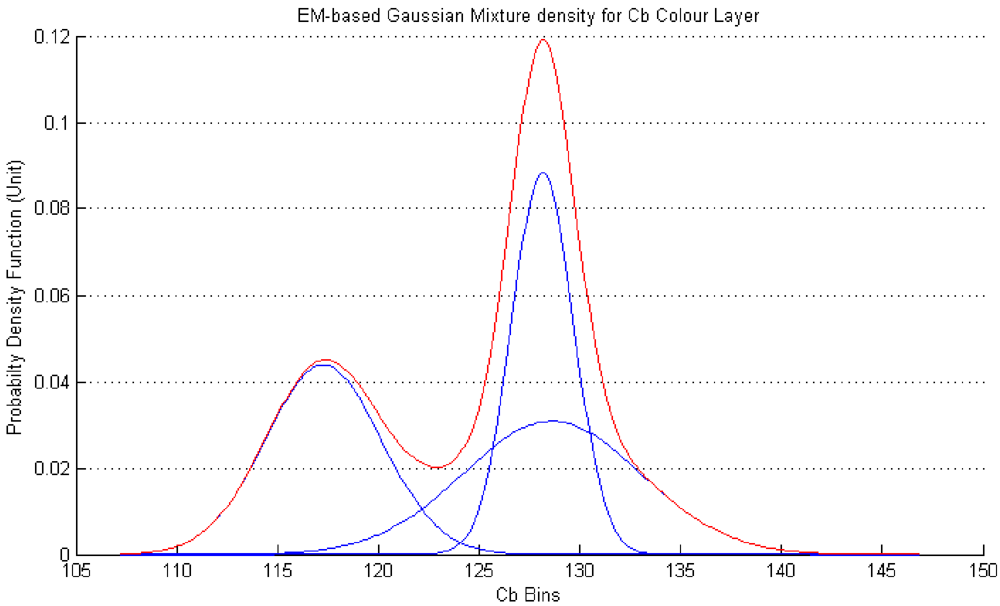


Figure 7.5: EM Extraction on Cb Layer

Figures 7.4 and 7.5 show the different Gaussian or normal distributions that can be extracted on the Cb and Cr layers respectively. Each normal distribution has its own mean and standard deviation and represents a specific colour. This information can then be used to segment regions according to the different colour regions based on the Cb and Cr layers. By having the necessary distribution means and standard deviations of each normal distribution, the pixels in the tracked target can be segmented into colour regions according to the parameters obtained. After the colour segmentation is done, the centroids of each segmented colour region can be obtained. The detected centroids are indicated by red and green crosshairs from the detected region in the Cb and Cr chrominance layers respectively.

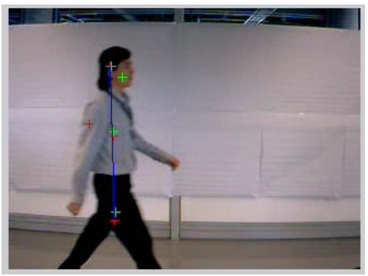


Figure 7.6: whole_seq1_1

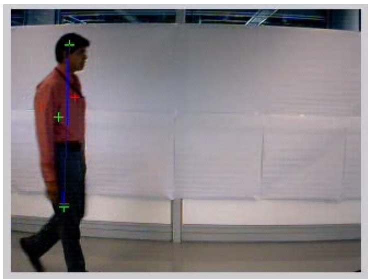


Figure 7.7: whole_seq1_2

Figure 7.8: whole_seq1_3

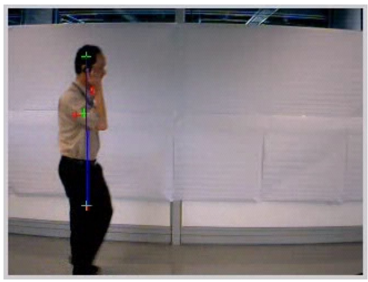


Figure 7.9: whole_seq1_4

Figure 7.10: whole_seq1_5

7.3. Parametric Tracking on Single Camera

The approach investigated here is the utilization of parameters between segmented colour regions. The target of interest, or person of interest, is first isolated from the background as a single blob. The blob, which is a region in an image where the image properties are constant, is identified by segmenting the foreground chromaticity against the average background pixel chromaticity [Zivkovic and Heijden, 2006, 773-780]. The single blob can be broken down into the different colour regions which means that a single blob can be broken down into a collection of blobs that represent the target of interest. The blob or the foreground image is segmented by using the EM algorithm in a GMM as mentioned in section 7.2. The accuracy of the segmentation is based on the probability of a pixel being assigned to the correct distribution and the correct number of normal distributions estimated to segment the colour region. The collection of different coloured blobs will have different locations on each of the colour layers or chrominance layers. Since a single target can now be represented by a group of different coloured blobs, the parameters that can be generated are based on the relative positions between detected blobs. The parameter investigated in this section was the vector between the top blob and the bottom blob. In addition to the vector parameter, an additional parameter investigated was the colour information of the blob used in generating the vector parameter.

7.3.1 Parameter Extraction

Observing a video in tracking a target, it can be seen that as the target moves across the camera view the detected regions have positional changes from frame to frame. With the change in the position of the colour regions from frame to frame, the parameters of interest will also change as the target moves across the camera view. Due to the variation in the positions of the detected colour regions, the parameters between colour regions such as vector magnitude, vector angle and colour information between regions, will generate a distribution of points representing the combination of parameters which is illustrated in Figure 7.11.

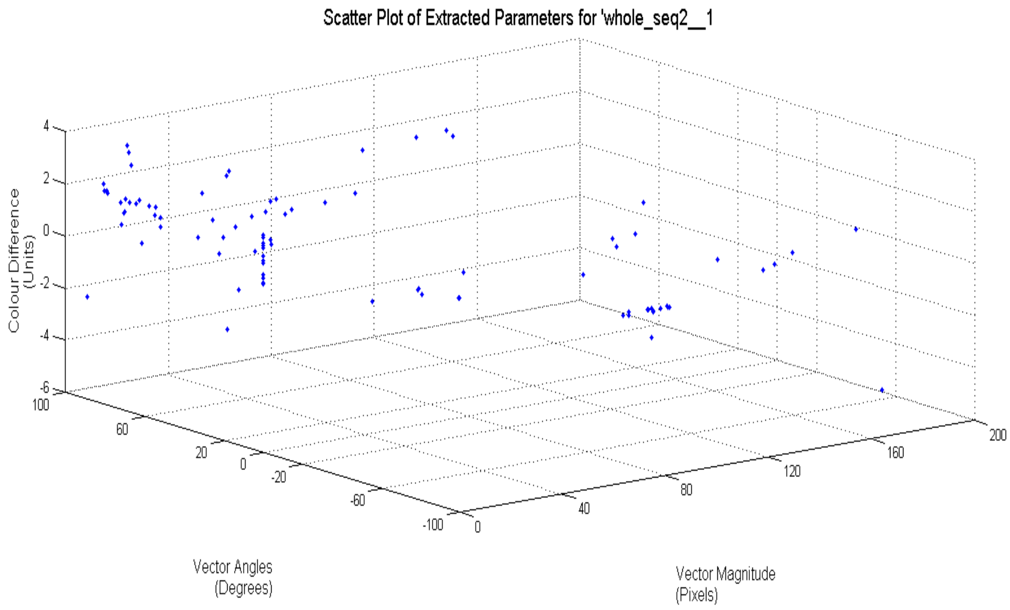


Figure 7.11: Scatter plot of extracted parameters

By looking at the scatter plot in Figure 7.11, the data points can be seen to be clustering at certain points which can lead us to generate parameters based on the clustering of the points [Alldrin, Smith et al., 2003]. The parameters to be generated are calculated utilizing the Expectation Maximization (EM) method to cluster the points using normal distribution as the reference distribution. The generated distributions on each axis would generate a Mixed Gaussian Distribution as can be seen in Figure 7.12. Tracking parameters are extracted automatically after the target of interest has passed through the view field of the camera that has been assigned to extract the necessary parameters.

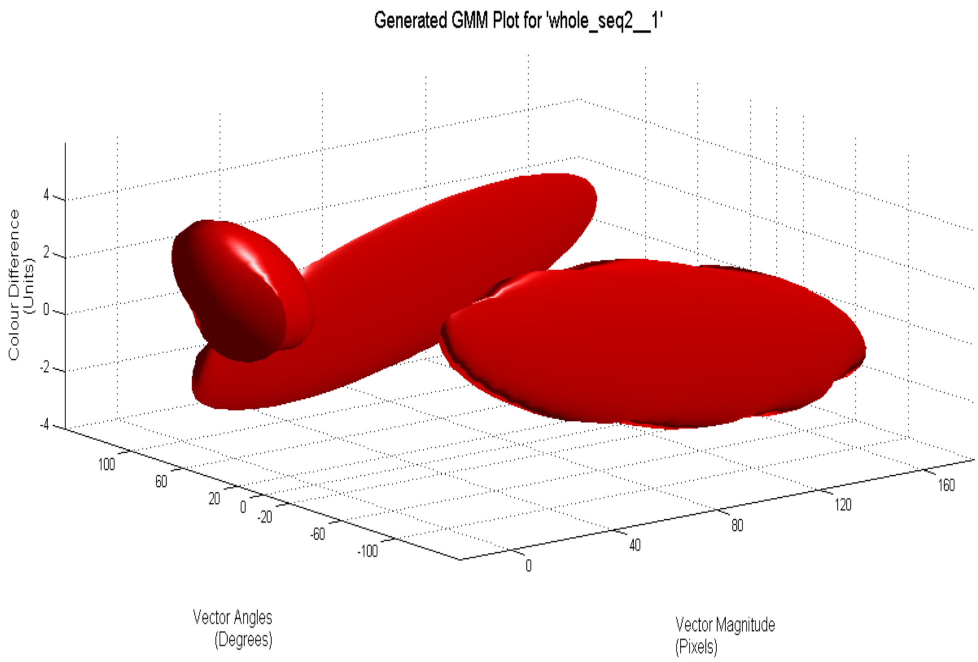


Figure 7.12: Mixed Gaussian Distribution of Extracted Parameters

Each axis represents a parameter used for tracking which in this case are the vector magnitude, vector angle and colour difference. Each parameter is then clustered using EM to obtain the normal distribution parameters for each detected normal distribution. The distribution parameters from each axis or tracking parameter is then combined to form a set of multiple regions. For each combination set of tracking parameters from each axis, individual regions of tracking parameters are generated. The regions are used as a means to detect if a track is successful or not by determining if the detected target parameters falls within the region. If the parameters were to fall within the region, it would be considered a successful track.

7.3.2 Vector Combination

The hypothesis investigated here is the viability of utilizing the vectors between detected colour regions on the different colour layers. A vector is typically defined as a quantity having a direction as well as a magnitude. In this instance, a vector would be a quantity between two different colour regions of interest. The vector parameters extracted for tracking a target are determined by computing all the possible combinations of vectors between the different colour regions. The generation of all possible combinations should hypothetically generate a set of parameters or a cluster of parameters that will be unique to each tracked target.

The tracking performance using the Cr and Cb layers alone can be seen in Tables 7.1 and 7.2 where the metric used is the TDR that has been discussed in Chapter 4 and used in Chapters 5 and 6. The detailed tracking results can be seen in Appendix 7 which lists the TDR and TC tracking metrics, which were mentioned in Chapter 4.

	TDR					TC				
	Input Video					Input Video				
Reference Histogram	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.58	0.23	0.04	0.40	0.08	0.09	0.03	0.00	0.06	0.01
whole_seq1_2	0.78	0.78	0.29	0.71	0.00	0.42	0.42	0.05	0.38	0.00
whole_seq1_3	1.00	0.74	0.93	1.00	0.09	1.00	0.18	0.59	1.00	0.00
whole_seq1_4	0.95	0.80	0.65	0.95	0.00	0.54	0.29	0.25	0.36	0.00
whole_seq1_5	0.94	0.76	0.81	1.00	0.06	0.37	0.33	0.34	1.00	0.00

Table 7.1: TDR rate for Cr layer tracking of vector combination

	TDR					TC				
	Input Video					Input Video				
Reference Histogram	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.65	0.10	0.42	0.81	0.00	0.09	0.01	0.08	0.27	0.00
whole_seq1_2	0.34	0.34	0.10	0.20	0.05	0.03	0.02	0.01	0.02	0.00
whole_seq1_3	0.77	0.53	0.18	0.91	0.05	0.15	0.04	0.03	0.48	0.00
whole_seq1_4	0.72	0.75	0.00	0.78	0.00	0.22	0.37	0.00	0.24	0.00
whole_seq1_5	1.00	0.09	0.22	1.00	0.06	1.00	0.00	0.05	1.00	0.00

Table 7.2: TDR rate for Cb layer tracking of vector combination

The results in Tables 7.1 and 7.2 show the tracking results of each input video against respective extracted reference parameters on Cr and Cb colour layers respectively. The results in Tables 7.1 and 7.2 indicate that there is no distinctive tracking performance as can be seen in the confusion matrix. The highlighted sections on Table 7.1 and 7.2 are supposed to be highest rates for each column. Results obtain in Tables 7.1 and 7.2 have shown results that had high tracking rates against incorrect reference parameters, which indicate that the vector combination parameters used were incapable of correctly tracking a target of interest. The high tracking rates of incorrect input video against respective reference video parameter are indications of incorrect tracking which, in this instance, is all instances as the incorrect tracking rates are higher than the correct tracking rates. Based on the results obtained, the utilization of the parameters of all possible vectors between all segmented regions has shown inconsistent tracking performance for the different input videos and reference videos.

7.3.3 Maximum Vector

The investigation done in this section is to determine the viability of the usage of information between the top and bottom colour regions of the tracked target. The information that can be generated between the colour regions are vector magnitude and vector angle between colour regions. An additional data point that can be extracted between the colour regions is the colour information or colour information difference between colour regions. The regions are obtained from the segmentation of the pixels based on the overall distribution of the pixels using the EM extraction on the Cb and Cr layers respectively. In this instance, the top most region that will be detected will be the head [Chang and Gong, 2001, 19-26; Siebel and Maybank, 2002, 373-387; Yan and Forsyth, 2005, 370-377] as this is a portion that is most easily detected. The reference point to obtain the maximum vector will be the limbs or the lower torso on the tracked target as the lower torso provides a point of detection that will be as far away as possible from the head [Diaz de Len and Sucar, 2002, 439 - 442; Ramanan, A. Forsyth et al., 2007, 65-81] . The complete tracking results, using the maximum vector between the top and bottom detected regions, can be seen in Appendix 8 where the data from Table 7.3 to Table 7.6 are obtained.

	TDR					TC				
	Input Video					Input Video				
Reference Histogram	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.83	1.00	0.00	0.10	0.00	0.31	1.00	0.00	0.01	0.00
whole_seq1_2	0.00	0.41	0.10	0.05	0.00	0.00	0.04	0.01	0.00	0.00
whole_seq1_3	0.00	0.78	0.50	0.04	0.15	0.00	0.39	0.09	0.00	0.01
whole_seq1_4	0.05	0.85	0.32	0.82	0.10	0.00	0.57	0.03	0.57	0.01
whole_seq1_5	0.00	0.98	0.02	0.48	0.83	0.00	0.96	0.00	0.18	0.28

Table 7.3: TDR rate for Cb layer tracking using maximum vector

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.69	0.00	0.42	0.00	0.04	0.21	0.00	0.08	0.00	0.00
whole_seq1_2	0.05	0.44	0.10	0.00	0.02	0.00	0.08	0.01	0.00	0.00
whole_seq1_3	0.30	0.07	0.92	0.00	0.12	0.02	0.00	0.34	0.00	0.00
whole_seq1_4	0.60	0.20	0.30	0.00	0.15	0.10	0.01	0.02	0.00	0.01
whole_seq1_5	0.61	0.00	0.37	0.37	0.98	0.17	0.00	0.04	0.12	0.83

Table 7.4: TDR rate for Cr layer tracking using maximum vector

Based on the data obtained, the tracking performance on different layers has different and contradicting tracking performances. This can be seen in Table 7.3 and 7.4 where related reference data from the related video runs gave contradicting tracking performances such as reference data from 'whole_seq1_2' and 'whole_seq1_4'. The data in the Tables 7.3 and 7.4 are TDR rates for each input video against related reference data extracted from each video. The TDR rate for the reference video of 'whole_seq1_2' and 'whole_seq1_4' compared against its own input video were expected to be the highest rate for its respective columns. However, in this instance for 'whole_seq1_2' and 'whole_seq1_4' as reference video parameters the tracking rate were not the highest rates for its own videos and the other video inputs had higher tracking rates on either the Cb or Cr colour layer. This raises the question of which colour layer should be used for tracking a target. Looking at the tracking performances on separate colour layers the next step is to determine the tracking performance using detected regions on the Cb and Cr layers. The vector is generated between the highest region on either colour layer to the lowest region on either colour layer. Tracking results using the vector between the different colour layers can be seen in Table 7.5.

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.38	0.00	0.08	0.50	0.00	0.11	0.00	0.01	0.08	0.00
whole_seq1_2	0.00	0.88	0.12	0.24	0.02	0.00	0.45	0.01	0.02	0.00
whole_seq1_3	0.00	0.15	0.82	0.46	0.03	0.00	0.01	0.12	0.05	0.00
whole_seq1_4	0.02	0.32	0.35	0.90	0.00	0.00	0.05	0.08	0.65	0.00
whole_seq1_5	0.04	0.61	0.96	0.54	0.20	0.00	0.11	0.83	0.10	0.04

Table 7.5: TDR rate for Cb and Cr Combination Tracking using maximum vector

Results from Table 7.5 indicate that the tracking performance using the regions detected on different colour layers has better tracking performance in terms of correct tracking to incorrect tracking performance. This can be inferred from the tracking rates that for every reference video, the respective input video had the highest tracking rate compared to the other input videos. The utilization of the different regions as a means to generate the vector and the colour information of the regions gives better tracking where the correct tracking performance is higher compared to incorrect tracking. By comparing the incorrect tracking performance listed in Table 7.3, 7.4 and 7.5, the utilization of different colour layers has lower incorrect tracking compared to single colour layer tracking. Results from single colour layers have shown to have inconsistent tracking results thus the combining of information from multiple colour layers such as the Cb and Cr colour layers would provide consistent tracking results. Parametric tracking is performed based on the vector between regions of interest and the colour information of the regions of interest. The tracking parameters are based on the vector between the regions of interest where the vector has two properties which are the magnitude and direction, and in addition to the vector parameter the colour difference between regions is used as an additional tracking parameter. Tracking performance using a vector parameter alone for tracking a target can be seen in Table 7.6 where colour difference between regions is not used as a tracking parameter.

	TDR					TC				
	Input Video					Input Video				
Reference Histogram	whole _seq1 _1	whole _seq1 _2	whole _seq1 _3	whole _seq1 _4	whole _seq1 _5	whole _seq1 _1	whole _seq1 _2	whole _seq1 _3	whole _seq1 _4	whole _seq1 _5
whole_seq1_1	0.90	0.00	0.73	0.54	0.02	0.34	0.00	0.19	0.08	0.00
whole_seq1_2	0.00	0.93	0.22	0.29	0.05	0.00	0.52	0.01	0.03	0.00
whole_seq1_3	0.03	0.15	0.91	0.53	0.23	0.00	0.01	0.16	0.05	0.01
whole_seq1_4	0.13	0.38	0.85	0.90	0.50	0.01	0.05	0.15	0.65	0.19
whole_seq1_5	0.04	0.61	0.96	0.54	0.94	0.00	0.11	0.83	0.10	0.54

Table 7.6: TDR rate for Cb and Cr Combination Tracking without colour information

Based on the results in Table 7.6, the correct tracking rate has shown a significant increase but at the same time the incorrect tracking rate had shown a significant increase in comparison to the results in Table 7.5 across all video sequences. In reducing the number of parameters used in video tracking, which in this case is the difference between colour regions, it would be expected that all track rates would increase that includes both the correct and incorrect track. In reducing the number of parameters for tracking, the tracking rates for both correct and incorrect tracks would increase as it would be just like loosening the specification of detection that would make more tracks be accepted. Utilization of colour information together with the vector between regions of interest gave better tracking ability in terms of correct tracking and lowered rates of incorrect tracking.

Looking at the different tables listing the tracking performances using the maximum vector between detected regions, the outcome from the different investigations lead to some conclusions. Among the conclusions obtained was that the utilization of a single colour layer for tracking purpose was not able to provide consistent tracking performance compared to tracking using multiple colour layers. Regions to be used as a means for determining the vector between them are based on the different colour layers. The parameters used for tracking a target will then be a combination of the vector between regions and the colour information of a region. This is where a contribution to the body of knowledge in tracking of a target is made where a target of interest is tracked based on the parameters generated between colour regions on different colour layers. A target is not tracked based on the overall colour distribution, height and size but rather by the information that can be generated between the colour regions that make up the target of interest. In determining the ability of using the vector for tracking a target of interest, a comparison of tracking performance can be seen in Table 7.7 that compares the vector tracking against normalized cross correlation and histogram similarity comparison.

Input Video	Normalized Cross Correlation	Histogram Similarity Comparison					Vector Tracking
		Histogram Intersection	Euclidean Intersection	Chi-Squared 1	Chi-Squared 2	Kullbeck-Leibler	
whole_seq1_1	0.00	0.29	0.19	0.31	0.25	0.27	0.92
whole_seq1_2	0.00	0.24	0.00	0.29	0.29	0.29	0.90
whole_seq1_3	0.00	0.20	0.00	0.15	0.16	0.12	0.95
whole_seq1_4	0.00	0.15	0.20	0.28	0.33	0.35	0.90
whole_seq1_5	0.00	0.20	0.00	0.19	0.28	0.19	0.94

Table 7.7: TDR rate comparison of different tracking methods

Based on the TDR rates of different tracking methods, it can be seen that the utilization of vectors between colour regions has higher tracking ability compared to normalized cross correlation and histogram similarity tracking methods.

7.3.4 Image Segmentation

The property to be investigated in this section is the effect of image segmentation based on pixel distribution on each colour layer. Colour segmentation is described in section 7.2 which breaks the overall pixel distribution into multiple Gaussian distributions. The pixels are separated into multiple Gaussian distributions using the EM method where each distribution has its own pixel mean. Each pixel was then classified or segmented into different regions based on the Gaussian distributions that were extracted from the overall pixel distribution. The investigation done here was to investigate the effect on the number of distributions or segmented regions on tracking capability. The number of distributions are decided based on extracted normal distribution parameters such as mean and variance. Criteria on the distribution parameters are that the means are not overlapping, which is to ensure that the distributions are not overlapping, and that the variance of each distribution is to be at minimum values. Tracking performance can be seen in the Table 7.8 and Table 7.9.

	TDR					TC				
	Reference Video					Reference Video				
Input Video	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.38	0.00	0.08	0.50	0.00	0.11	0.00	0.01	0.08	0.00
whole_seq1_2	0.00	0.88	0.12	0.24	0.02	0.00	0.45	0.01	0.02	0.00
whole_seq1_3	0.00	0.15	0.82	0.46	0.03	0.00	0.01	0.12	0.05	0.00
whole_seq1_4	0.02	0.32	0.35	0.90	0.00	0.00	0.05	0.08	0.65	0.00
whole_seq1_5	0.04	0.61	0.96	0.54	0.20	0.00	0.11	0.83	0.10	0.04

Table 7.8: TDR rate for Cb and Cr Combination Tracking with 5 colour segments

	TDR					TC				
	Reference Video					Reference Video				
Input Video	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.92	0.00	0.31	0.60	0.00	0.43	0.00	0.05	0.10	0.00
whole_seq1_2	0.00	0.90	0.24	0.22	0.05	0.00	0.48	0.02	0.02	0.23
whole_seq1_3	0.03	0.15	0.95	0.47	0.39	0.00	0.01	0.45	0.04	0.06
whole_seq1_4	0.08	0.32	0.45	0.90	0.47	0.00	0.05	0.04	0.59	0.09
whole_seq1_5	0.04	0.61	0.39	0.54	0.94	0.00	0.11	0.04	0.11	0.68

Table 7.9: TDR rate for Cb and Cr Combination Tracking with 6 colour segments

Looking at tracking performance listed in Table 7.8 and 7.9 for 5 and 6 colour regions segmentation respectively, tracking performance has different rates when targets of interest are segmented into a different number of colour regions. Looking at performance data in

Table 7.8 and 7.9, segmenting targets of interest changes the tracking rates especially for video sequence 'whole_seq1_3'. For this video sequence, when the target is segmented into 5 colour regions the correct track rate was at 0.82 as listed in Table 7.8. This result is overshadowed by the result of the tracking rate on the 'whole_seq1_5' at 0.96 and this result shows that the tracking algorithm was not able to distinguish between both targets and thus creates an incorrect tracking scenario. This scenario changes when the number of colour region segmentation is increased to 6 and the results are listed in Table 7.9. In this instance, the correct tracking rate had increased to 0.95 and the highest incorrect tracking rate was at 0.45. The increase in the number of colour segmented regions had improved the correct tracking performance while reducing the incorrect tracking rate, to ensure the robustness of the tracking algorithm. However, in the same set of data it can be seen that for 'whole_seq1_1' and 'whole_seq1_5' the tracking rates were at 0.38 and 0.20 respectively when the number of colour regions was set at 5. These rates had increased to 0.92 and 0.94 when the number of colour regions segmentation was set to 6. This increase in the correct tracking rates is a desired result but this increase also has to take into account the increase in the incorrect tracking rates. This can be seen in the video sequence of 'whole_seq1_5' where the correct track had changed from 0.20 to 0.94 but at the same time the incorrect track had changed from 0.02 to 0.47. These results show that by increasing the number of colour regions or colour region segmentation the tracking results improve but increase in the number of segments could also increase the incorrect tracking rates. Therefore, further investigation would need to be done to determine if the reduction in the segmented colour regions have better tracking performance.

7.4. Parametric Tracking on Multiple Cameras

The hypothesis investigated here is that parametric tracking of a target on multiple cameras will improve detection reliability. The work done here is an extension of previous work done in the area of target tracking using multiple cameras. The initial work done on multiple camera tracking was that multiple cameras were observing the same area to ensure that correct target tracking was being identified [Cai and Aggarwal, 1996, 68-72; Ellis, 2002, 228-233]. Such tracking methodology used the information of trajectory, colour and size of the target blob as parameters in tracking targets of interest [Choi and Yoo, 2013, 125-126; Ellis, 2002, 228-233]. Apart from general blob parameters used in tracking targets, other parameters such as histogram of gradients, colour histogram and reference set of images were used as parameters in tracking targets in a multi camera environment [Chen, An et al., 2015, 1; Hsu, Yang et al., 2013]. The investigation-reported here is the utilization of parameters as a means of tracking a target across an array of non-overlapping cameras [Chen, An et al., 2015, 1]. The investigation looks into the parameters that are present between colour blobs that are the representation of the target person. In investigating the parameters for tracking a person of interest across an array of cameras the setup of the system comprises three cameras where the first camera has a uniform background, the second camera has an uneven background and camera three has a noisy environment. The utilization of the different environment setup was to evaluate the tracking in a real-life condition which would have different background conditions. The cameras were set up where targets could be tracked as they move across the view field of Camera 1 to Camera 2 and Camera 3. The targets or persons being tracked are based on the parameters extracted from a camera that is placed before the first tracking camera, camera 1, and the parameters that were extracted were the vector between regions of interest and the colour difference information between the regions. The targets are then tracked using the same parameters as the targets move across the view fields of Camera 2 and Camera 3. The tracking of targets on multiple cameras is based on the parameters and segmentation settings used in tracking a target on a single camera setting.

A sample of the raw tracking performance data using multiple cameras is listed in the following tables:

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.70	0.44	0.36	0.17	0.60	0.06	0.73	0.33	0.20
whole_seq2_2	0.87	0.97	0.87	0.38	0.69	0.11	0.90	0.84	0.34
whole_seq2_3	0.95	0.87	0.96	0.68	0.83	0.60	0.94	0.86	0.57
whole_seq2_4	1.00	1.00	1.00	0.92	0.91	0.29	0.96	0.96	0.53
whole_seq2_5	0.17	0.13	0.13	0.03	0.26	0.00	0.23	0.08	0.06
whole_seq2_6	0.78	0.92	0.77	0.56	0.62	0.55	0.83	0.84	0.23
whole_seq2_7	0.65	0.61	0.57	0.23	0.71	0.04	0.80	0.42	0.10
whole_seq2_8	0.89	0.97	0.92	0.44	0.73	0.21	0.92	0.97	0.48
whole_seq2_9	0.97	1.00	0.99	0.64	0.89	0.35	0.88	0.97	0.96

Table 7.10: TDR rate of Camera 1

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.66	0.44	0.41	0.27	0.00	0.08	0.58	0.38	0.27
whole_seq2_2	0.70	0.84	0.84	0.33	0.00	0.21	0.86	0.57	0.31
whole_seq2_3	0.93	0.96	0.91	0.65	0.00	0.30	0.82	0.88	0.39
whole_seq2_4	0.93	1.00	0.86	0.59	0.00	0.33	1.00	0.90	0.50
whole_seq2_5	0.40	0.13	0.12	0.60	0.00	0.00	0.40	0.10	0.05
whole_seq2_6	0.75	0.98	0.89	0.49	0.00	0.30	0.85	0.81	0.43
whole_seq2_7	0.65	0.63	0.51	0.21	0.00	0.07	0.88	0.37	0.18
whole_seq2_8	0.93	0.86	0.95	0.63	0.00	0.36	0.93	0.93	0.51
whole_seq2_9	0.86	0.99	0.94	0.38	0.00	0.17	0.90	0.92	0.74

Table 7.11: TDR rate of Camera 2

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.78	0.18	0.18	0.10	0.00	0.01	0.76	0.14	0.11
whole_seq2_2	0.91	0.89	0.86	0.39	0.00	0.29	0.98	0.82	0.46
whole_seq2_3	0.73	0.94	0.86	0.41	0.00	0.30	0.95	0.82	0.25
whole_seq2_4	0.93	0.97	0.95	0.73	0.00	0.14	0.98	0.92	0.55
whole_seq2_5	0.88	0.57	0.55	0.25	0.00	0.16	0.90	0.40	0.24
whole_seq2_6	0.65	0.92	0.75	0.21	0.00	0.00	0.86	0.77	0.34
whole_seq2_7	0.68	0.66	0.66	0.23	0.00	0.10	0.73	0.32	0.29
whole_seq2_8	0.80	0.92	0.87	0.38	0.00	0.20	0.92	0.80	0.34
whole_seq2_9	0.80	0.84	0.99	0.49	0.00	0.45	0.83	0.83	0.60

Table 7.12: TDR rate of Camera 3

A summary of the detailed tracking performance based on data from Tables 7.10 to 7.12 is illustrated in the following Table 7.13(a) and 7.13(b). The summary tracking data in Table 7.13 summarizes the tracking of a target against its own set of parameters to indicate the correct tracking rate for each target. A detailed list of the all the tracking results can be seen in Appendix 9.

Ref. Video	Number of EM Segmentation Distributions											
	raw6			raw5			raw4			raw3		
	cam 1	cam 2	cam 3	cam 1	cam 2	cam 3	cam 1	cam 2	cam 3	cam 1	cam 2	cam 3
whole_seq2_1				0.76	0.7	0.78						
whole_seq2_2												
whole_seq2_3				0.92	0.69	0.59						
whole_seq2_4				0.93	0.58	0.57	0.84	0.16	0.23	0.20	0.17	0.10
whole_seq2_5												
whole_seq2_6												
whole_seq2_7				0.83	0.82	0.76						
whole_seq2_8												
whole_seq2_9	0.96	0.74	0.60				0.88	0.83	0.63	0.68	0.42	0.40

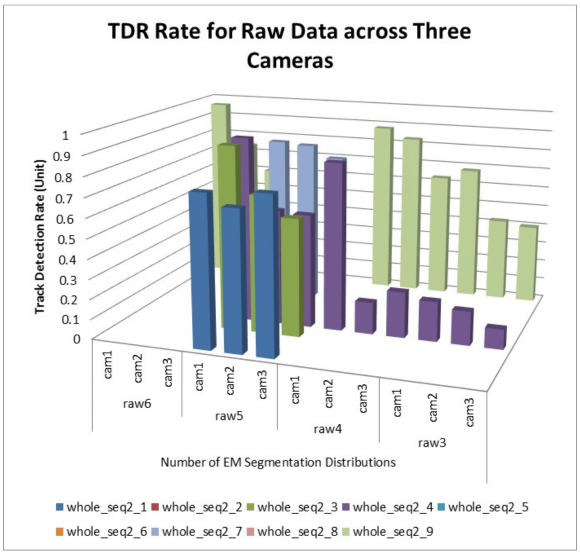
a) TDR rates with raw data

Ref. Video	Number of EM Segmentation Distributions											
	out6			out5			out4			out3		
	cam 1	cam 2	cam 3	cam 1	cam 2	cam 3	cam 1	cam 2	cam 3	cam 1	cam 2	cam 3
whole_seq2_1				0.76	0.7	0.82						
whole_seq2_2												
whole_seq2_3				0.89	0.61	0.57	0.94	0.62	0.62	0.64	0.61	0.37
whole_seq2_4	0.87	0.66	0.68	0.92	0.44	0.54				0.19	0.14	0.15
whole_seq2_5												
whole_seq2_6				0.71	0.59	0.26						
whole_seq2_7												
whole_seq2_8												
whole_seq2_9	0.88	0.67	0.55				0.84	0.77	0.57	0.68	0.42	0.40

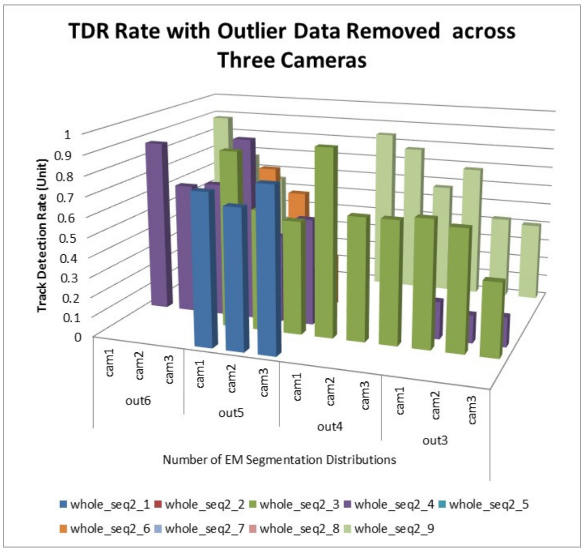
b) TDR rates with outlier data removed

Table 7.13: TDR Rate across 3 cameras

Table 7.13 summarizes the tracking performance of tracking a target across 3 cameras by listing the number of colour region segmentation for each run or target from Table 7.10, 7.11 and 7.12. The data in Table 7.13(a) and (b) are illustrated in Figure 7.13(a) and (b) respectively.



(a) Raw data illustration



(b) Outlier Data Removed Illustration

Figure 7.13: TDR Rate across Three Cameras

The empty cells in the table and chart indicate that the correct tracking of a target was not achieved across the 3 cameras. The cells with tracking data rates indicate that the target was correctly tracked and had the highest tracking rate. The random videos used for testing tracking showed that tracking data from whole_seq2_1, whole_seq2_3, whole_seq2_4, whole_seq2_6, whole_seq2_7 and whole_seq2_9 had parameters that enabled a target to be tracked from camera1 through camera3. The videos of whole_seq2_2, whole_seq2_5 and whole_seq2_8 did not generate the necessary tracking parameters to enable the related target to be tracked across three cameras. These runs did not have any related tracking across three cameras due to the unavailability of related vectors between the top and bottom regions. In the case of whole_seq2 and whole_seq2_8, the top and bottom regions were detected in the same colour space layer. The tracking performance when using regions from the same colour layer can be seen to be not consistent which is seen in initial evaluations of tracking on a single colour layer in Table 7.3 and 7.4. The results in those tables indicated that tracking in a single layer had inconsistent tracking performance. In this instance, the results from whole_seq2_2 and whole_seq2_8 have confirmed that the utilization of a single colour layer for region detection and tracking would not be able to track targets correctly and consistently. For whole_seq2_5, there was only one region detected across both colour space layers. This was due to the fact that the whole target was generally a single colour and that there were no different colour regions. In this instance there are no colour regions to detect and thus no vector to be used as a means for detection and tracking a target.

In addition to the summary of the tracking performance across three cameras, Table 7.12 also shows summarized tracking data based on the number of EM segmentation distributions, which indicate that the colour pixels were segmented into the respective number of distributions ranging from 6 to 3 distributions. In addition, the tracking performance is further segregated into groups indicated with ‘raw’ and ‘out’, where the data indicated as ‘raw’ used all parameters that were generated between the colour regions of interest as the target moved across the camera view field to obtain the tracking parameters. This is in contrast to the data indicated as ‘out’ that had some outlier data removed to overcome the problem of incorrect parameters being generated due to incorrect segmentation of colour regions.

7.5. Summary

This chapter presented a method by which targets may be tracked based on the parameters between different colour segments. The first step in the process was the segmenting of the image into regions with similar pixel values. This segmentation is done based on the parameters extracted from the overall pixel distribution into a fixed number of normal distributions, based on the EM method. After the EM segmentation is done, the input image was then separated into different colour regions with each region having its own colour properties or values. This made a person that is being tracked a collection of different colour regions. With each person or target of interest as a collection of colour regions, the assumption is that the relative position of all regions would remain the same throughout the movement of the target across the camera view field. The relative positions of the different colour regions are also assumed to be similar across the different cameras. It has to be noted that tracking of colour segments detected on a single colour layer does not provide accurate enough information for tracking a target. As such, the investigation done in this chapter is based on the utilization of the information between the regions such as the vector between regions of interest and the colour information of the related regions. From related experiments, it was observed that the parameters of vector magnitude, vector angle and colour difference between colour regions provided better tracking accuracy compared to vector magnitude and angle alone. This initial work was on a single camera view. Investigation of parametric tracking was furthered by increasing the number of cameras for a target to be tracked. In this instance the tracking of targets was done across three cameras and the results indicated that targets could be tracked using the parameters of interest alone with results listed in Table 7.12. The investigation reported in this chapter indicates that the colour regions of interest and even the parameters that exist between these colour regions are consistent across the view field of single cameras and also across the view field of an array of cameras.

CHAPTER 8

Parametric Multi Camera Tracking with Spatial Estimation

8.1. Introduction

This chapter describes a tracking method based on a combination of general blob parameters and parameters between colour regions of interest. Together with blob information and parameters between colour regions, this can be used as a means for determining the spatial location of a target of interest or the relative location of the cameras.

8.2. Multi-camera Tracking.

In a multi-camera surveillance system, different configurations of multi-camera systems are employed such as overlapping [Behera, Kharade et al., 2012, 102-108] and non-overlapping configurations [Hommel, Grimm et al., 2012, 175 - 180; Hsu, Yang et al., 2013]. Depending on the application of the surveillance system, either one of the surveillance configurations may be used. Distributed, non-overlapping surveillance systems are typically applied in large public areas such as airports [Hommel, Grimm et al., 2012, 175 - 180], walkways [Bredereck, Jiang et al., 2012, 1 - 6] and streets [Wang, Velipasalar et al., 2011, 1 - 6]. In tracking targets in a surveillance system, the typical approach is to implement the tracking methodology on a single camera first and then implement the methodology across a distributed video surveillance system [Chen, Cao et al., 2014, 2329-2333]. Among the common tracking methodologies used in a distributed or multi-camera video surveillance system are colour mean-shift tracking [Behera, Kharade et al., 2012, 102-108; Hsu, Yang et al., 2013; Shen, Zhao et al., 2010, V3-409 - V403-412; Wang, Liu et al., 2009], colour and texture [Aziz, Mustafah et al., 2014, 243-246; Hommel, Grimm et al., 2012, 175 - 180], object size [Bredereck, Jiang et al., 2012, 1 - 6; Mehmood, 2009, 1-6; Wang, Velipasalar et al., 2011, 1 - 6], height of target [Behera, Kharade et al., 2012, 102-108], colour histogram [Hassan, Bangalore et al., 2011, 289-294; Shen, Zhao et al., 2010, V3-409 - V403-412] and trajectory [Chen, Chen et al., 2011, 1-4; Eldib, Bo et al., 2014, 378-382; Komlen, Lombarovic et al., 2012, 1798-1802].

Looking at the tracking methodologies used in tracking a target across a multi-camera system, it can be seen that a number of the methods are based on the general parameters extracted from a blob such as size, height, vector and trajectory of the blob itself. Other detailed parameters that can be extracted for use in tracking a target are colour and texture information of the blob. The colour information is also used as a means for generating a histogram for tracking.

In this chapter, some of the tracking methodologies described earlier are used to complement the tracking methodology described in Chapter 7 that tracks a target based on the parameters between colour regions. These tracking parameters are used together with general blob parameters to track targets across a multi-camera system.

8.3. Parametric Tracking on Multiple Camera Systems

Investigating parametric tracking across multiple cameras, a further series of experiments was conducted using the segmented regions parametric tracking method, previously described. The experiment was conducted based on 4 different scenarios which are listed as follows:

Scenario 1 – this scenario is based on tracking a target, either Target1 or Target2, across a three camera tracking system. The following figure will better illustrate the setup of the scenario and the expected tracking of the target.

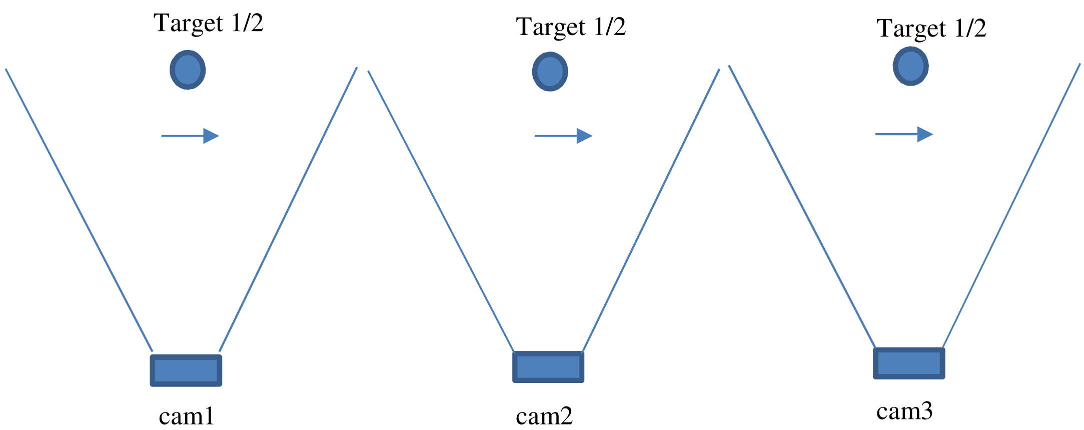


Figure 8.1: Scenario 1

Scenario 2 – this scenario is based on tracking a target across two cameras in a three camera tracking system. This scenario is to determine if the target is missing in the camera view field. The following figure illustrates the setup of Scenario 2.

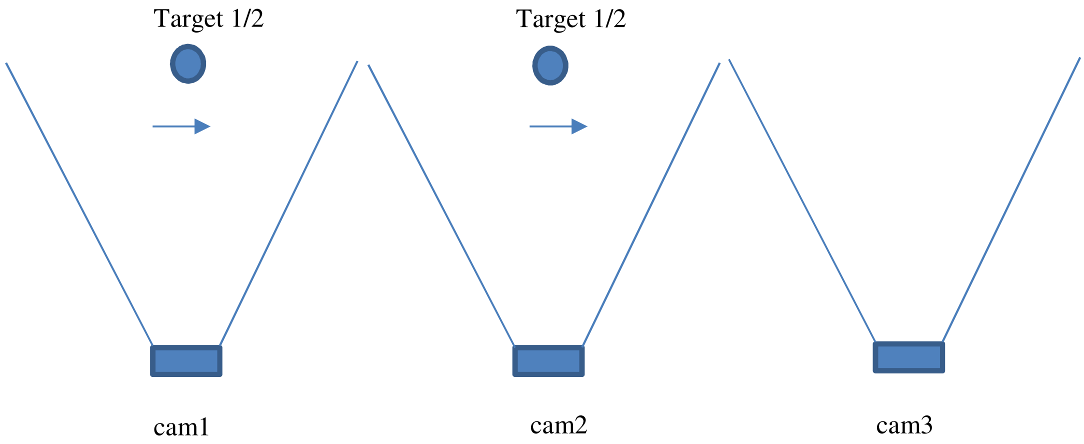


Figure 8.2: Scenario 2

Scenario 3 – this scenario is based on tracking a specific target as two targets walk across the view field of cameras. The following figure illustrates the tracking of specific target across an array of cameras.

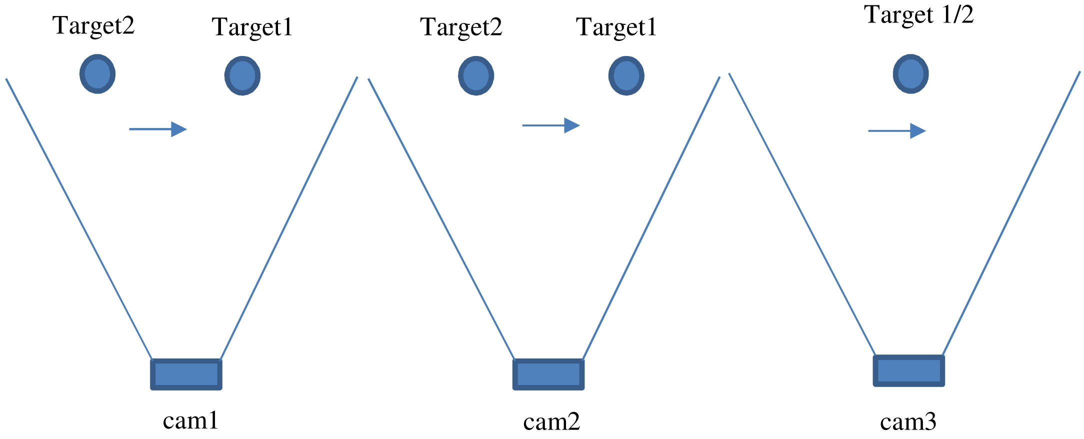


Figure 8.3: Scenario 3

Scenario 4 – this scenario is based on tracking a specific target as two targets move in opposite directions. There will be a short period where the target of interest will be occluded by the second moving object within the camera view field. This scenario will also determine if the tracking methodology is able to continue tracking the target of interest after the occlusion.

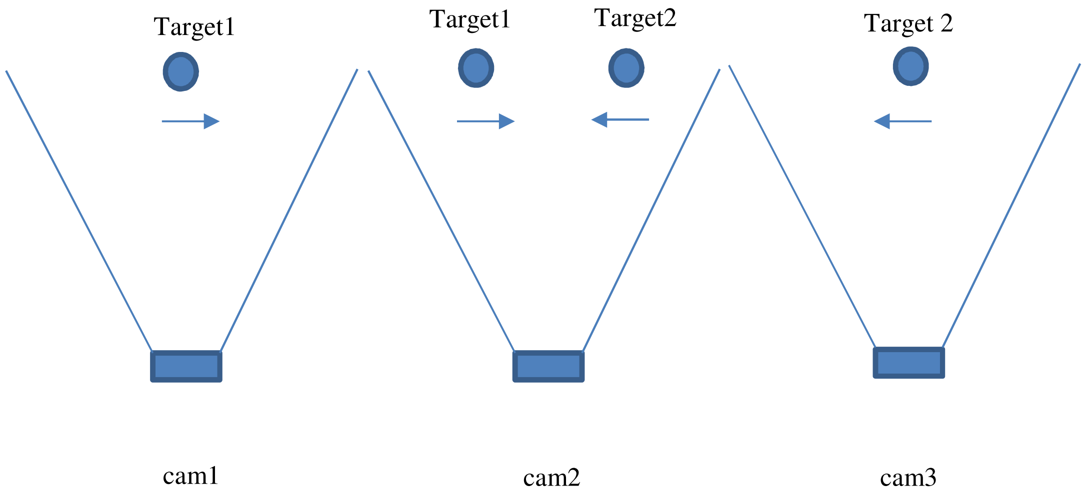
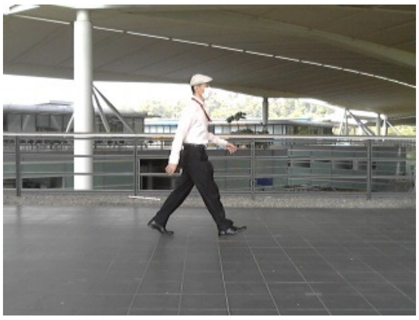
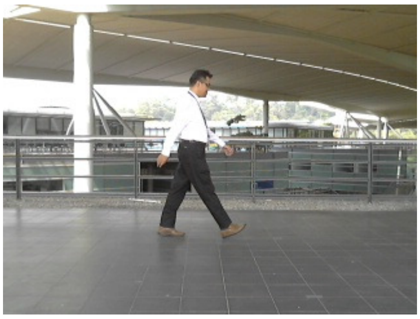


Figure 8.4: Scenario 4

The setup of the experiments required that two different targets be used for the different scenarios. Sample images from the video sequences for each scenario can be seen in the following figures.



(a) Target1 from cam1



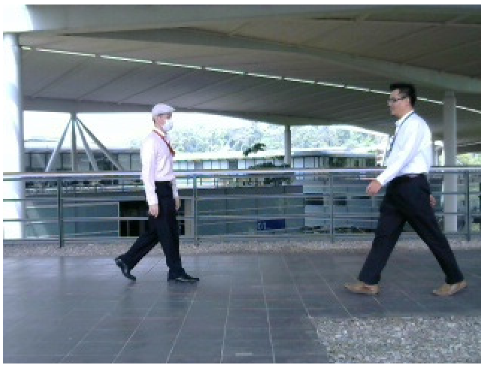
(b) Target2 from cam1

Figure 8.5: Sample Images from Scenario 1

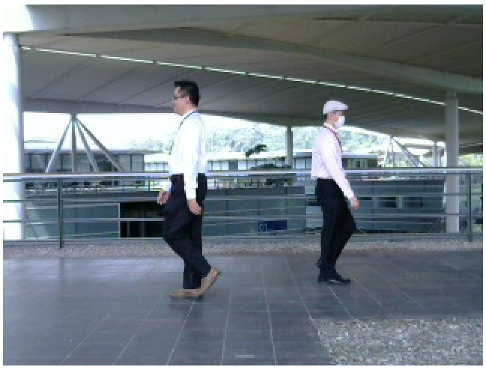
(a) Sample from cam1

(b) Sample from cam1

Figure 8.6: Sample Images from Scenario 3



(a) Sample from cam2



(b) Sample from cam2

Figure 8.7: Sample Images from Scenario 4

The volunteer targets had worn multiple colour clothing which was as requested in obtaining the latest set of sample video sequences. The first target, who is also designated as ‘whole_seq3_1’, had worn a light pink shirt with dark pants. In addition to the clothing worn he had also a cap and a face mask worn on the day of the video capture. The second target, who can be designated as ‘whole_seq3_2’, had worn a white shirt and dark pants. In the sample snap shots seen in Figures 8.5, 8.6 and 8.7, the images show that both targets have similar coloured clothing although the colour of the clothing or shirts worn were different. The purpose of the different colour clothing was to ensure that the segmentation of the colour regions would be better as compared to targets that have all pieces of clothing to be dark or a single colour. It has to be kept in mind that the tracking method being evaluated was the utilization of parameters between the top and bottom colour regions.

The cameras used in these scenarios were the same type of cameras, which in this case were the Logitech 270 where the capture frame rate was set to 30 frames per second and at a resolution of 320x240 pixels. The cameras were not calibrated or tuned to each other as they were used as they were taken out of their boxes. The cameras were placed in locations where the camera view fields were not overlapping. The data that is transferred between camera tracking stations were vector magnitude, vector angle, colour difference between regions, average velocity of target, last observed coordinates and average height of target. In this instance, the setup for the video processing and tracking algorithm were done on a single computing platform and the data is transferred from one tracking program to the next program. Table 8.1 lists the input video settings used in evaluating the tracking methodology.

Input video	Camera 1		Uneven Background		Noisy Environment	
	Time (s)	Number of Frames	Time (s)	Number of Frames	Time (s)	Number of Frames
whole_seq3_1	10.20	306	7.33	220	14.03	421
whole_seq3_2	9.50	285	8.76	263	11.50	345

Table 8.1: Summary of Input Video Specifications for outdoor tracking

8.3.1 Segmented Parameter Tracking

The results from the segmented parametric tracking across an array of cameras can be seen in the following table that summarizes the tracking performance across 3 cameras in a multi-camera surveillance system for each scenario.

Input Video Sequence	3 segment			4 segment			5 segment			6 segment		
	cam1	cam2	cam3	cam1	cam2	cam3	cam1	cam2	cam3	cam1	cam2	cam3
whole_seq3_1				0.92	0.90	0.68						
whole_seq3_2	0.82	0.23	0.34	0.81	0.22	0.28				0.93	0.24	0.53

Table 8.2: TDR rate for Scenario 1

Input Video Sequence	3 segment			4 segment			5 segment			6 segment		
	cam1	cam2	cam3	cam1	cam2	cam3	cam1	cam2	cam3	cam1	cam2	cam3
whole_seq3_1				0.33	0.67	0.00						
whole_seq3_2				0.06	0.23	0.00						

Table 8.3: TDR rate for Scenario 2

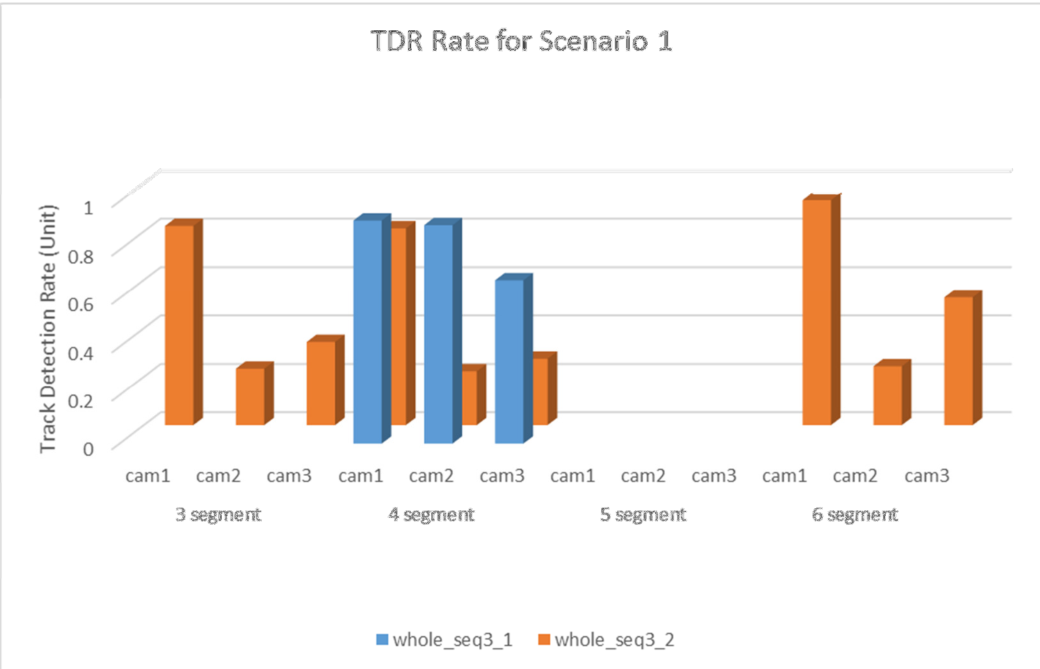
Input Video Sequence	3 segment			4 segment			5 segment			6 segment		
	cam1	cam2	cam3	cam1	cam2	cam3	cam1	cam2	cam3	cam1	cam2	cam3
whole_seq3_1				0.39	0.41	0.44						
whole_seq3_2				0.12	0.14	0.24						

Table 8.4: TDR rate for Scenario 3

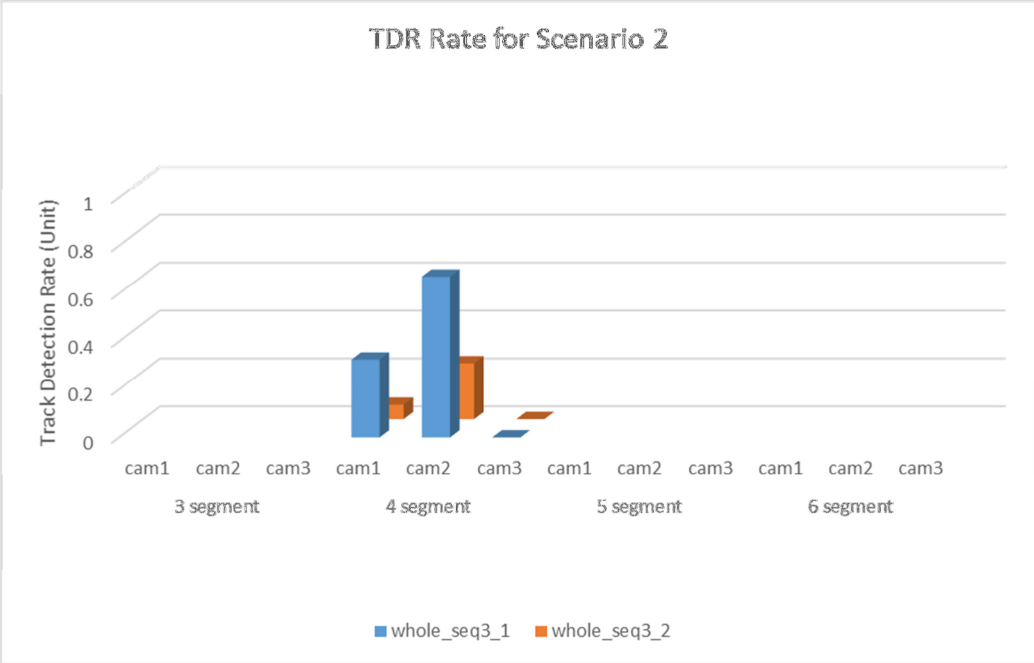
Input Video Sequence	3 segment			4 segment			5 segment			6 segment		
	cam1	cam2	cam3	cam1	cam2	cam3	cam1	cam2	cam3	cam1	cam2	cam3
whole_seq3_1				0.28	0.39	0.36						
whole_seq3_2				0.40	0.13	0.16						

Table 8.5: TDR rate for Scenario 4

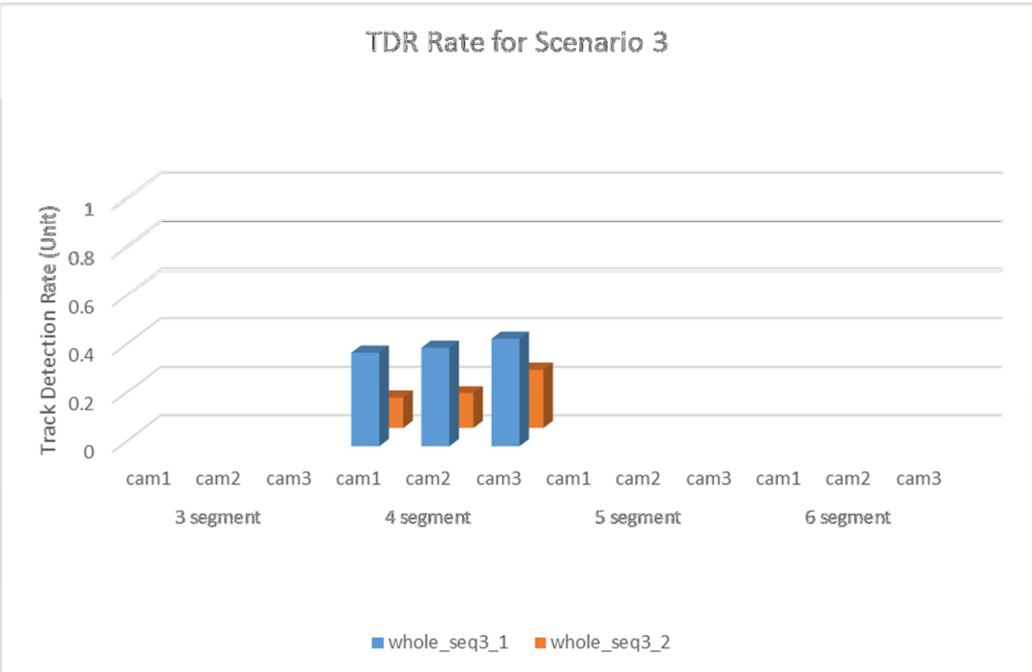
The data in Tables 8.2, 8.3, 8.4 and 8.5 can be illustrated in Figure 8.8.



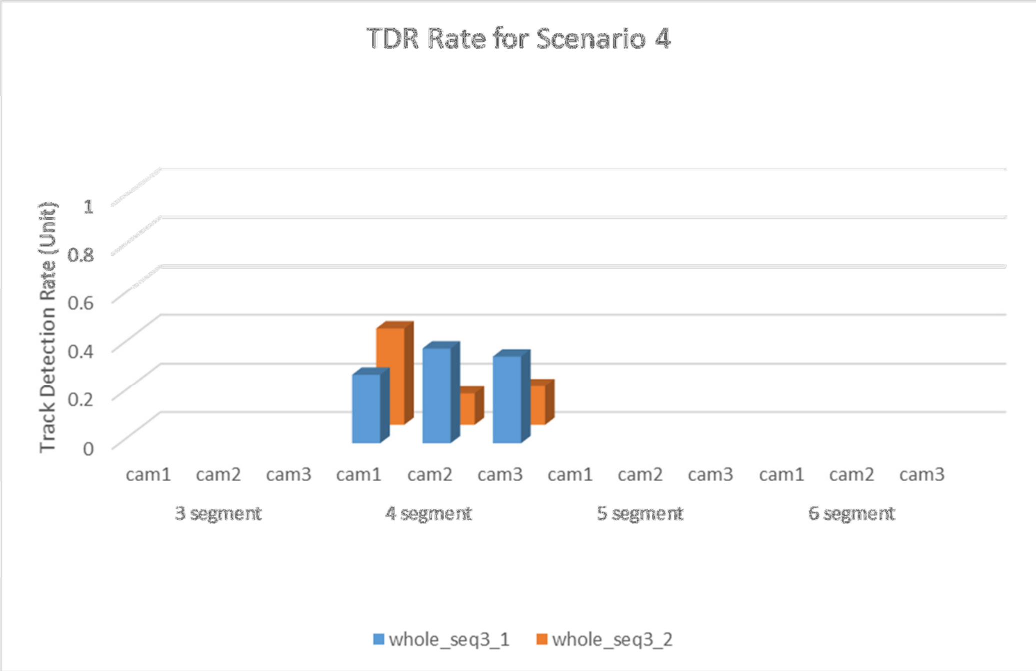
a) TDR Rate for Scenario 1



b) TDR Rate for Scenario 2



c) TDR Rate for Scenario 3



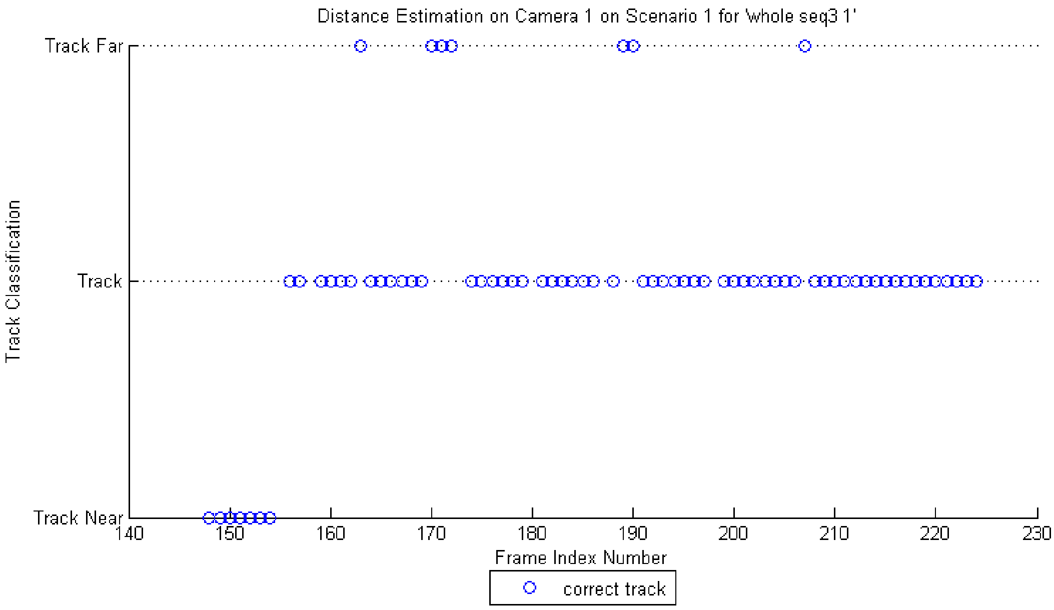
d) TDR Rate for Scenario 4

Figure 8.8: TDR Rates across different Scenarios

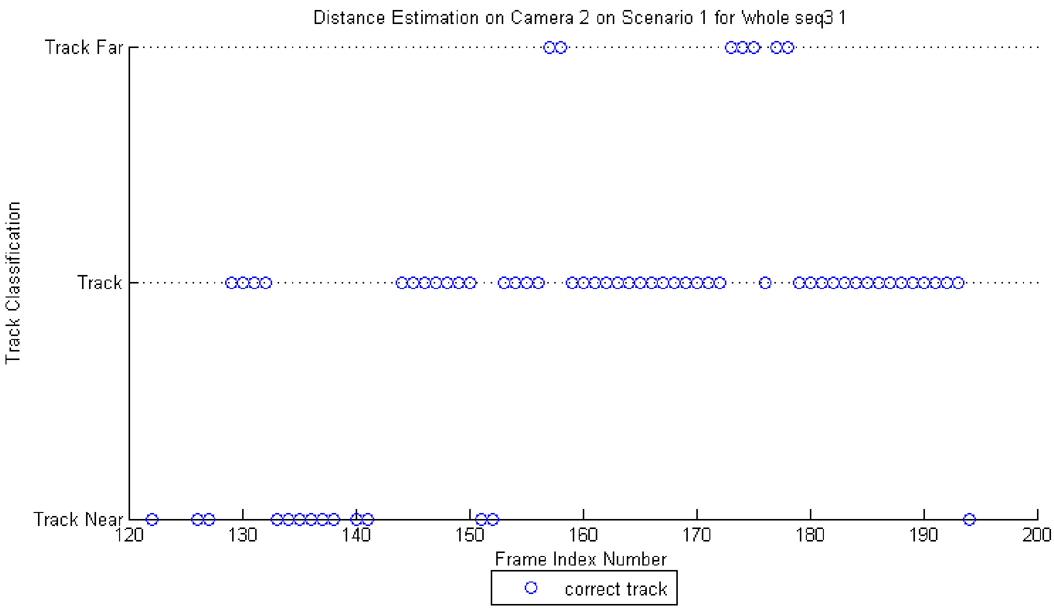
Based on the data shown in the tables and in the graphs, it has been shown that tracking of targets across a multi-camera system is possible, based on the parameters that exist between top and bottom colour regions. In this instance, when the colour image is segmented into 4 colour regions it provides consistent performance across a multi-camera system for tracking. This finding further confirms the results obtained in Chapter 7, section 7.4 that had targets tracked across an array of cameras based on parameters between segmented regions.

8.3.2 Segmented Parameter Tracking with Distance Estimation

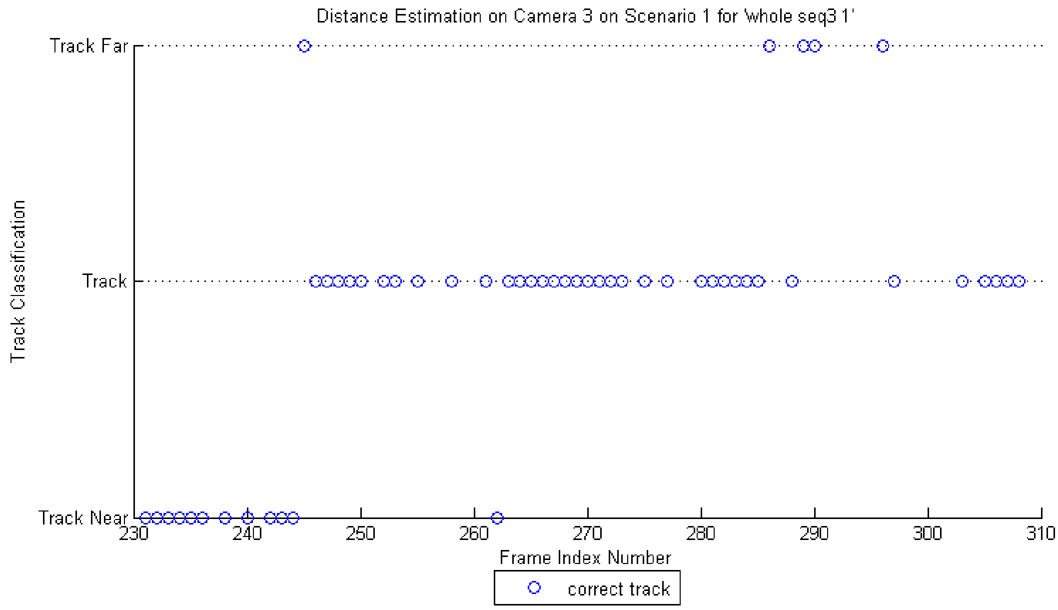
Apart from using parameters between segmented regions, the tracking of targets can be complemented with distance estimation of targets by using the general blob parameters. The blob parameters to be used are the height and area of the blobs. The blob parameters can be used as a means for estimating if the blob is closer to or further away from a camera as compared to a reference camera. An example of the tracking with estimated distance for Scenario 1 is as follows:



(a) Camera 1



(b) Camera 2

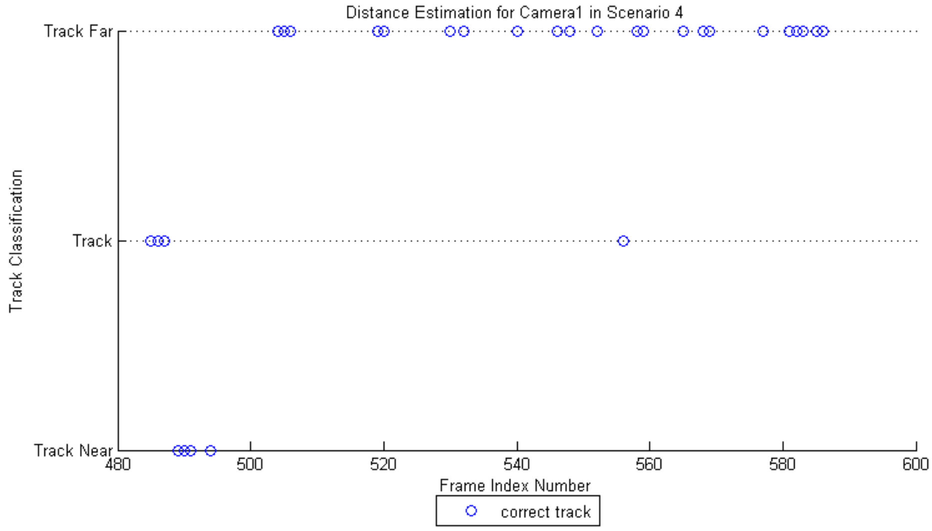


(c) Camera 3

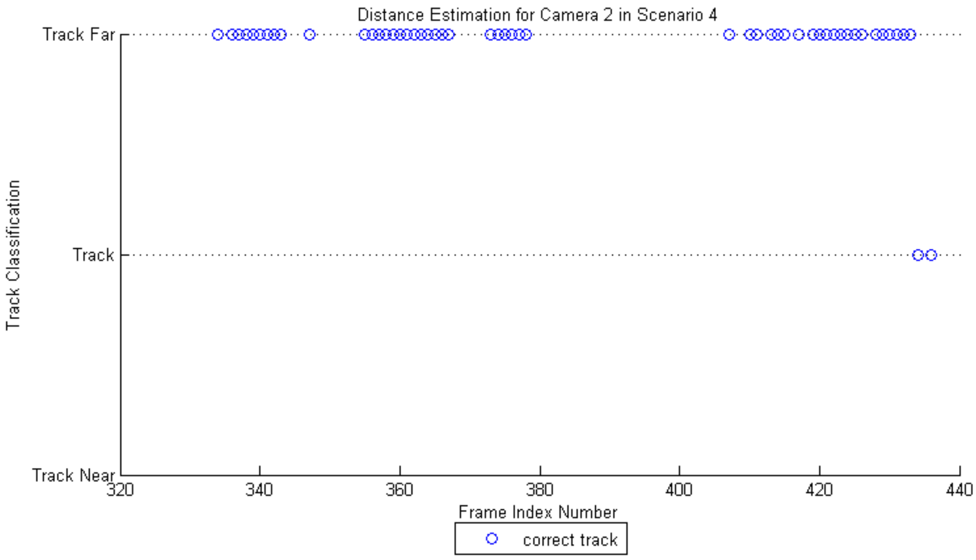
Figure 8.9: Distance Estimation on Three Separate Cameras in Scenario 1

Figure 8.9 shows that the tracking of target with the distance estimation of the target being labelled as either 'Track Near' or 'Track Far' as the indication that the tracked target is being nearer to or further from the camera. In Scenario 1, the target being tracked is at the same distance from the camera across the three cameras. This can be seen in Figure 8.9 where each successful track is indicated as a circle. For the tracking done in Scenario 1, where the targets are at the same distance from the camera, the majority of the tracked points are plotted on the 'Track' category with minimal points plotted on the 'Track Far' and 'Track Near' categories. For some of the points that are plotted on the 'Track Far' or 'Track Near' category, these are due to incorrect segmentation where some of the targets would seem to be taller or shorter as compared to the reference height information passed across cameras. Incorrect segmentation from background are the cause for the incorrect height measurement where the initial tracks were due to having shadows that were classified as foreground objects. The shadows happened to be formed due to changes in the background brightness added with reflection from the floor surface had created an object that was different from the background.

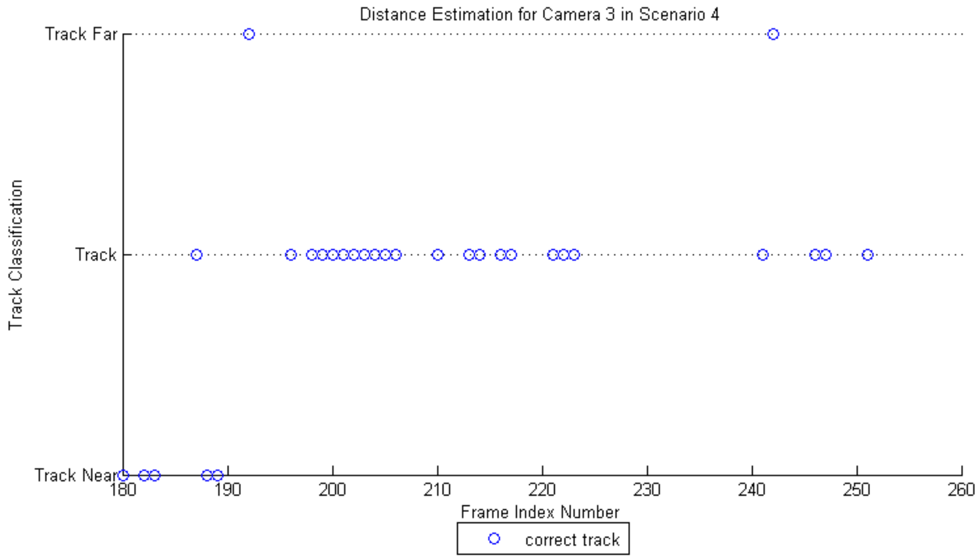
In Scenario 4, distance estimation had determined that the target of interest had been further away from the camera in the view field of Camera 2. It also showed that the target of interest was tracked to be at the same distance in Camera 1 and Camera 3. The distance estimation in Scenario 4 is seen in Figure 8.10.



(a) Camera 1



(b) Camera 2



(c) Camera 3

Figure 8.10: Distance Estimation on Three Cameras in Scenario 4

Based on the tracking data illustrated in Figure 8.10, the target of interest is categorized to be at the same distance from cameras in the view field of Camera 1 and Camera 2 based on measured height. The assumption that the targets of interest are at the same distance from Camera 1 and 2 are that the majority of correct tracking points in the view of Camera 1 and Camera 2 are on the ‘Track Far’ category. The ‘Track Far’ category was verified by checking the videos captured on Camera 1 and Camera 2 where targets moving from left to right have taken a step further away from the camera and targets moving from right to left have taken a step closer to the cameras. For scenario 4 as seen in Figure 8.7, Camera 2 is the camera view field that has two targets moving in two opposite directions. The target of interest moves across the camera view field from left to right. In Camera 2 view field, the target of interest has already moved one step further away from the camera while the occluding target or object had moved one step closer to the camera where these were steps taken to avoid a collision. At the point of occlusion, the target of interest is occluded by the second target and thus the target of interest is not detected and that is the point where there is a break in target tracking. In Figure 8.10 (b), it can be observed that there is a break in the tracking action from frame 390 to frame 410. The break in the tracking action, which is verified by manual inspection of the video sequence, that the tracked target is temporarily occluded or masked by another target within the camera view field. Based on the tracking operation illustrated in Figure 8.10(b), tracking was able to be continued after the occlusion. The tracking done indicates that using the height and area can help in determining if a target is nearer or further to a camera.

8.3.3 Segmented Parameter Tracking with Spatial Estimation

Taking the findings from section 8.3.2, we can approximate if the target of interest is nearer to or further from the camera as the target moves across the view field of an array of cameras. To further add spatial information or spatial awareness, the distance between cameras can be estimated by determining the time that has elapsed between the last detected image of target on the first camera and first detected target image on the following camera against target speed. The following charts provide a sample of the horizontal location of a target as it moves across a multi-camera system.

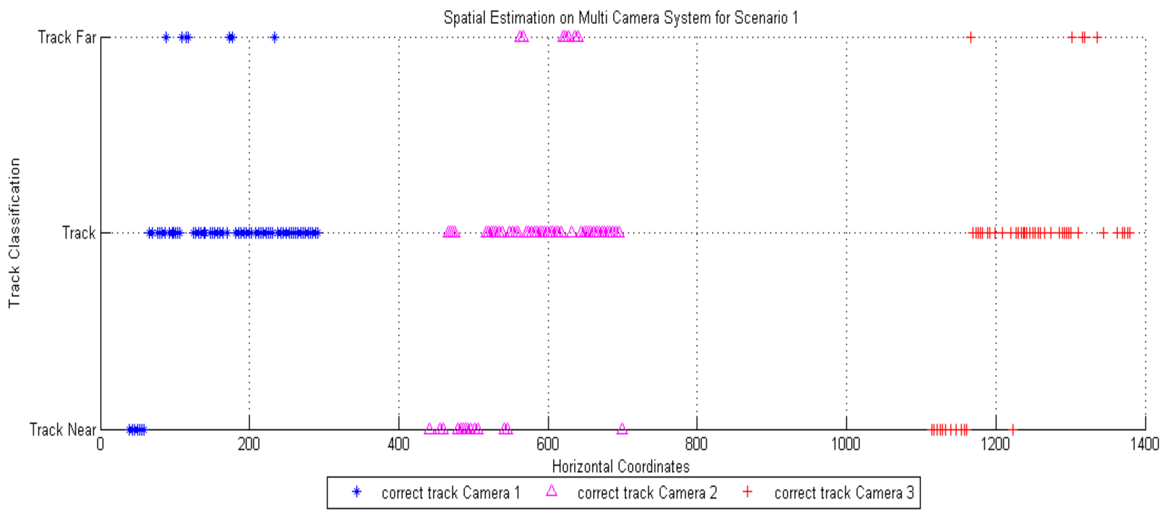


Figure 8.11: Location Estimation of Tracked Target across Three Cameras in Scenario 1

Figure 8.11 shows that the tracking of the target of interest as it moves along the view field of the array cameras. The data illustrated in Figure 8.11 is the compilation of the tracking data as the target is tracked from one camera to another while data is transferred from one camera to another. Camera 1 tracks targets from horizontal pixel location 0 to 300 and once the target is out of Camera 1 view it passes data to Camera 2 to look and continue tracking of the target of interest. Camera 2 tracks targets within its view field and estimates the horizontal location to be between 400 to 700 pixels based on the data received from Camera 1 which are the GMM colour region parameters, speed and measured height of targets. Camera 2 would pass the same data to the Camera 3 for it to continue tracking which are estimated to be from horizontal pixel location 1100 to 1400. The gaps between tracked points show that the position of the cameras have gaps or spacing between them and that the camera views are not overlapping. The gap in the tracked points are the estimated distance based on time and average speed of target as data is transferred from one camera to another camera. In addition to the gap in the estimated location of tracks, another data being passed from camera to camera is used in the categorization of the estimated distance of target from camera that is based on measured height of target. Resulting tracked positions were confirmed against the experimental set up and observation of captured videos. The tracked position was compared against the ground truth or reference position that had been observed and determined as stated in Section 4.2.1. The tracking in other scenarios can also be seen in the following figures for Scenario 3 and Scenario 4.

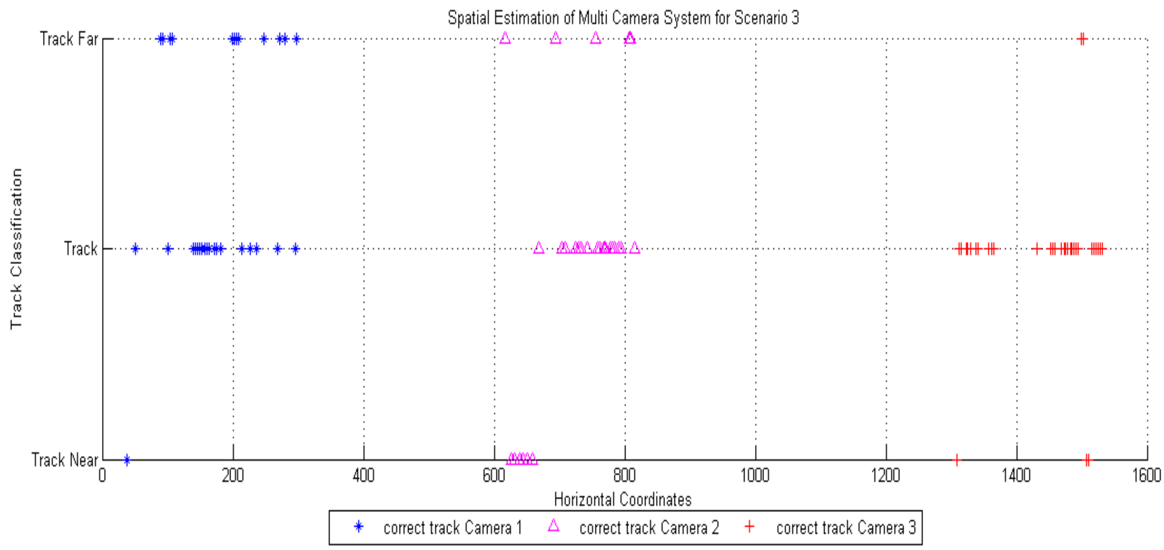


Figure 8.12: Location Estimation of Tracked Target across Three Cameras in Scenario 3

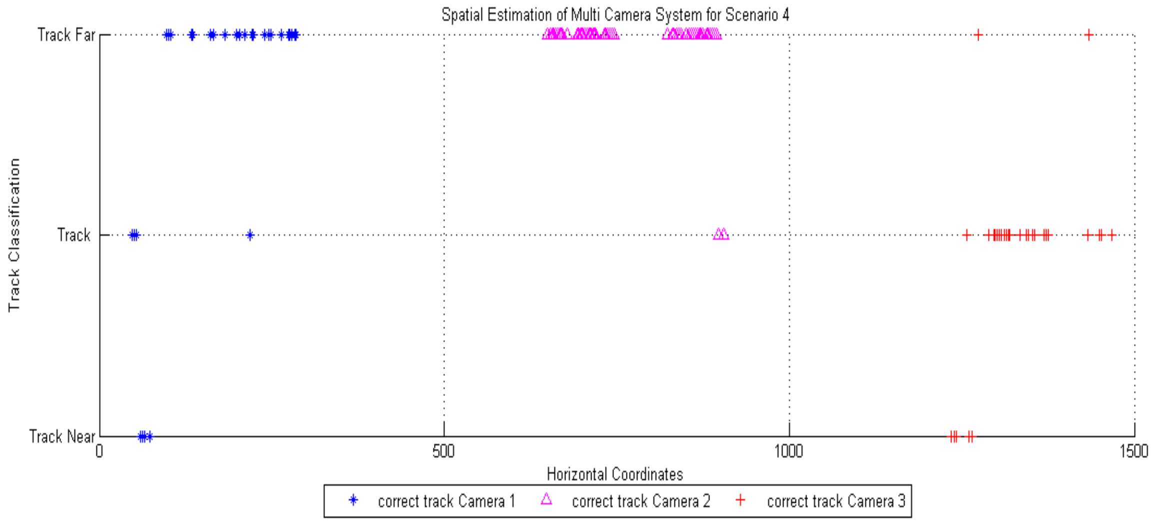


Figure 8.13: Location Estimation of Tracked Target across Three Cameras in Scenario 4

Based on Figures 8.12 and 8.13, the tracking of a target as it moves across the view field of a multi-camera system is able to estimate the location and estimate or approximate the relative distance of the target from the camera.

8.4. Summary

By looking at the results from the different experiments and analysis of the data in this chapter, it can be seen that the tracking performance of the method using the parameters between different colour segments was repeated and the results showed that tracking of the target across multiple cameras was able to be done. Tracking a target of interest across an array of cameras is based on a set of parameters which are the vector magnitude and vector angle between the top and bottom colour regions together with colour difference between the same colour regions being passed from camera to camera. In addition to the parameters used for determining that the target of interest is being tracked, another set of parameters are used for spatial estimation of target distance between cameras as the target moves across camera view fields and estimated distance of target to camera where the target is classified as being either closer or further. In addition to tracking of the target across multiple cameras, additional information can be extracted to determine the relative position between cameras and the approximate distance of the target from the camera. The investigation done in this chapter was done with some of the blob information such as blob height, area and speed to determine the relative positions and the depth or distance of the targets from cameras. The information of distance of the targets from the cameras and relative position of targets as they move across the view fields of a multi-camera system is able to generate a map of the target as it moves across view fields of a multi-camera surveillance system. The horizontal position information and the distance estimation of the target from the cameras gives this system a basic spatial awareness where the position of targets of interest can be estimated. This type of information is important where a model or movement of a target can be obtained by tracking a specific target as it moves across an array of cameras and also that the estimated spatial location of targets of interest could be estimated to have more robust tracking and determination of location of targets of interest.

CHAPTER 9

Discussion and Conclusions

9.1. Introduction

In the introduction to this thesis the problems associated with tracking a moving target especially a human target moving across a camera view were discussed. Tracking of targets in the various colour spaces was also discussed. Experiments were conducted to investigate the minimum information required for tracking targets. The minimal tracking information was then used in tracking a target across an array of cameras.

9.2. Colour Space Factor

Chapter 2 discusses the various differences between a number of colour spaces of potential application to tracking. Each colour space has different information stored on different layers. The utilization of different colour spaces were discussed in the different implementations of tracking a target as it moves across a camera view field. The utilization of different colour spaces did give different tracking performances as discussed in Chapter 3 and 4. The main findings concerning colour space as a factor in tracking targets showed that illumination or brightness information was not the parameter to be used as the data to track a target. Although initial tracking work was done in grayscale and RGB colour space, these colour spaces have the brightness information embedded together with colour information, especially the RGB colour space. Therefore, the colour spaces of interest in this investigation were the colour spaces that had colour information separated from the brightness or illumination data, which in this instance was the YCbCr and HSV colour spaces. The development of tracking methods were to take advantage or utilize the information stored in the YCbCr or HSV colour space which were the colour information itself without the illumination information. The work done in this research had initially used both HSV and YCbCr colour spaces for tracking of targets. Initial work done had managed to use specific colour layers for tracking purposes such as HS colour layers from the HSV colour space and CbCr colour layers from the YCbCr colour spaces based on the results in Chapter 4.

9.3. Video Tracking Methodologies

The discussion on video tracking methodologies was reported in Chapter 4. Video tracking methodologies were developed based on different features and parameters. Tracking methods that were found to be used in tracking targets ranged from window tracking based methods, feature based tracking and colour based tracking. The window tracking method utilizes a reference image as the point or region to be detected within the input image. Feature based tracking uses specific features detected within an image for tracking purposes. Colour based tracking uses the colour properties of the input image as the basis for tracking a target. The work discussed in this section described the different tracking methods developed based on the different colour spaces used. The colour space that was typically used in window and feature based tracking was the grayscale colour space. In the colour based methods, the parameters used in tracking a target were based on the colour information, specific colour region sizes, location and vector.

From video tracking methodologies that have been used in different papers, colour information for the target of interest is the main information to be used for tracking and a number of methods were tested in the grayscale colour space. Therefore, YCbCr and HSV colour spaces were chosen due to the minimal work that had been done in those colour spaces. Among the methods used for tracking objects was the utilization of colour information of the target by comparing the colour histogram against a reference histogram for tracking a target. Histogram comparison was chosen as the method to be tested due to the perceived idea that comparing colour histograms would provide a robust tracking method.

9.4. Video Tracking Metrics

Chapter four discusses the methods utilized in measuring video tracking performances. Metrics are an essential requirement for measuring the performance of any new tracking method. The development of video tracking metrics moved together with the development of the various video tracking methods. This was due to the need of proving the viability or advantages of the newly developed tracking methods. The metrics range from contingency tables to qualitative to quantitative measures. From the different performance metrics, one common point was the utilization of a reference point called ground truth. Ground truth points are points, features or objects which are identified as reference points or reference objects. The metrics developed from different sources, such as ETISEO and Ellis, have relevant similar metrics to show the tracking performance of any algorithm such as the Precision and Tracking Detection Rate which gave the same result in testing with reference videos from the ETISEO group. The similarity in some the metrics from different groups and researchers indicate that the understanding and method of measuring tracking are coming to a point of convergence where metrics are reporting the same parameters being measured.

In addition to the similarity of metrics, another metric was developed for measuring the consistency of tracking a target. This measure is a metric that supplements or assists in the metric that measures the detection rate of a target in a video. A majority of the metrics that have been developed and used have been utilized to determine the performance of a tracking algorithm in tracking the target of interest. The consistency metric measures the consistency of tracking a target where the tracking of a target is supposed to be continuous from one frame to the next. Currently there are no properly defined metrics to measure the tracking consistency. As such there is a need for the consistency metric to be developed to supplement the reporting of tracking performance.

9.5. Tracking Methods

9.5.1. Correlation Tracking

Chapter 5 reports on an investigation into the utilization of normalized cross correlation in tracking a target. The normalized cross correlation method is dependent on the image used as reference for detecting the reference image within the input video frame. The cross correlation method was tested on two different types of image namely the face and a whole person. The normalized cross correlation tracking was done on a single camera setup first, and the results analysed before testing on a multiple camera setup. The initial results for normalized cross correlation tracking indicated that tracking a face showed favourable results. The face tracking also indicated that the colour spaces of YCbCr and HSV had better tracking performances compared to RGB and grayscale colour spaces. RGB and grayscale colour spaces are affected by brightness levels and this makes tracking less robust as small changes in illumination would affect tracking performance. This is in contrast to YCbCr and HSV colour spaces that have the brightness information on a separate layer whereas the colour information is placed on different colour layers. The separation of the colour information makes tracking of the face more robust as the skin colour is within a narrow colour range. Tracking in the different colour spaces also indicated that tracking of targets could be done on colour layers alone without the illumination or brightness information together.

The results from the facial tracking tests gave the drive to continue the normalized cross correlation tracking to track a whole person. The results in tracking a whole person using normalized cross correlation however did not yield the high tracking performances seen in tracking a face. The results from tracking of a whole person in the different colour spaces indicated that the tracking performance was contrary to the results obtained in tracking a face. The initial results indicated that the grayscale and RGB colour spaces had better tracking performances compared to the YCbCr and HSV colour spaces. This is the opposite of the results when tracking a face. This is probably due to the number of colour regions available or used in tracking the different targets whereas in face tracking there is only one major region for tracking. When dealing with a whole person there are many colour regions to be used for tracking. In addition to the multiple colour regions used in the correlation tracking, the tracking of a whole person also would have to deal with a target that is moving which would have the input image constantly changing its shape which could lead to inaccurate tracking. The inaccurate tracking is due to the method of correlation tracking that uses a reference image for detecting the target image. From the observations of utilizing correlation as a method of tracking a target, it can be seen that tracking for a single colour region target has a better tracking performance as compared to a multiple colour region target.

9.5.2. Histogram Tracking

In order to address the problems encountered when tracking a target with multiple colour regions using normalized cross correlation, a histogram comparison was investigated. The method described in Chapter 6 uses the distribution of pixels of each target as a means for detecting and tracking a target as the target moves across the camera view field. The utilization of the pixel distribution generates a histogram that encompasses the different colour regions making up the overall person. The hypothesis tested here was that the generated histogram for each target would be unique enough to be used as a reference for detecting and tracking that target. Based on results obtained, the utilization of the colour histogram for tracking had enough unique capability for tracking a single colour target but for a multiple colour target histogram comparison could not provide robust tracking performance. It has to be mentioned here that as a target moves across a camera view field, the pixel histogram of the target will change accordingly due to some of the limbs being obscured.

The utilization of a reference histogram is the main question in this tracking method. The reference histogram was obtained via three measures which are the average histogram, the composite image histogram and the segmented regions average histogram. The average histogram methods use the average histogram value as the target moves across the camera view field to generate the reference histogram. The composite image histogram uses all images as the target moves across the camera view field to generate a composite image and then obtain the histogram from the composite image.

One outcome that can be used from the investigation of histogram tracking is the performance difference between colour spaces. The results show that the HSV and YCbCr colour spaces had better tracking performances compared to grayscale. This is probably due to the different colour information stored on the layers compared to the illumination information stored on the grayscale colour space. In further comparing the HSV and YCbCr colour space, it could be seen that the YCbCr colour space had better tracking performance in terms of correct tracking and lower incorrect tracking. It was observed that tracking of targets using the histogram of pixel distribution could be used as a means for tracking but the overall tracking performance was seen to be on the lower side. This was in comparison to the segmented histogram tracking method. The results from the segmented histogram tracking method did show that tracking performance was significantly higher but at the same time it also showed higher incorrect tracking rates. This was to be expected where there would be segments from other targets that would have similar or statistically similar histograms. The correct tracking rates were observed to be significantly higher compared to the incorrect tracking rates for the different methods of obtaining the reference histogram.

9.5.3. Parametric Tracking of Multiple Segmented Regions

In Chapter 7, the investigation continued on the tracking method of a target based on the parameters that can extracted after the target is segmented into multiple colour regions. The initial target is segmented into the different colour regions using the EMM method of clustering pixels according to estimated parameters of normal distributions. The segmenting of the input image breaks the image into groups or clusters of similar colour regions. From the colour groups that are generated, a set of parameters can be generated between the different colour regions. The parameters used in the tracking of a target were the vector magnitude, vector angle and colour difference between the regions of interest. These parameters are recorded as a target moves across the camera view field and generates a cluster or multiple clusters which should be unique to each target. These data point clusters are then analysed with EMM to obtain parameters representing the data points within the cluster.

The parameters obtained in this section are between the highest colour region to the lowest region. The regions used were the head and lower torso as these are the regions that can be deemed to be more consistent or robust to movement [Chang and Gong, 2001, 19-26; Lu, Wang et al., 2004, 188 - 191; Siebel and Maybank, 2002, 373-387; Yan and Forsyth, 2005, 370-377]. There are other regions in between the head and lower torso that could be used for tracking but were not used in the current investigation of tracking across a set of cameras. The usage of the regions such as the arms and legs would not be feasible as compared to the torso as the segmentation and detection rates of the arms and legs are lower [Ramanan, A. Forsyth et al., 2007, 65-81]. The arms and legs are not used because of the movement of the limbs makes segmentation inaccurate. One of the main points that can be inferred from the tracking results using the parameters between selected regions is that the tracking of a target can be done across cameras but there is still a rate of incorrect tracking. However, the incorrect tracking rate is always lower compared to the correct tracking rate for the images tested, when using the optimum set of parameters. The optimum set of parameters in this situation is based on an experimental result that evaluates the number of colour regions that a target can be segmented into. The number of colour regions are decided by having the minimum number of regions, non-overlapping mean value of each normal distribution and minimum variance for each normal variance. The utilization of statistical testing to determine correct tracking of a target is dependent on the unique set of parameters that can be extracted from the different clusters of points from each target. Using statistical testing as a means for testing will always have a possibility of having incorrect tracking. This method could be further tested or expanded to include additional regions and points in between the top and bottom regions as additional parameters for tracking. By using other colour regions between the top and bottom colour regions of interest, additional parameters between different colour regions, from the top to the bottom colour region, could be used to generate a set of parameters that could be used for accurate and robust tracking.

9.6. Contributions

This section summarises the contributions made by this thesis. The contributions are classified into main and subsidiary contributions based on the work carried out in this thesis and contributions based on previous work. The references in parenthesis refer to the corresponding sections in the thesis.

9.6.1. Main Contributions

i. Parametric tracking method used together with blob parameters to have spatial awareness to determine the relative position of a target of interest as the target moves across the view field of multiple cameras. (Chapter 8).

ii. A parametric tracking method based on multiple segmented regions was proposed and shown to function reliably across multiple cameras (Chapter 7).

iii. A metric for measuring the consistency of tracking a target was developed. (Chapter 4).

iv. The relative performance of different algorithms for detecting and tracking a face in different colour spaces has been established from experimental work. (Section 5.3.2).

9.6.2. Subsidiary Contributions

i. It has been shown from experimental work that using histogram measurement is a potential method of detecting and tracking a target. (Section 6.3).

ii. It has been established that colour spaces such as HSV and YCbCr have better tracking performance compared to grayscale colour space. (Chapter 6).

iii. It has been shown from experimental work that the utilization of EMM can be used as a method of segmenting an image into a group of multiple colour regions (Section 7.2).

iv. It has been shown from experimental work that parameters can be extracted from clusters of data points for use of tracking a target on a single camera (Section 7.3).

v. It has been shown, with the aid of experimental work, that the parameters extracted using EMM are able to track the same target across three cameras with backgrounds ranging from uniform, to uneven and finally cluttered (Section 7.4).

9.7. Future Work

Based on the work reported in this thesis, this section puts forward some ideas for future work.

9.7.1. Multiple regions and multiple parameters

As discussed in section 8.4, the implementation of statistical analysis and hypothesis testing on the parameters between top and bottom colour regions indicate that the chosen colour regions did have enough unique parameters for the purpose of video tracking. Based on the initial results of tracking a target from the parameters of two regions, there should be further investigation on the tracking of a target using multiple regions. The multiple colour regions that make up a person could provide enough parameters between colour regions to give each target a unique signature that would enable a target to be tracked across a number of cameras. Therefore, an extensive set of tests or a method of statistical analysis would be needed to investigate the type of connection between the different colour regions to generate the necessary parameters or necessary connection to be a unique fingerprint to track targets across different cameras.

9.7.2. Intelligent Network Application for Parametric Tracking

As seen in the discussion in section 8.4, there is a need to investigate the implementation or utilization of a neural network or intelligent network system to extract the necessary parameters for tracking. The current method of determining the best set of parameters would be to check or utilize all possible combinations of segmentation to be used for tracking a target. This method is a time consuming and a resource consuming method of determining the best set of parameters for tracking a target across an array of cameras.

In conclusion, this thesis demonstrates the viability of using parameters between colour regions that represent a person. Colour region segmenting is an important step that is determined by the type of colour space that is used which is one of the findings of this thesis. Resulting from colour space investigations, it was demonstrated that a whole person can be represented or is made up of a number of colour regions. Based on the collection of colour regions representing a person, parameters between the colour regions were used as a means of tracking a person. Parametric tracking between colour regions was demonstrated by tracking targets across an array of cameras.

References

Abbas, 2008	Abbas, O. A. (2008). "*Comparison Between Data Clustering Algorithms*." The International Arab Journal of Information Technology **5**(3).
Adebesi and Olusayo, 2012	Adebisi, A. A., O. E. Olusayo, et al. (2012). "*An Exploratory Study of K-Means and Expectation Maximization Algorithms*." British Journal of Mathematics and Computer Science **2**(2): 62-71
Ali and Indupalli, 2006	Ali, M. A., S. Indupalli, et al. (2006). *Tracking Multiple People for Video Surveillance*. First International Workshop on Video Processing for Security, Quebec City , Canada.
Alldrin and Smith, 2003	Alldrin, N., A. Smith, et al. (2003). *Clustering with EM and K-Means*. San Diego, University of California
Amiri and Rabiee, 2003	Amiri, M., H. R. Rabiee, et al. (2003). *A New Wavelet Domain Block Matching Algorithm for Real-Time Object Tracking*. International Conference on Image Processing 2003 (ICIP 2003)
Atsushi et al, 2002	Aziz, N. N. A., Y. M. Mustafah, et al. (2014). *Features Selection for Multi-camera Tracking*. Computer and Communication Engineering (ICCCE), 2014 International Conference on, Kuala Lumpur
Aziz, N. N. A., Y. M. Mustafah, et al., 2014	Atsushi, N., K. Hirokazu, et al. (2002). *Tracking multiple people using distributed vision systems*. IEEE International Conference on Robotics and Automation, 2002., Washington, DC, USA
Bakowski and Jones, 1999	Bakowski, A. and G. A. Jones (1999). *Video Surveillance Tracking Using Colour Region Adjacency Graphs*. Image Processing and its Applications
Bashir and Porikli, 2006	Bashir, F. and F. Porikli (2006). *Performance Evaluation of Object Detection and Tracking Systems*. Ninth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2006), New York, USA.
Beetz et al, 2006	Beetz, M., B. Radig, et al. (2006). *A Person and Context Specific Approach for Skin Color Classification*. 18th International Conference on Pattern Recognition, 2006 (ICPR 2006), Hong Kong
Behera, R. K., P. Kharade, et al. , 2012	Behera, R. K., P. Kharade, et al. (2012). *Multi-camera based surveillance system*. Information and Communication Technologies (WICT), 2012 World Congress on, Trivandrum.

Bird et al, 2005	Bird, N. D., O. Masoud, et al. (2005). *Detection of loitering individuals in public transportation areas*. IEEE Transactions on Intelligent Transportation Systems
Black et al, 2003	Black, J., T. Ellis, et al. (2003). *A Novel Method for Video Tracking Performance Evaluation*. Joint IEEE Intl. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, Nice, France
Bodor et al, 2003	Bodor, R., B. Jackson, et al. (2003). *Vision-Based Human Tracking and Activity Recognition*. 11th Mediterranean Conference on Control and Automation
Bredereck, M., X. Jiang, et al. ,2012	Bredereck, M., X. Jiang, et al. (2012). *Data association for multi-object Tracking-by-Detection in multi-camera networks*. Distributed Smart Cameras (ICDSC), 2012 Sixth International Conference on, Hong Kong.
Cai and Aggarwal et al, 1996	Cai, Q. and J. K. Aggarwal (1996). *Tracking human motion using multiple cameras*. 13th International Conference on Pattern Recognition, 1996, Vienna, Austria.
Cai and Aggarwal et al, 1998	Cai, Q. and J. K. Aggarwal (1998). *Automatic tracking of human motion in indoor scenes across multiple synchronized video streams*. Sixth International Conference on Computer Vision 1998, Bombay, India.
Cai and Aggarwal et al, 1999	Cai, Q. and J. K. Aggarwal (1999). "*Tracking human motion in structured environments using a distributed camera system*." IEEE Transactions on Pattern Analysis and Machine Intelligence **21**(11): 1241-1247
Castleman, 1996	Castleman, K. R. (1996). *Digital Image Processing*, Prentice Hall.
Chang and Gong , 2001	Chang, T. H. and S. Gong (2001). *Tracking multiple people with a multi-camera system*. Proceedings 2001 IEEE Workshop on Multi-Object Tracking, Vancouver, BC, Canada
Chen, C.-H., T.-Y. Chen, et al. , 2011	Chen, C.-H., T.-Y. Chen, et al. (2011). *People Tracking in the Multi-camera Surveillance System*. Innovations in Bio-inspired Computing and Applications (IBICA), 2011 Second International Conference on, Shenzen.
Chen, W., L. Cao, et al. , 2014	Chen, W., L. Cao, et al. (2014). *A novel solution for multi-camera object tracking*. Image Processing (ICIP), 2014 IEEE International Conference on, Paris.
Chen, X. and L. An, 2015	Chen, X., L. An, et al. (2015). *Multi-Target Tracking in Non-overlapping Cameras Using a Reference Set*. Sensors Journal, IEEE **PP**(99): 1.

Chen et al , 2005	Chen, Y.-J., Y.-C. Li, et al. (2005). *The Implementation of a Stand-alone Video Tracking and Analysis System for Animal Behavior Measurement in Morris Water Maze*. 27th Annual International Conference on Engineering in Medicine and Biology Society 2005 (IEEE-EMBS 2005), Shanghai
Chippendale, 2006	Chippendale, P. (2006). *Towards Automatic Body Language Annotation*. 7th International Conference on Automatic Face and Gesture Recognition 2006 (FGR 2006)
Choi and Yoo, 2013	Choi, J.-W. and J.-H. Yoo (2013). *Real-time multi-person tracking in fixed surveillance camera environment.* Consumer Electronics (ICCE), 2013 IEEE International Conference on. Las Vegas, NV, IEEE**:** 125-126.
Dahmane and Meunier, 2005	Dahmane, M. and J. Meunier (2005). *Real-time video surveillance with self-organizing maps*. The 2nd Canadian Conference on Computer and Robot Vision, 2005, Montreal, Canada
Dargazany, A. and M. Nicolescu, 2012	Dargazany, A. and M. Nicolescu (2012). *Human Body Parts Tracking Using Torso Tracking: Applications to Activity Recognition. Information Technology: New Generations (ITNG), 2012 Ninth International Conference on*. Las Vegas, NV**:** 646-651.
Denman and Fookes, 2009	Denman, S., C. Fookes, et al. (2009). *Dynamic Performance Measures for Object Tracking Systems*. Advanced Video and Signal Based Surveillance, 2009. AVSS '09. Sixth IEEE International Conference on Genova
Diaz and Sucar, 2002	Diaz de Len, R. and L. E. Sucar (2002). *Continuous activity recognition with missing data*. 16th International Conference on Pattern Recognition 2002
Doermann and Mihalcik, 2000	Doermann, D. and D. NMihalcik (2000). *Tools and techniques for video performance evaluation*. 15th International Conference on Pattern Recognition 2000, Barcelona
Dutta and Sarma, 2014	Dutta, A. and K. K. Sarma (2014). *SAR image segmentation using wavelets and Gaussian mixture model.* Signal Processing and Integrated Networks (SPIN), 2014 International Conference on. Noida, IEEE**:** 466-770.
Eldib, M., N. B. Bo, et al., 2014	Eldib, M., N. B. Bo, et al. (2014). *A low resolution multi-camera system for person tracking*. Image Processing (ICIP), 2014 IEEE International Conference on, Paris.

Ellis, 2002	Ellis, T. (2002). *Performance Metrics and Methods for Tracking in Surveillance*. Third IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Copenhagen, Denmark
Ellis, 2000	Ellis, T. (2002). *Multi-camera Video surveillance*. 36th Annual 2002 International Carnahan Conference on Security Technology 2002.
Erdem et al, 2003	Erdem, C. E., B. Sankur, et al. (2003). *Performance Measures for Video Object Segmentation and Tracking*. Proceedings of the SPIE Visual Communications and Image Processing 2003
Ferman et al, 2002	Ferman, A. M., A. M. Tekalp, et al. (2002). "*Robust Color Histogram Descriptors for Video Segment Retrieval and Identification*." IEEE Transactions on Image Processing **11**(5)
Gonzales and Woods, 2002	Gonzalez, R. C. and R. E. Woods (2002). *Digital Image Processing*. New Jersey, Prentice Hall
Gonzales et al, 2004	Gonzalez, R. C., R. E. Woods, et al. (2004). *Digital Image Processing using Matlab*, Pearson Prentice Hall
Greenspan et al., 2002	Greenspan, H., S. Gordon, et al. (2002). *Probabilistic Models for Generating, Modelling and Matching Image Categories*. 16th International Conference on Pattern Recognition, IEEE Computer Society Washington, DC, USA
Han et al., 2004	Han, M., A. Sethi, et al. (2004). *A detection-based multiple object tracking method*. 2004 International Conference on Image Processing
Haritaoglu et al., 2002	Haritaoglu, I., D. Harwood, et al. (2000). "*W4: real-time surveillance of people and their activities*." IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(8): 809-830
Hassan, W., N. Bangalore, et al. (2011).	Hassan, W., N. Bangalore, et al. (2011). *Object tracking in a multi camera environment*. Signal and Image Processing Applications (ICSIPA), 2011 IEEE International Conference on, Kuala Lumpur.
Hidaka, K. , 2009	Hidaka, K. (2009). *Robust color tracking based on mean-shift under illuminance change*. ICCAS-SICE, 2009, Fukuoka
Hii et al., 2006	Hii, A., C. E. Hann, et al. (2006). "*Fast normalized cross correlation for motion tracking using basis functions*." Computer Methods and Programs in Biomedicine **82**: 144-156.

Hommel, S., M. A. Grimm, et al., 2012	Hommel, S., M. A. Grimm, et al. (2012). *An intelligent system architecture for multi-camera human tracking at airports*. Computational Intelligence and Informatics (CINTI), 2012 IEEE 13th International Symposium on, Budapest.
Horprasert et al., 1999	Horprasert, T., D. Harwood, et al. (1999). *A statistical Approach for Real-time Robust Background Subtraction and Shadow Detection*. IEEE ICCV '99 Frame Rate Workshop, Kerkyra, Greece.
Hsu, Yang, et. al., 2013	Hsu, H.-H., W.-M. Yang, et al. (2013, 1-8 Jan. 2013). *People tracking in a multi-camera environment.*Conference Anthology, IEEE Retrieved 1-8 Jan. 2013, 2013
Huang, Q., J. Yang, et al. , 2012	Huang, Q., J. Yang, et al. (2012). *Person re-identification across multi-camera system based on local descriptors*. Distributed Smart Cameras (ICDSC), 2012 Sixth International Conference on, Hong Kong.
Innria, 2006	Inrria, S.-. (2006). *Metrics Definition*, ETISEO
Ioochi and Bollees, 2005	Iocchi, L. and R. C. Bollees (2005). *Integrated Plan-View Tracking and Color-based person Models for Multiple People Tracking*. International Conference on Image Processing 2005.
Jack, 2007	Jack, K. (2007). *Video Demystified*, Newnes
Ju, Wang, et. al., 2014	Ju, Z., Y. Wang, et al. (2014). *A modified EM algorithm for hand gesture segmentation in RGB-D data*. Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on. Beijing, IEEE**:** 1736 - 1742.
KaewTraKulPong and Bowden, 2001	KaewTraKulPong, P. and R. Bowden (2001). *An Improved Adaptive Background Mixture Model for Realtime Tracking with Shadow Detection*. 2nd European Workshop on Advanced Video Based Surveillance Systems, Kluwer Academic Publishers
Kawato and Ohya, 2000	Kawato, S. and J. Ohya (2000). *Automatic skin-color distribution extraction for face detection and tracking*. 5th International Conference on Signal Processing Proceedings 2000 (WCCC-ICSP 2000), Beijing
Khan and Shah, 2003	Khan, S. and M. Shah (2003). "*Consistent labeling of tracked objects in multiple cameras with overlapping fields of view*." IEEE Transactions on Pattern Analysis and Machine Intelligence **25**(10): 1355-1360.
Kim et al., 2004	Kim, W.-S., D.-S. Cho, et al. (2004). *Interplane prediction for RGB video coding*. 2004 International Conference on Image Processing (ICIP '04)

Kitani et al, 2005	Kitani, K. M., Y. Sato, et al. (2005). *Deleted Interpolation Using a Hierarchical Bayesian Grammar Network for Recognizing Human Activity*. 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance 2005
Kobayashi et al, 1999	Kobayashi, M., M. Noma, et al. (1999). *Lossless compression for RGB color still images*. 1999 International Conference on Image Processing (ICIP99), Kobe, Japan
Komlen, D., T. Lombarovic, et al. , 2012	Komlen, D., T. Lombarovic, et al. (2012). *Multi-camera person tracking system*. MIPRO, 2012 Proceedings of the 35th International Convention, Opatija
Koschan et al, 2002	Koschan, A., S. K. Kang, et al. (2002). *Video Object Tracking Based on Extended Active Shape Models with Color Information*. 1st European Conference on Color in Graphics, Imaging, and Vision CGIV2002, Poitiers, France
Leo et al, 2004	Leo, M., T. D'Orazio, et al. (2004). *Complex human activity recognition for monitoring wide outdoor environments*. 17th International Conference on Pattern Recognition 2004
Lewis, 1995	Lewis, J. P. (1995). Fast Normalized Cross-Correlation. Vision Interface, Canadian Image Processing and Pattern Recognition Society**:** 120-123
Li et al, 2002	Li, J., C. S. Chua, et al. (2002). *Color based multiple people tracking*. 7th International Conference on Control, Automation, Robotics and Vision 2002. ICARCV 2002
Liang, D., Q. Huang, et al. ,2007	Liang, D., Q. Huang, et al. (2007). *Mean-Shift Blob Tracking with Adaptive Feature Selection and Scale Adaptation. Image Processing, 2007. ICIP 2007. IEEE International Conference on (Volume:3).* San Antonio, TX**:** III - 369 - III – 372
Lim et al, 2000	Lim, J., C. H.K., et al. (2000). *An Improved Video Object Tracking Algorithm Based on Motion Re-estimation*. 2000 International Conference on Image Processing, Vancouver , BC, Canada
Lu et al, 2004	Lu, H., Y. Wang, et al. (2004). *A crowd detecting and tracking algorithm based on projection of grayscale image*. Third International Conference on Image and Graphics, 2004
Mariano et al, 2002	Mariano, V. Y., J. Min, et al. (2002). *Performance evaluation of object detection algorithms*. 16th International Conference on Pattern Recognition 2002

Matsuzawa and Kumazawa, 2000	Matsuzawa, Y. and I. Kumazawa (2000). *Object tracking with shape representation network using color information*. 2000 International Conference on Image Processing, Vancouver, BC, Canada
McErlean 2006	McErlean, M. (2006). *Hierarchical Motion Estimation for Embedded Object Tracking*. 2006 IEEE International Symposium on Signal Processing and Information Technology
Mehmood, M. O., 2009	Mehmood, M. O. (2009). *Multi-camera based human tracking with non-Overlapping fields of view*. Application of Information and Communication Technologies, 2009. AICT 2009. International Conference on, Baku
Moritomo et al., 2005	Moritomo, T., O. Kiriyama, et al. (2005). *Object tracking in video pictures based on image segmentation and pattern matching*. IEEE International Symposium on Circuits and Systems 2005.
Nascimento and Marques, 2006	Nascimento, J. C. and J. S. Marques (2006). "*Performance Evaluation of object detection algorithm for video surveillance*." IEEE Transactions on Multimedia **8**(4): 761
Navas and Sasikumar, 2011	Navas, K. and M. Sasikumar (2011). "*Image Fidelity Metric: Future Directions*." IETE Technical Review **28**(1): 50-56
Needham and Boyle, 2003	Needham, C. J. and R. D. Boyle (2003). Performance Evaluation Metrics and Statistics for Positional Tracker Evaluation. Computer Vision Systems: Third International Conference, ICVS 2003, Graz, Austria, April 1-3, 2003. Proceedings, Springer Berlin / Heidelberg. **2626/2003:** 278-289.
Ngheim et al., 2007	Nghiem, A. T., F. Bremond, et al. (2007). *A New Evaluation Approach for Video Processing Algorithms*. Motion and Video Computing, 2007. WMVC '07. IEEE Workshop on Austin, Texas, US.
Ngheim et al, 2007	Nghiem, A. T., F. Bremond, et al. (2007). *ETISEO, performance evaluation for video surveillance*. Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on London.
Nillius and Eklundh, 2002	Nillius, P. and J.-O. Eklundh (2002). *Fast Block Matching with Normalized Cross-Correlation using Walsh Transforms*, Royal Institute of Technology
Niu and Abdel-Mottalen, 2004	Niu, F. and M. Abdel-Mottaleb (2004). *View-invariant human activity recognition based on shape and motion feature*. IEEE Sixth International Conference on Multimedia Software Engineering 2004.

Niu and Long, 2004	Niu, W., J. Long, et al. (2004). *Human Activity detection and recognition for video surveillance*. IEEE International Conference on Multimedia and Expo 2004.
Omkar et al, 2008	Omkar, S. N., J. Senthilnath, et al. (2008). "*Crop classification using biologically-inspired techniques with high resolution satellite image*." Journal of the Indian Society of Remote Sensing **36**(2): 175-182.
Park et al., 2000	Park, J., J. Seo, et al. (2000). *Detection of Human Faces using skin color and eyes*. 2000 IEEE International Conference on Multimedia and Expo 2000 (ICME 2000), New York, NY
Park and Aggarwal, 2002	Park, S. and J. K. Aggarwal (2002). *Segmentation and Tracking of interacting human body parts under occlusion and shadowing*. Workshop on Motion and Video Computing 2002.
Popoola and Amer, 2008	Popoola, J. and A. Amer (2008). *Performance evaluation for tracking algorithm using object labels*. IEEE International Conference on Acoustics, Speech and Signal Processing 2008, Las Vegas, NV
Poynton, 2012	Poynton, C. (2012). *Digital Video and HD Algorithms and Interfaces*. Burlington, Mass., Elsevier/Morgan Kaufmann
Ramanan, A. Forsyth et al, 2007	Ramanan, D., D. A. Forsyth, et al. (2007). "*Tracking People by Learning Their Appearance*." IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(1): 65-81
Sato and Aggarwal, 2001	Sato, K. and J. K. Aggarwal (2001). *Tracking and recognizing two-person interactions in outdoor image sequence*. 2001 IEEE Workshop on Multi-object tracking, Vancouver, BC, Canada.
Schairer, 1998	Schairer, G. (1998). *Improving Ground Truth Data Collection Efforts Using ArcView. National GAP Conference*. Santa Barbara.
Schogl et al., 2005	Schlogl, T., C. Beleznai, et al. (2005). *Performance Evaluation Metrics for Motion Detection and Tracking*. 17th International Conference on Pattern Recognition (ICPR'04).
Sebastian and Vooi Voon, 2007	Sebastian, P. and Y. Vooi Voon (2007). *Tracking Using Normalized Cross Correlation and Color Space*. Internation Conference on Intelligent and Advanced Systems (ICIAS 2007), Kuala Lumpur, Malaysia.
Sebastian and Vooi Voon, 2009	Sebastian, P. and Y. Vooi Voon (2009). *Tracking Consistency Metric for Video Surveillance Tracking*. International Conference on Signal Processing Systems, Singapore.

Sebastian and Vooi Voon, 2008	Sebastian, P., Y. Vooi Voon, et al. (2008). *The Effect of Color Space on Tracking Robustness*. 3rd IEEE Conference on Industrial Electronics and Applications, Singapore
Sebastian, Vooi Voon et al., 2010	Sebastian, P., Y. Vooi Voon, et al. (2010). "*Colour Space Effect on Tracking in Video Surveillance*." International Journal on Electrical Engineering and Informatics **2**(4): 298-312.
Sebastian and Vooi Voon, 2011	Sebastian, P., Y. Vooi Voon, et al. (2011). "*Performance Evaluation Metrics for Video Tracking*." IETE Technical Review **28**(6): 493-502.
Shao et. al. , 2014	Shao, G., J. Gao, et al. (2014). *Image Segmentation Based on Spatially Coherent Gaussian Mixture Model*. Digital lmage Computing: Techniques and Applications (DlCTA), 2014 International Conference on. Wollongong, NSW**:** 1-6.
Sharma, V. , 2012	Sharma, V. (2012). *A blob representation for tracking robust to merging and fragmentation*. Applications of Computer Vision (WACV), 2012 IEEE Workshop on, Breckenridge, CO..
Shen, G., G. Zhao, al. ,2010	Shen, G., G. Zhao, et al. (2010). *Method of motion detection and tracking based on multi-camera*. Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on, Chengdu.
Short, 2009	Short, N. (2009). *The LANDSAT tutorial workbook: Basics of satellite remote sensing*, NASA Goddard Space Flight Center
Siebel and Mayban 2009	Siebel, N. T. and S. Maybank (2002). *Fusion of Multiple Tracking Algorithms for Robust People Tracking*. In Proceedings of the 7th European Conference on Computer Vision (ECCV 2002).
Soriano and Martinkauppi, 2000	Soriano, M., B. Martinkauppi, et al. (2000). Skin Detection in vid under changing illumination conditions. 15th Internatio Conference on Pattern Recognition 2000, Barcelona.
Stern and Efros, 2002	Stern, H. and B. Efros (2002). *Adaptive Color Space Switching for Face Tracking in Multi-Colored Lighting Environments*. Fifth IEEE Conference on Automatic Face and Gesture Recognition 2002, Washington, D.C.
Sural, Qian et al. 2002	Sural, S., G. Qian, et al. (2002). *Segmentation and histogram generation using the HSV color space for image retrieval*. 2002 International Conference on Image Processing.
Tatiraju and Mehta, 2008	Tatiraju, S. and A. Mehta (2008). "*Image Segmentation using k-means clustering, EM and Normalized Cuts*."

Teknomo, Takeyama, et al., 2001	Teknomo, K., Y. Takeyama, et al. (2001). *Frame-based tracing of multiple objects*. 2001 IEEE Workshop on Multi-object Tracking, Vancouver, BC, Canada.
Thacket, Clark, et al, 2008	Thacker, N. A., A. F. Clark, et al. (2008). "*Performance characterization in computer vision: A guide to best practices*." Computer Vision and Image Understanding **109**(3): 305-334.
Trucco and Plakas, 2006	Trucco, E. and K. Plakas (2006). "*Video Tracking: A Concise Survey*." IEEE Journal of Oceanic Engineering **31**(2): 520-529.
Umbaugh, 1998	Umbaugh, S. E. (1998). *Computer Vision and Image Processing*, Prentice Hall.
Union, 2002	Union, I. T. (2002). *Recommendation ITU-R BT.709-5. Parameter values for the HDTV standards for production and international programme exchange* Electronic Publisher.
Union, 2011	Union, I. T. (2011). *Recommendation ITU-R BT.601-7. Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios*, Electronic Publication.
Wahl, Hillebrand et al., 2003	Wahl, E., U. Hillenbrand, et al. (2003). *Surflet-Pair-Relation Histograms: A statistical 3D-shape Representation for Rapid classification*. Fourth International Conference on 3-D Imaging and Modelling 2003.
Wang, H., C. Liu, et al., 2009	Wang, H., C. Liu, et al. (2009). *Multi-Camera Tracking Based on Information Fusion in Video Surveillance*. Image and Signal Processing, 2009. CISP '09. 2nd International Congress on, Tianjin.
Wang, Sang et. al., 2014	Wang, H., N. Sang, et al. (2014). *Real-Time Tracking Combined with Object Segmentation*. Pattern Recognition (ICPR), 2014 22nd International Conference on. Stockholm**:** 4098 - 4103.
Wang, Bebis et al., 2006	Wang, J., G. Bebis, et al. (2006). *Robust Video-Based Surveillance by Integrating Target Detection with Tracking*. 2006 Conference on Computer Vision and Pattern Recognition Workshop.
Wang and Liu, 2006	Wang, Q. and J. Liu (2006). *The Improved Particle Filter for Object Tracking*. 6th World Congress on Intelligent Control and Automation 2006, Dalian , China.
Wang, Phillips et al., 2001	Wang, Y., I. T. Phillips, et al. (2001). *Automatic table ground truth generation and a background-analysis-based table structure extraction method*. Sixth International Conference on Document Analysis and Recognition 2001, Seattle, WA.

Wang, Y., S. Velipasalar, et al., 2011	Wang, Y., S. Velipasalar, et al. (2011). *Wide-area multi-object tracking with non-overlapping camera views*. Multimedia and Expo (ICME), 2011 IEEE International Conference on, Barcelona.
Wu, Sankaranarayan et al., 2010	Wu, H., A. C. Sankaranarayana, et al. (2010). "*Online Empirical Evaluation of Tracking Algorithms*." Pattern Analysis and Machine Intelligence, IEEE Transactions on **32**(8): 1443
Yan and Forsyth, 2005	Yan, W. and D. A. Forsyth (2005). *Learning the Behavior of Users in a Public Space through Video Tracking*. Seventh IEEE Workshops on Application of Computer Vision 2005, WACV/MOTIONS '05.
Yang, Huang et al., 2006	Yang, L., W. Huang, et al. (2006). "*Semi-automatic ground truth generation for chart image recognition*." Lecture Notes in computer science(3872): 324-335.
Yang, Doherty et al. 2000	Yang, Y., J. F. Doherty, et al. (2000). *Moving Object Tracking in Video*. 29th Applied Imagery Pattern Recognition Workshop 2000, Washington, D.C.
Yiming and Guirong, 2014	Yiming, Q. and W. Guirong (2014). *Lung Nodule Segmentation Using EM Algorithm.* Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2014 Sixth International Conference on. Hangzhou, IEEE. **1:** 20 - 23.
Young and Ferryman, 2005	Young, D. P. and J. M. Ferryman (2005). *PETS metrics: on-line performance evaluation service*. Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on
Yunji, Z., Z. Bin, et al. ,2014	Yunji, Z., Z. Bin, et al. (2014). *Meanshift blob tracking with target model adaptive update*. Control Conference (CCC), 2014 33rd Chinese, Nanjing.
Zhai, Shafique et al., 2006	Zhai, Y., K. Shafique, et al. (2006). *KNIGHT: Object Detection and Tracking in Surveillance Videos*. CLEAR Evaluation and Workshop, Southampton, UK.
Zhao, Bu et al., 2002	Zhao, M., J. Bu, et al. (2002). *Robust background subtraction in HSV color space*. Proceedings of Multimedia Systems and Applications V, Boston, USA.
Zhu et. al., 2014	Zhu, L. and C.-q. Zhu (2014). *An algorithm for human face detection in color image based on Skin color segmentation.* Electronics, Computer and Applications, 2014 IEEE Workshop on. Ottawa, ON, IEEE**:** 101 - 104.

Zhu, Y.-f., 2009	Zhu, Y.-f. (2009). *Background Subtraction and Color Clustering Based Moving Objects Detection*. Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on
Zivkovic and Heijden, 2006	Zivkovic, Z. and F. v. d. Heijden (2006). "*Efficient adaptive density estimation per image pixel for the task of background subtraction*." Pattern Recognition Letters **27**: 773-780.
Zivkovic and Krose, 2004	Zivkovic, Z. and B. Krose (2004). *An EM-like algorithm for color-hostogram-based object tracking*. Proceedings of the 2004 IEEE Computer Vision Society Conference on Computer Vision and Pattern Recognition, 2004, CVPR 2004.

Appendix 1

This appendix contains the tabulated data for normalized cross correlation tracking performance using different layers in different colour spaces.

Input Video	Reference				
	whole _seq_1	whole _seq_2	whole _seq_3	whole _seq_4	whole _seq_5
whole_seq1_1	0.32	0.30	0.00	0.45	0.28
whole_seq1_2	0.20	0.18	0.00	0.23	0.23
whole_seq1_3	0.08	0.11	0.01	0.08	0.11
whole_seq1_4	0.34	0.53	0.00	0.25	0.34
whole_seq1_5	0.19	0.36	0.00	0.08	0.13

Normalized Cross Correlation Tracking Performance using Grayscale Colour Space

Input Video	Reference				
	whole _seq_1	whole _seq_2	whole _seq_3	whole _seq_4	whole _seq_5
whole_seq1_1	0.15	0.04	0.00	0.28	0.23
whole_seq1_2	0.13	0.00	0.00	0.15	0.15
whole_seq1_3	0.12	0.00	0.00	0.27	0.25
whole_seq1_4	0.12	0.02	0.00	0.41	0.10
whole_seq1_5	0.13	0.02	0.00	0.43	0.08

Normalized Cross Correlation Tracking Performance using RGB Colour Space

Input Video	Reference				
	whole _seq_1	whole _seq_2	whole _seq_3	whole _seq_4	whole _seq_5
whole_seq1_1	0.00	0.00	0.00	0.02	0.04
whole_seq1_2	0.00	0.03	0.00	0.03	0.05
whole_seq1_3	0.00	0.00	0.01	0.04	0.04
whole_seq1_4	0.05	0.09	0.00	0.05	0.10
whole_seq1_5	0.00	0.15	0.00	0.04	0.08

Normalized Cross Correlation Tracking Performance using HSV Colour Space

Input Video	Reference				
	whole _seq_1	whole _seq_2	whole _seq_3	whole _seq_4	whole _seq_5
whole_seq1_1	0.13	0.00	0.00	0.04	0.02
whole_seq1_2	0.00	0.08	0.00	0.05	0.08
whole_seq1_3	0.00	0.00	0.03	0.04	0.00
whole_seq1_4	0.03	0.19	0.00	0.12	0.10
whole_seq1_5	0.04	0.13	0.00	0.00	0.19

Normalized Cross Correlation Tracking Performance using HS layers in HSV Colour Space

Input Video	Reference				
	whole _seq_1	whole _seq_2	whole _seq_3	whole_ seq_4	whole _seq_5
whole_seq1_1	0.04	0.04	0.00	0.02	0.04
whole_seq1_2	0.05	0.05	0.00	0.03	0.05
whole_seq1_3	0.08	0.08	0.00	0.04	0.10
whole_seq1_4	0.10	0.10	0.00	0.05	0.10
whole_seq1_5	0.09	0.09	0.00	0.06	0.09

Normalized Cross Correlation Tracking Performance using YCbCr Colour Space

Input Video	Reference				
	whole _seq_1	whole _seq_2	whole _seq_3	whole _seq_4	whole _seq_5
whole_seq1_1	0.04	0.04	0.00	0.02	0.04
whole_seq1_2	0.05	0.05	0.00	0.03	0.05
whole_seq1_3	0.08	0.10	0.00	0.04	0.10
whole_seq1_4	0.10	0.10	0.00	0.05	0.10
whole_seq1_5	0.09	0.09	0.00	0.06	0.09

Normalized Cross Correlation Tracking Performance using CbCr layers in YCbCr Colour Space

Appendix 2

This appendix contains the tabulated data for normalized cross correlation tracking performance using different layers in different colour spaces with different background scenarios for whole person tracking.

Uneven Illumination

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.70	0.09	0.02	0.23	0.22	0.13	0.90	0.85	0.22
whole_seq2_2	0.75	0.17	0.09	0.28	0.18	0.13	0.84	0.69	0.24
whole_seq2_3	0.12	0.12	0.08	0.15	0.08	0.02	0.44	0.56	0.12
whole_seq2_4	0.34	0.05	0.02	0.02	0.00	0.00	0.93	0.73	0.25
whole_seq2_5	0.87	0.69	0.08	0.76	0.29	0.07	0.88	0.68	0.44
whole_seq2_6	0.80	0.24	0.10	0.54	0.04	0.08	0.83	0.82	0.01
whole_seq2_7	0.00	0.13	0.05	0.03	0.11	0.05	0.89	0.71	0.42
whole_seq2_8	0.00	0.00	0.00	0.20	0.00	0.00	0.20	0.80	0.20
whole_seq2_9	0.23	0.15	0.00	0.13	0.15	0.28	0.60	0.72	0.17

Normalized Cross Correlation Tracking Performance using Grayscale Colour Space

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.20	0.13	0.81	0.19	0.78	0.87	0.17	0.21	0.05
whole_seq2_2	0.16	0.06	0.75	0.09	0.86	0.91	0.18	0.21	0.03
whole_seq2_3	0.00	0.00	0.90	0.00	0.80	0.97	0.02	0.31	0.00
whole_seq2_4	0.00	0.00	0.93	0.11	0.84	0.92	0.00	0.21	0.00
whole_seq2_5	0.18	0.14	0.78	0.25	0.66	0.87	0.08	0.20	0.02
whole_seq2_6	0.18	0.09	0.72	0.21	0.84	0.86	0.13	0.23	0.05
whole_seq2_7	0.00	0.00	0.90	0.03	0.81	0.92	0.13	0.23	0.00
whole_seq2_8	0.15	0.00	1.00	0.04	0.72	0.93	0.05	0.40	0.00
whole_seq2_9	0.11	0.00	0.81	0.05	0.84	0.71	0.05	0.27	0.00

Normalized Cross Correlation Tracking Performance using RGB Colour Space

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
whole_seq2_2	0.00	0.01	0.03	0.00	0.03	0.00	0.00	0.00	0.00
whole_seq2_3	0.00	0.00	0.92	0.00	0.00	0.00	0.00	0.00	0.02
whole_seq2_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_5	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00
whole_seq2_6	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_7	0.00	0.00	0.08	0.00	0.13	0.00	0.00	0.00	0.00
whole_seq2_8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_9	0.00	0.00	0.04	0.00	0.07	0.00	0.00	0.00	0.00

Normalized Cross Correlation Tracking Performance using HSV Colour Space

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.04	0.02	0.26	0.03	0.01	0.05	0.18	0.01	0.19
whole_seq2_2	0.25	0.70	0.52	0.79	0.09	0.53	0.43	0.01	0.30
whole_seq2_3	0.21	0.12	0.96	0.56	0.02	0.65	0.06	0.02	0.48
whole_seq2_4	0.05	0.86	0.16	0.98	0.00	0.29	0.38	0.07	0.00
whole_seq2_5	0.00	0.00	0.01	0.00	0.02	0.02	0.09	0.00	0.00
whole_seq2_6	0.01	0.31	0.21	0.42	0.21	0.82	0.03	0.00	0.00
whole_seq2_7	0.02	0.00	0.08	0.03	0.06	0.00	0.31	0.00	0.02
whole_seq2_8	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.40	0.00
whole_seq2_9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01

Normalized Cross Correlation Tracking Performance using HS layers in HSV Colour Space

Input Video	Reference								
	whole _seq2_ 1	whole _seq2_ 2	whole _seq2_ 3	whole _seq2_ 4	whole _seq2_ 5	whole _seq2_ 6	whole _seq2_ 7	whole _seq2_ 8	whole _seq2_ 9
whole_seq2 _1	0.04	0.00	0.00	0.07	0.13	0.06	0.19	0.03	0.00
whole_seq2 _2	0.33	0.00	0.00	0.37	0.10	0.13	0.23	0.17	0.00
whole_seq2 _3	0.00	0.00	0.12	0.25	0.00	0.00	0.08	0.06	0.00
whole_seq2 _4	0.09	0.00	0.00	0.71	0.00	0.32	0.02	0.68	0.00
whole_seq2 _5	0.28	0.00	0.00	0.04	0.06	0.08	0.24	0.13	0.00
whole_seq2 _6	0.07	0.00	0.00	0.83	0.06	0.08	0.07	0.13	0.00
whole_seq2 _7	0.13	0.00	0.02	0.02	0.03	0.10	0.15	0.06	0.00
whole_seq2 _8	0.20	0.00	0.00	0.80	0.00	0.00	0.60	0.00	0.00
whole_seq2 _9	0.03	0.00	0.00	0.67	0.19	0.21	0.25	0.21	0.00

Normalized Cross Correlation Tracking Performance using YCbCr Colour Space

Input Video	Reference								
	whole _seq2_ 1	whole _seq2_ 2	whole _seq2_ 3	whole _seq2_ 4	whole _seq2_ 5	whole _seq2_ 6	whole _seq2_ 7	whole _seq2_ 8	whole _seq2_ 9
whole_seq2 _1	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00
whole_seq2 _2	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00
whole_seq2 _3	0.00	0.00	0.00	0.00	0.08	0.00	0.02	0.00	0.00
whole_seq2 _4	0.00	0.00	0.00	0.00	0.07	0.00	0.02	0.00	0.00
whole_seq2 _5	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00
whole_seq2 _6	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00
whole_seq2 _7	0.00	0.00	0.03	0.00	0.15	0.00	0.02	0.00	0.00
whole_seq2 _8	0.00	0.00	0.00	0.00	0.07	0.00	0.02	0.00	0.00
whole_seq2 _9	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00

Normalized Cross Correlation Tracking Performance using CbCr layers in YCbCr Colour Space

Uneven Background

Input Video	Reference								
	whole _seq2_ 1	whole _seq2_ 2	whole _seq2_ 3	whole _seq2_ 4	whole _seq2_ 5	whole _seq2_ 6	whole _seq2_ 7	whole _seq2_ 8	whole _seq2_ 9
whole_seq2 _1	0.76	0.23	0.33	0.53	0.29	0.79	0.97	0.93	0.01
whole_seq2 _2	0.61	0.14	0.26	0.49	0.43	0.82	0.93	0.94	0.00
whole_seq2 _3	0.00	0.00	0.00	0.00	0.21	0.43	0.17	0.60	0.00
whole_seq2 _4	0.30	0.00	0.19	0.13	0.07	0.52	0.90	1.00	0.00
whole_seq2 _5	0.76	0.62	0.28	0.87	0.33	0.20	0.99	0.89	0.01
whole_seq2 _6	0.90	0.00	0.14	0.80	0.04	0.36	1.00	0.92	0.00
whole_seq2 _7	0.00	0.00	0.04	0.02	0.14	0.49	0.65	0.93	0.02
whole_seq2 _8	0.00	0.00	0.00	0.07	0.00	0.47	0.73	1.00	0.00
whole_seq2 _9	0.00	0.00	0.03	0.01	0.12	0.79	0.51	0.92	0.00

Normalized Cross Correlation Tracking Performance using Grayscale Colour Space

Input Video	Reference								
	whole _seq2_ 1	whole _seq2_ 2	whole _seq2_ 3	whole _seq2_ 4	whole _seq2_ 5	whole _seq2_ 6	whole _seq2_ 7	whole _seq2_ 8	whole _seq2_ 9
whole_seq2 _1	0.13	0.00	0.00	0.06	0.01	0.00	0.09	0.21	0.00
whole_seq2 _2	0.18	0.00	0.00	0.07	0.03	0.03	0.22	0.22	0.00
whole_seq2 _3	0.07	0.00	0.00	0.00	0.00	0.00	0.72	0.19	0.00
whole_seq2 _4	0.17	0.00	0.00	0.00	0.00	0.00	0.30	0.28	0.00
whole_seq2 _5	0.14	0.06	0.02	0.23	0.06	0.00	0.10	0.22	0.00
whole_seq2 _6	0.06	0.00	0.00	0.04	0.00	0.00	0.00	0.34	0.00
whole_seq2 _7	0.13	0.00	0.00	0.00	0.00	0.00	0.40	0.23	0.00
whole_seq2 _8	0.10	0.00	0.00	0.00	0.00	0.00	0.20	0.40	0.00
whole_seq2 _9	0.10	0.00	0.00	0.00	0.01	0.00	0.67	0.25	0.00

Normalized Cross Correlation Tracking Performance using RGB Colour Space

Input Video	Reference								
	whole _seq2_ 1	whole _seq2_ 2	whole _seq2_ 3	whole _seq2_ 4	whole _seq2_ 5	whole _seq2_ 6	whole _seq2_ 7	whole _seq2_ 8	whole _seq2_ 9
whole_seq2_ 1	0.00	0.00	0.00	0.00	0.10	0.04	0.03	0.04	0.00
whole_seq2_ 2	0.00	0.00	0.00	0.00	0.11	0.00	0.03	0.00	0.00
whole_seq2_ 3	0.00	0.00	0.36	0.00	0.02	0.02	0.03	0.02	0.00
whole_seq2_ 4	0.00	0.00	0.00	0.00	0.01	0.01	0.15	0.01	0.00
whole_seq2_ 5	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
whole_seq2_ 6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_ 7	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq2_ 8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_ 9	0.00	0.00	0.00	0.00	0.03	0.00	0.01	0.04	0.00

Normalized Cross Correlation Tracking Performance using HSV Colour Space

Input Video	Reference								
	whole _seq2_ 1	whole _seq2_ 2	whole _seq2_ 3	whole _seq2_ 4	whole _seq2_ 5	whole _seq2_ 6	whole _seq2_ 7	whole _seq2_ 8	whole _seq2_ 9
whole_seq2_ 1	0.09	0.00	0.23	0.10	0.07	0.08	0.19	0.02	0.06
whole_seq2_ 2	0.36	0.72	0.39	0.88	0.60	0.42	0.78	0.06	0.18
whole_seq2_ 3	0.05	0.00	0.48	0.45	0.12	0.17	0.29	0.00	0.07
whole_seq2_ 4	0.00	0.13	0.04	1.00	0.06	0.06	0.81	0.19	0.01
whole_seq2_ 5	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.01
whole_seq2_ 6	0.00	0.00	0.10	0.60	0.52	0.62	0.10	0.00	0.00
whole_seq2_ 7	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00
whole_seq2_ 8	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00
whole_seq2_ 9	0.00	0.00	0.00	0.00	0.01	0.03	0.04	0.00	0.00

Normalized Cross Correlation Tracking Performance using HS layers in HSV Colour Space

Input Video	Reference								
	whole _seq2_ 1	whole _seq2_ 2	whole _seq2_ 3	whole _seq2_ 4	whole _seq2_ 5	whole _seq2_ 6	whole _seq2_ 7	whole _seq2_ 8	whole _seq2_ 9
whole_seq2_ 1	0.11	0.00	0.00	0.53	0.10	0.11	0.48	0.04	0.00
whole_seq2_ 2	0.72	0.00	0.00	0.25	0.07	0.38	0.07	0.90	0.00
whole_seq2_ 3	0.00	0.00	0.00	0.57	0.10	0.12	0.03	0.05	0.00
whole_seq2_ 4	0.52	0.00	0.00	0.30	0.00	0.81	0.00	0.99	0.00
whole_seq2_ 5	0.37	0.00	0.00	0.38	0.15	0.11	0.24	0.21	0.00
whole_seq2_ 6	0.00	0.00	0.00	1.00	0.00	0.04	0.00	0.06	0.00
whole_seq2_ 7	0.05	0.00	0.00	0.58	0.00	0.04	0.12	0.12	0.00
whole_seq2_ 8	0.00	0.00	0.00	0.67	0.00	0.00	0.07	0.07	0.00
whole_seq2_ 9	0.30	0.00	0.00	0.70	0.01	0.33	0.00	0.26	0.00

Normalized Cross Correlation Tracking Performance using YCbCr Colour Space

Input Video	Reference								
	whole _seq2_ 1	whole _seq2_ 2	whole _seq2_ 3	whole _seq2_ 4	whole _seq2_ 5	whole _seq2_ 6	whole _seq2_ 7	whole _seq2_ 8	whole _seq2_ 9
whole_seq2_ 1	0.28	0.00	0.00	0.28	0.10	0.22	0.09	0.63	0.00
whole_seq2_ 2	0.17	0.67	0.00	0.74	0.00	0.19	0.13	0.46	0.00
whole_seq2_ 3	0.10	0.59	0.00	0.79	0.00	0.10	0.22	0.14	0.00
whole_seq2_ 4	0.10	0.75	0.00	1.00	0.01	0.27	0.01	0.28	0.00
whole_seq2_ 5	0.18	0.00	0.00	0.07	0.17	0.15	0.15	0.26	0.00
whole_seq2_ 6	0.66	0.00	0.00	0.86	0.54	0.74	0.04	0.96	0.00
whole_seq2_ 7	0.21	0.00	0.00	0.25	0.07	0.09	0.12	0.16	0.00
whole_seq2_ 8	1.00	0.00	0.00	0.33	0.00	0.20	0.00	1.00	0.00
whole_seq2_ 9	0.16	0.12	0.00	0.99	0.05	0.27	0.00	0.27	0.00

Normalized Cross Correlation Tracking Performance using CbCr layers in YCbCr Colour Space

Cluttered Foreground and Background

Input Video	Reference								
	whole _seq2_ 1	whole _seq2_ 2	whole _seq2_ 3	whole _seq2_ 4	whole _seq2_ 5	whole _seq2_ 6	whole _seq2_ 7	whole _seq2_ 8	whole _seq2_ 9
whole_seq2_ 1	0.21	0.00	0.00	0.00	0.36	0.57	0.00	0.21	0.14
whole_seq2_ 2	0.00	0.00	0.00	0.00	0.08	0.79	0.15	0.11	0.43
whole_seq2_ 3	0.00	0.00	0.00	0.00	0.13	0.30	0.00	0.00	0.00
whole_seq2_ 4	0.00	0.00	0.00	0.00	0.10	0.79	0.00	0.00	0.00
whole_seq2_ 5	0.00	0.00	0.00	0.00	0.45	0.45	0.03	0.07	0.07
whole_seq2_ 6	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.44
whole_seq2_ 7	0.00	0.00	0.00	0.00	0.33	0.63	0.15	0.02	0.11
whole_seq2_ 8	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.48
whole_seq2_ 9	0.00	0.00	0.00	0.00	0.17	0.84	0.07	0.26	0.00

Normalized Cross Correlation Tracking Performance using Grayscale Colour Space

Input Video	Reference								
	whole _seq2_ 1	whole _seq2_ 2	whole _seq2_ 3	whole _seq2_ 4	whole _seq2_ 5	whole _seq2_ 6	whole _seq2_ 7	whole _seq2_ 8	whole _seq2_ 9
whole_seq2_ 1	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.14	0.00
whole_seq2_ 2	0.00	0.00	0.95	0.00	0.12	0.14	0.00	0.00	0.00
whole_seq2_ 3	0.00	0.00	0.99	0.00	0.10	0.03	0.00	0.00	0.00
whole_seq2_ 4	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_ 5	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_ 6	0.00	0.00	1.00	0.00	0.19	0.10	0.00	0.00	0.00
whole_seq2_ 7	0.00	0.00	1.00	0.00	0.13	0.06	0.07	0.00	0.00
whole_seq2_ 8	0.00	0.00	1.00	0.00	0.09	0.11	0.00	0.00	0.00
whole_seq2_ 9	0.00	0.00	1.00	0.00	0.14	0.03	0.06	0.26	0.00

Normalized Cross Correlation Tracking Performance using RGB Colour Space

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_3	0.00	0.00	0.12	0.12	0.00	0.02	0.00	0.00	0.05
whole_seq2_4	0.00	0.00	0.00	0.28	0.00	0.00	0.00	0.00	0.00
whole_seq2_5	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00
whole_seq2_6	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00
whole_seq2_7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_8	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00
whole_seq2_9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Normalized Cross Correlation Tracking Performance using HSV Colour Space

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00
whole_seq2_2	0.19	0.40	0.24	0.79	0.62	0.49	0.14	0.17	0.19
whole_seq2_3	0.16	0.00	0.24	0.30	0.52	0.48	0.00	0.04	0.55
whole_seq2_4	0.17	0.45	0.14	1.00	0.62	0.59	0.03	0.10	0.03
whole_seq2_5	0.00	0.00	0.00	0.00	0.24	0.03	0.00	0.00	0.00
whole_seq2_6	0.00	0.00	0.04	0.29	0.23	0.83	0.04	0.00	0.00
whole_seq2_7	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.02
whole_seq2_8	0.00	0.00	0.00	0.00	0.39	0.00	0.00	0.04	0.00
whole_seq2_9	0.00	0.00	0.00	0.01	0.11	0.03	0.00	0.00	0.01

Normalized Cross Correlation Tracking Performance using HS layers in HSV Colour Space

Input Video	Reference								
	whole _seq2_ 1	whole _seq2_ 2	whole _seq2_ 3	whole _seq2_ 4	whole _seq2_ 5	whole _seq2_ 6	whole _seq2_ 7	whole _seq2_ 8	whole _seq2_ 9
whole_seq2_ 1	0.00	0.00	0.00	0.00	0.21	0.14	0.00	0.00	0.00
whole_seq2_ 2	0.80	0.00	0.00	0.08	0.20	0.85	0.26	0.67	0.00
whole_seq2_ 3	0.00	0.00	0.69	0.19	0.09	0.21	0.25	0.07	0.00
whole_seq2_ 4	1.00	0.00	0.00	0.34	0.82	1.00	0.17	0.97	0.00
whole_seq2_ 5	0.38	0.00	0.00	0.00	0.21	0.00	0.31	0.14	0.00
whole_seq2_ 6	0.62	0.00	0.00	0.75	0.12	0.87	0.21	0.27	0.00
whole_seq2_ 7	0.22	0.00	0.00	0.11	0.35	0.19	0.09	0.13	0.00
whole_seq2_ 8	0.43	0.00	0.00	0.33	0.26	0.63	0.15	0.87	0.00
whole_seq2_ 9	0.00	0.00	0.00	0.64	0.34	0.83	0.21	0.11	0.00

Normalized Cross Correlation Tracking Performance using YCbCr Colour Space

Input Video	Reference								
	whole _seq2_ 1	whole _seq2_ 2	whole _seq2_ 3	whole _seq2_ 4	whole _seq2_ 5	whole _seq2_ 6	whole _seq2_ 7	whole _seq2_ 8	whole _seq2_ 9
whole_seq2_ 1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_ 2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_ 3	0.00	0.00	0.18	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_ 4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_ 5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_ 6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_ 7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_ 8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq2_ 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Normalized Cross Correlation Tracking Performance using CbCr layers in YCbCr Colour Space

Appendix 3

This appendix contains the tabulated data for average histogram tracking performance using different layers in different colour spaces.

Histogram Intersection

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_3	0.00	0.07	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_4	0.00	0.10	0.00	0.07	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Euclidean intersection

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.37	0.15	0.13	0.00	0.00	0.04	0.01	0.01	0.00
whole_seq1_3	0.00	0.22	0.04	0.03	0.00	0.00	0.02	0.00	0.00	0.00
whole_seq1_4	0.00	0.15	0.00	0.07	0.02	0.00	0.02	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.01	0.00

Chi Squared 1

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.24	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
whole_seq1_3	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.05	0.00	0.08	0.00	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Chi Squared 2

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.24	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00
whole_seq1_3	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.05	0.00	0.08	0.00	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Kullbeck Leibler

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.34	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00
whole_seq1_3	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.07	0.00	0.07	0.00	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Average Histogram Tracking Performance using Grayscale Colour Space

Histogram Intersection

Reference Histogram	TDR Input Video whole_ seq1_1	TDR Input Video whole_ seq1_2	TDR Input Video whole_ seq1_3	TDR Input Video whole_ seq1_4	TDR Input Video whole_ seq1_5	TC Input Video whole_ seq1_1	TC Input Video whole_ seq1_2	TC Input Video whole_ seq1_3	TC Input Video whole_ seq1_4	TC Input Video whole_ seq1_5
whole_seq1_1	0.29	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.34	0.00	0.10	0.00	0.00	0.04	0.00	0.01	0.00
whole_seq1_3	0.00	0.00	0.24	0.00	0.11	0.00	0.00	0.01	0.00	0.01
whole_seq1_4	0.00	0.10	0.00	0.33	0.00	0.00	0.01	0.00	0.02	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.01

Euclidean intersection

Reference Histogram	TDR Input Video whole_ seq1_1	TDR Input Video whole_ seq1_2	TDR Input Video whole_ seq1_3	TDR Input Video whole_ seq1_4	TDR Input Video whole_ seq1_5	TC Input Video whole_ seq1_1	TC Input Video whole_ seq1_2	TC Input Video whole_ seq1_3	TC Input Video whole_ seq1_4	TC Input Video whole_ seq1_5
whole_seq1_1	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.37	0.15	0.13	0.00	0.00	0.04	0.01	0.01	0.00
whole_seq1_3	0.00	0.22	0.04	0.03	0.00	0.00	0.02	0.00	0.00	0.00
whole_seq1_4	0.00	0.15	0.00	0.07	0.02	0.00	0.02	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.01	0.00

Chi Squared 1

Reference Histogram	TDR Input Video whole_ seq1_1	TDR Input Video whole_ seq1_2	TDR Input Video whole_ seq1_3	TDR Input Video whole_ seq1_4	TDR Input Video whole_ seq1_5	TC Input Video whole_ seq1_1	TC Input Video whole_ seq1_2	TC Input Video whole_ seq1_3	TC Input Video whole_ seq1_4	TC Input Video whole_ seq1_5
whole_seq1_1	0.33	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.32	0.00	0.18	0.00	0.00	0.04	0.00	0.02	0.00
whole_seq1_3	0.00	0.00	0.23	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.10	0.00	0.28	0.00	0.00	0.01	0.00	0.02	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.01

Chi Squared 2

Reference Histogram	TDR Input Video whole_ seq1_1	TDR Input Video whole_ seq1_2	TDR Input Video whole_ seq1_3	TDR Input Video whole_ seq1_4	TDR Input Video whole_ seq1_5	TC Input Video whole_ seq1_1	TC Input Video whole_ seq1_2	TC Input Video whole_ seq1_3	TC Input Video whole_ seq1_4	TC Input Video whole_ seq1_5
whole_seq1_1	0.23	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.34	0.00	0.28	0.00	0.00	0.03	0.00	0.05	0.00
whole_seq1_3	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.10	0.00	0.25	0.00	0.00	0.01	0.00	0.02	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.01

Kullbeck Leibler

Reference Histogram	TDR Input Video whole_ seq1_1	TDR Input Video whole_ seq1_2	TDR Input Video whole_ seq1_3	TDR Input Video whole_ seq1_4	TDR Input Video whole_ seq1_5	TC Input Video whole_ seq1_1	TC Input Video whole_ seq1_2	TC Input Video whole_ seq1_3	TC Input Video whole_ seq1_4	TC Input Video whole_ seq1_5
whole_seq1_1	0.35	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.34	0.00	0.20	0.00	0.00	0.06	0.00	0.02	0.00
whole_seq1_3	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.10	0.00	0.23	0.00	0.00	0.01	0.00	0.01	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00

Average Histogram Tracking Performance using H-layer in HSV Colour Space

Histogram Intersection

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.00	0.00	0.04	0.05	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.10	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
whole_seq1_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.27	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq1_5	0.02	0.02	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00

Euclidean intersection

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.00	0.17	0.00	0.10	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_2	0.00	0.00	0.09	0.00	0.28	0.00	0.00	0.00	0.00	0.02
whole_seq1_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.13	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Chi Squared 1

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.00	0.12	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Chi Squared 2

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.00	0.07	0.00	0.10	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_2	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.33	0.00	0.00
whole_seq1_5	0.00	0.12	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00

Kullbeck Leibler

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Average Histogram Tracking Performance using S-layer in HSV Colour Space

Histogram Intersection

Reference Histogram	TDR Input Video whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	TC Input Video whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.29	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.24	0.00	0.13	0.00	0.00	0.02	0.00	0.01	0.00
whole_seq1_3	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.01

Euclidean intersection

Reference Histogram	TDR Input Video whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	TC Input Video whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.19	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.27	0.00	0.03	0.30	0.00	0.02	0.00	0.00	0.08
whole_seq1_3	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.01	0.00	0.01
whole_seq1_4	0.00	0.32	0.00	0.20	0.33	0.00	0.04	0.00	0.01	0.12
whole_seq1_5	0.00	0.00	0.35	0.00	0.24	0.00	0.00	0.02	0.00	0.02

Chi Squared 1

Reference Histogram	TDR Input Video whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	TC Input Video whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.31	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.29	0.00	0.17	0.00	0.00	0.03	0.00	0.01	0.00
whole_seq1_3	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.28	0.00	0.00	0.00	0.00	0.02	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.01

Chi Squared 2

Reference Histogram	TDR Input Video whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	TC Input Video whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.25	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.29	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
whole_seq1_3	0.00	0.00	0.16	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.04	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.28	0.00	0.00	0.00	0.00	0.02

Kullbeck Leibler

Reference Histogram	TDR Input Video whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	TC Input Video whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.27	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.29	0.00	0.20	0.00	0.00	0.03	0.00	0.03	0.00
whole_seq1_3	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.35	0.00	0.00	0.00	0.00	0.03	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.01

Average Histogram Tracking Performance using Cb-layer in YCbCr Colour Space

Histogram Intersection

Reference Histogram	TDR Input Video					TC Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.29	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.24	0.00	0.13	0.00	0.00	0.02	0.00	0.01	0.00
whole_seq1_3	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.01

Euclidean intersection

Reference Histogram	TDR Input Video					TC Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.19	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.27	0.00	0.03	0.30	0.00	0.02	0.00	0.00	0.08
whole_seq1_3	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.01	0.00	0.01
whole_seq1_4	0.00	0.32	0.00	0.20	0.33	0.00	0.04	0.00	0.01	0.12
whole_seq1_5	0.00	0.00	0.35	0.00	0.24	0.00	0.00	0.02	0.00	0.02

Chi Squared 1

Reference Histogram	TDR Input Video					TC Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.31	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.29	0.00	0.17	0.00	0.00	0.03	0.00	0.01	0.00
whole_seq1_3	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.28	0.00	0.00	0.00	0.00	0.02	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.01

Chi Squared 2

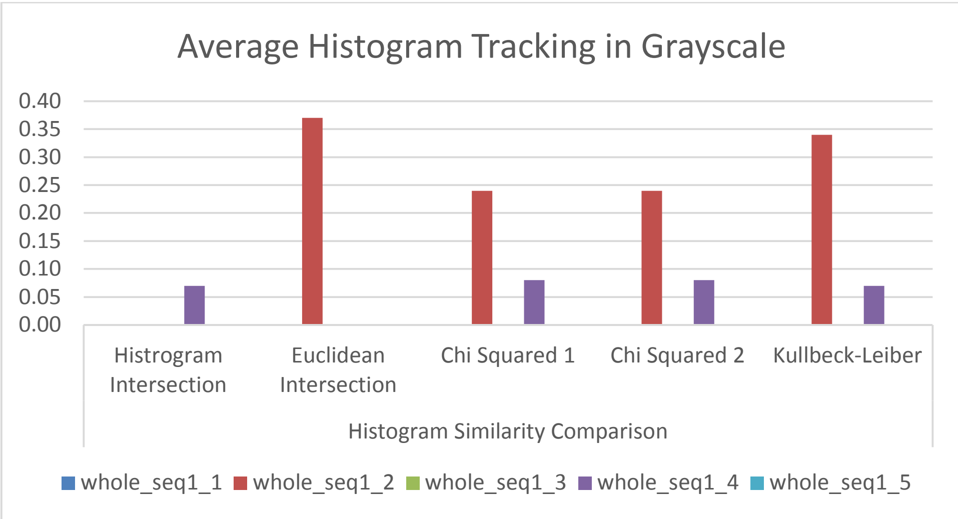
Reference Histogram	TDR Input Video					TC Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.25	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.29	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
whole_seq1_3	0.00	0.00	0.16	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.04	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.28	0.00	0.00	0.00	0.00	0.02

Kullbeck Leibler

Reference Histogram	TDR Input Video					TC Input Video				
	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.27	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.29	0.00	0.20	0.00	0.00	0.03	0.00	0.03	0.00
whole_seq1_3	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.35	0.00	0.00	0.00	0.00	0.03	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.01

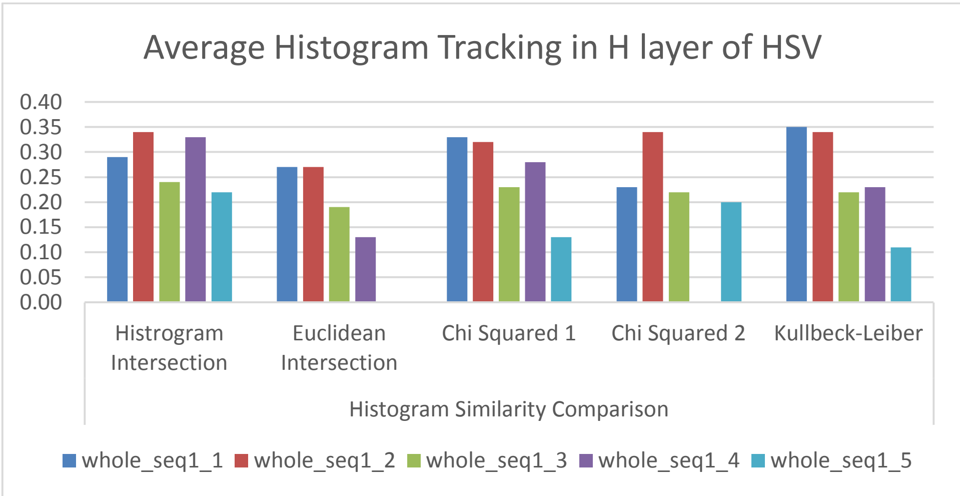
Average Histogram Tracking Performance using Cr-layer in YCbCr Colour Space

Input Video	Histogram Similarity Comparison				
	Histrogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.37	0.24	0.24	0.34
whole_seq1_3	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.07	0.00	0.08	0.08	0.07
whole_seq1_5	0.00	0.00	0.00	0.00	0.00



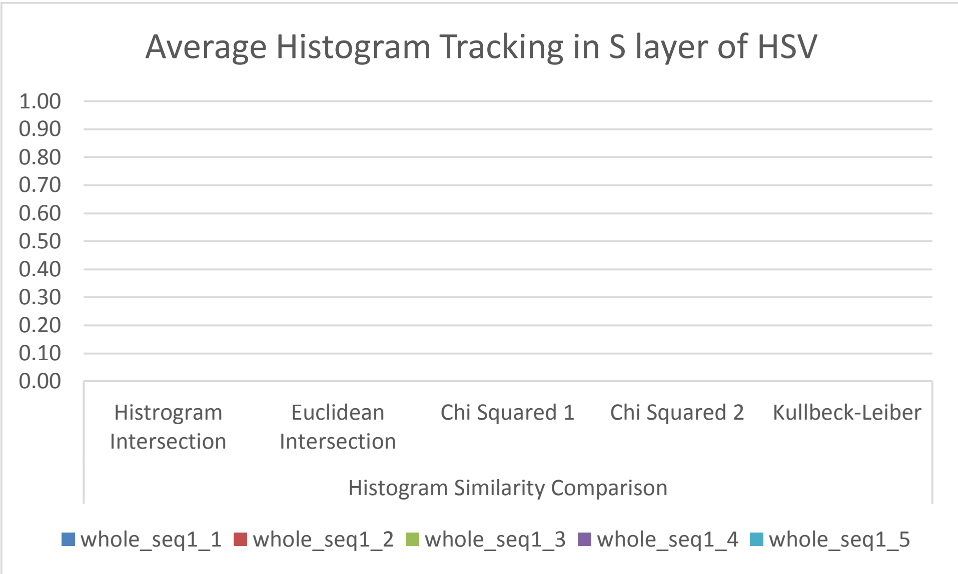
Summary of Average Histogram Tracking in Grayscale

Input Video	Histogram Similarity Comparison				
	Histrogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.29	0.27	0.33	0.23	0.35
whole_seq1_2	0.34	0.27	0.32	0.34	0.34
whole_seq1_3	0.24	0.19	0.23	0.22	0.22
whole_seq1_4	0.33	0.13	0.28	0.00	0.23
whole_seq1_5	0.22	0.00	0.13	0.20	0.11



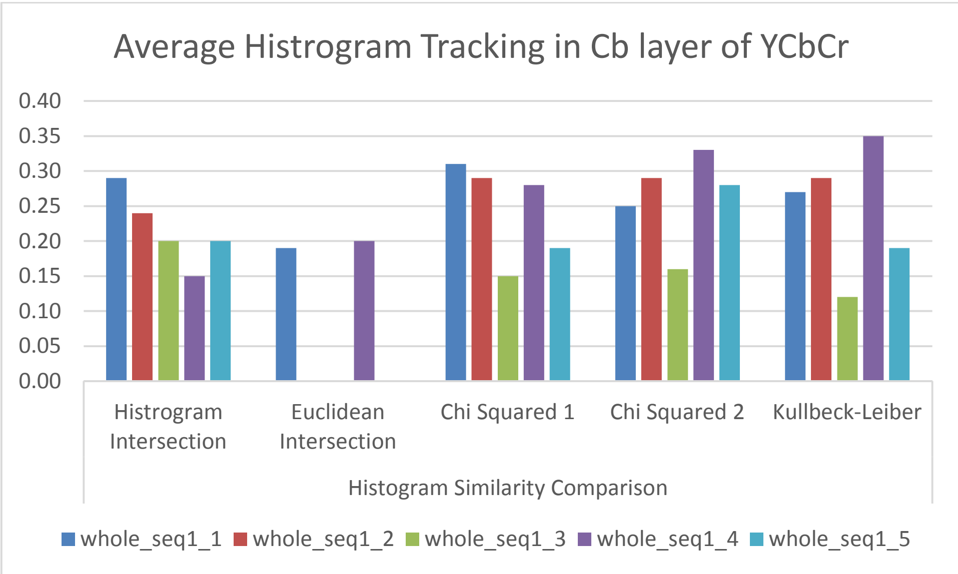
Summary of Average Histogram Tracking in H layer of HSV Colour Space

Input Video	Histogram Similarity Comparison				
	Histrogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.00	0.00	0.00	0.00
whole_seq1_3	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.00



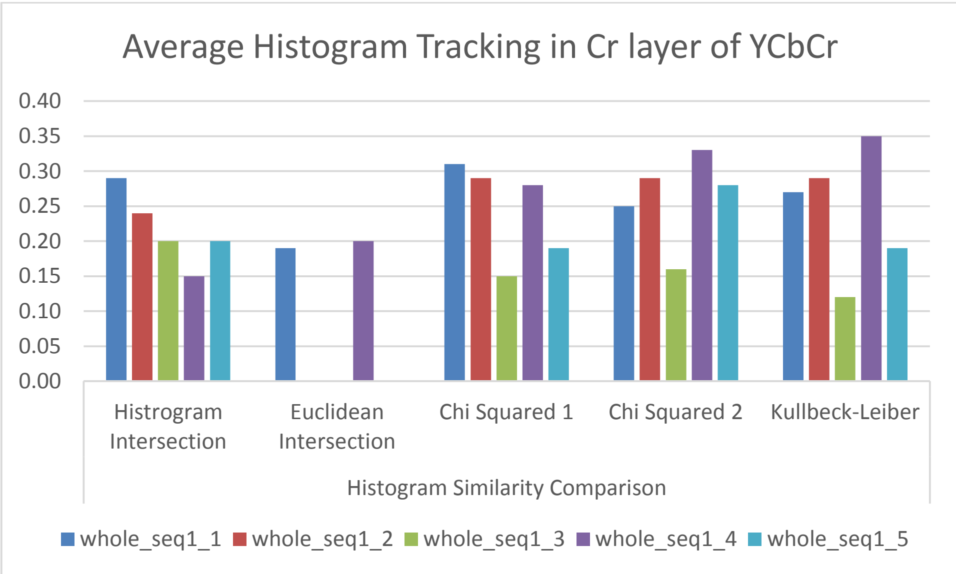
Summary of Average Histogram Tracking in S Layer of HSV Colour Space

Input Video	Histogram Similarity Comparison				
	Histrogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leilber
whole_seq1_1	0.29	0.19	0.31	0.25	0.27
whole_seq1_2	0.24	0.00	0.29	0.29	0.29
whole_seq1_3	0.20	0.00	0.15	0.16	0.12
whole_seq1_4	0.15	0.20	0.28	0.33	0.35
whole_seq1_5	0.20	0.00	0.19	0.28	0.19



Summary of Average Histogram Tracking in Cb Layer of YCbCr Colour Space

Input Video	Histogram Similarity Comparison				
	Histrogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.29	0.19	0.31	0.25	0.27
whole_seq1_2	0.24	0.00	0.29	0.29	0.29
whole_seq1_3	0.20	0.00	0.15	0.16	0.12
whole_seq1_4	0.15	0.20	0.28	0.33	0.35
whole_seq1_5	0.20	0.00	0.19	0.28	0.19



Summary of Average Histogram Tracking in Cr Layer of YCbCr Colour Space

Appendix 4

This appendix contains the tabulated data for Alpha-Trimmed histogram tracking performance using different layers in different colour spaces.

Histogram Intersection

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.23	0.21	0.04	0.19	0.06	0.02	0.01	0.00	0.01	0.00
whole_seq1_2	0.24	0.17	0.27	0.15	0.15	0.02	0.02	0.02	0.01	0.01
whole_seq1_3	0.06	0.06	0.03	0.05	0.04	0.01	0.01	0.00	0.01	0.01
whole_seq1_4	0.03	0.03	0.02	0.02	0.01	0.02	0.02	0.01	0.02	0.00
whole_seq1_5	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.02	0.01	0.01

Euclidean intersection

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.13	0.21	0.21	0.02	0.21	0.01	0.02	0.02	0.00	0.02
whole_seq1_2	0.07	0.32	0.32	0.12	0.32	0.00	0.03	0.02	0.01	0.02
whole_seq1_3	0.16	0.19	0.15	0.04	0.14	0.01	0.01	0.00	0.00	0.00
whole_seq1_4	0.03	0.05	0.08	0.10	0.05	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.22	0.28	0.19	0.00	0.17	0.02	0.02	0.01	0.00	0.01

Chi Squared 1

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.21	0.21	0.00	0.08	0.71	0.03	0.04	0.00	0.00	0.23
whole_seq1_2	0.22	0.24	0.07	0.20	0.93	0.02	0.03	0.00	0.03	0.46
whole_seq1_3	0.18	0.20	0.20	0.32	0.85	0.03	0.03	0.01	0.04	0.26
whole_seq1_4	0.13	0.12	0.10	0.12	0.73	0.01	0.01	0.00	0.01	0.34
whole_seq1_5	0.06	0.07	0.07	0.13	0.52	0.00	0.01	0.00	0.01	0.07

Chi Squared 2

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.04	0.31	0.06	0.19	0.10	0.00	0.02	0.00	0.01	0.01
whole_seq1_2	0.20	0.32	0.07	0.34	0.22	0.01	0.03	0.01	0.04	0.02
whole_seq1_3	0.01	0.18	0.16	0.22	0.27	0.00	0.02	0.01	0.01	0.02
whole_seq1_4	0.00	0.03	0.07	0.02	0.20	0.00	0.00	0.00	0.00	0.02
whole_seq1_5	0.07	0.11	0.17	0.20	0.30	0.00	0.01	0.01	0.02	0.02

Kullbeck Leibler

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.15	0.19	0.00	0.13	0.04	0.01	0.02	0.00	0.01	0.00
whole_seq1_2	0.22	0.24	0.12	0.22	0.22	0.02	0.03	0.01	0.03	0.02
whole_seq1_3	0.18	0.18	0.31	0.15	0.12	0.03	0.03	0.02	0.01	0.00
whole_seq1_4	0.12	0.15	0.10	0.12	0.12	0.01	0.01	0.01	0.01	0.01
whole_seq1_5	0.04	0.07	0.04	0.17	0.24	0.00	0.01	0.00	0.01	0.01

Average Histogram Tracking Performance using Grayscale Colour Space

Histogram Intersection

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.21	0.02	0.00	0.06	0.04	0.02	0.00	0.00	0.00	0.00
whole_seq1_2	0.02	0.32	0.15	0.07	0.02	0.00	0.04	0.01	0.00	0.00
whole_seq1_3	0.03	0.11	0.16	0.30	0.08	0.00	0.01	0.01	0.01	0.00
whole_seq1_4	0.07	0.02	0.13	0.23	0.15	0.00	0.00	0.01	0.01	0.01
whole_seq1_5	0.24	0.04	0.04	0.04	0.37	0.02	0.00	0.00	0.00	0.03

Euclidean intersection

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.15	0.13	0.02	0.15	0.15	0.01	0.01	0.00	0.02	0.02
whole_seq1_2	0.27	0.12	0.10	0.15	0.12	0.03	0.01	0.01	0.01	0.01
whole_seq1_3	0.10	0.21	0.16	0.18	0.14	0.00	0.01	0.01	0.01	0.00
whole_seq1_4	0.17	0.17	0.10	0.10	0.12	0.01	0.01	0.00	0.01	0.02
whole_seq1_5	0.13	0.33	0.22	0.17	0.13	0.01	0.04	0.02	0.01	0.01

Chi Squared 1

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.19	0.04	0.29	0.08	0.13	0.01	0.00	0.03	0.01	0.01
whole_seq1_2	0.07	0.17	0.07	0.27	0.02	0.00	0.02	0.00	0.03	0.00
whole_seq1_3	0.22	0.01	0.15	0.18	0.16	0.01	0.00	0.01	0.01	0.01
whole_seq1_4	0.33	0.05	0.13	0.25	0.15	0.06	0.00	0.01	0.03	0.01
whole_seq1_5	0.09	0.00	0.11	0.04	0.30	0.01	0.00	0.01	0.00	0.02

Chi Squared 2

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.42	0.02	0.25	0.13	0.15	0.05	0.00	0.02	0.01	0.01
whole_seq1_2	0.37	0.10	0.02	0.41	0.05	0.04	0.01	0.00	0.06	0.00
whole_seq1_3	0.34	0.04	0.11	0.21	0.30	0.02	0.00	0.00	0.00	0.01
whole_seq1_4	0.32	0.10	0.15	0.20	0.12	0.03	0.01	0.01	0.02	0.01
whole_seq1_5	0.13	0.04	0.17	0.02	0.35	0.01	0.00	0.01	0.00	0.02

Kullbeck Leibler

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.21	0.04	0.31	0.08	0.08	0.01	0.00	0.03	0.01	0.00
whole_seq1_2	0.05	0.17	0.07	0.22	0.02	0.00	0.02	0.01	0.02	0.00
whole_seq1_3	0.23	0.04	0.18	0.15	0.15	0.01	0.00	0.01	0.01	0.01
whole_seq1_4	0.37	0.07	0.15	0.25	0.15	0.06	0.01	0.01	0.03	0.01
whole_seq1_5	0.09	0.00	0.11	0.02	0.28	0.01	0.00	0.01	0.00	0.01

Average Histogram Tracking Performance using H-layer in HSV Colour Space

Histogram Intersection

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.21	0.00	0.02	0.02	0.04	0.01	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.12	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_3	0.00	0.00	0.23	0.33	0.05	0.00	0.00	0.01	0.02	0.00
whole_seq1_4	0.00	0.00	0.10	0.18	0.08	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.17	0.00	0.02	0.02	0.11	0.01	0.00	0.00	0.00	0.01

Euclidean Intersection

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.25	0.00	0.06	0.25	0.00	0.03	0.00	0.00	0.03	0.00
whole_seq1_2	0.00	0.20	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_3	0.33	0.00	0.16	0.15	0.10	0.03	0.00	0.01	0.01	0.00
whole_seq1_4	0.42	0.02	0.10	0.20	0.00	0.05	0.00	0.00	0.01	0.00
whole_seq1_5	0.00	0.00	0.04	0.00	0.26	0.00	0.00	0.00	0.00	0.01

Chi Squared 1

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.35	0.00	0.06	0.00	0.21	0.03	0.00	0.00	0.00	0.01
whole_seq1_2	0.00	0.15	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_3	0.11	0.00	0.15	0.00	0.03	0.01	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.15	0.00	0.00	0.00	0.20	0.02	0.00	0.00	0.00	0.03

Chi Squared 2

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.19	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
whole_seq1_2	0.02	0.22	0.00	0.27	0.00	0.00	0.05	0.00	0.02	0.00
whole_seq1_3	0.00	0.00	0.18	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.10	0.00	0.03	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_5	0.00	0.06	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.02

Kullbeck Leibler

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.17	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.32	0.00	0.15	0.12	0.00	0.03	0.00	0.01	0.01
whole_seq1_3	0.01	0.00	0.16	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.08	0.00	0.20	0.08	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.00	0.24	0.00	0.00	0.07	0.00	0.03	0.00	0.00	0.00

Average Histogram Tracking Performance using S-layer in HSV Colour Space

Histogram Intersection

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.25	0.00	0.04	0.00	0.00	0.03	0.00	0.00	0.00	0.00
whole_seq1_2	0.02	0.22	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
whole_seq1_3	0.00	0.00	0.16	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00

Euclidean Intersection

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.06	0.06	0.10	0.13	0.13	0.00	0.01	0.01	0.01	0.01
whole_seq1_2	0.12	0.10	0.15	0.22	0.17	0.01	0.01	0.01	0.02	0.01
whole_seq1_3	0.12	0.05	0.14	0.08	0.18	0.00	0.00	0.01	0.00	0.01
whole_seq1_4	0.23	0.05	0.20	0.05	0.13	0.02	0.00	0.01	0.00	0.01
whole_seq1_5	0.17	0.11	0.22	0.17	0.19	0.01	0.01	0.01	0.01	0.01

Chi Squared 1

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.19	0.06	0.23	0.04	0.15	0.01	0.01	0.01	0.00	0.01
whole_seq1_2	0.02	0.17	0.00	0.32	0.02	0.00	0.01	0.00	0.04	0.00
whole_seq1_3	0.22	0.00	0.12	0.01	0.07	0.01	0.00	0.00	0.00	0.00
whole_seq1_4	0.32	0.07	0.05	0.00	0.13	0.03	0.00	0.00	0.00	0.01
whole_seq1_5	0.44	0.00	0.09	0.06	0.20	0.04	0.00	0.00	0.00	0.01

Chi Squared 2

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.25	0.88	0.25	0.13	0.15	0.02	0.25	0.02	0.02	0.01
whole_seq1_2	0.02	0.80	0.05	0.07	0.00	0.00	0.37	0.00	0.01	0.00
whole_seq1_3	0.22	0.67	0.18	0.00	0.15	0.01	0.12	0.01	0.00	0.01
whole_seq1_4	0.17	1.00	0.10	0.03	0.23	0.05	1.00	0.00	0.00	0.02
whole_seq1_5	0.48	0.96	0.24	0.06	0.26	0.09	0.54	0.03	0.00	0.02

Kullbeck Leibler

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.19	0.00	0.21	0.04	0.08	0.01	0.00	0.01	0.00	0.01
whole_seq1_2	0.05	0.17	0.02	0.24	0.05	0.00	0.01	0.00	0.02	0.00
whole_seq1_3	0.19	0.00	0.08	0.00	0.07	0.01	0.00	0.00	0.00	0.00
whole_seq1_4	0.37	0.12	0.07	0.00	0.15	0.03	0.02	0.00	0.00	0.01
whole_seq1_5	0.46	0.00	0.07	0.06	0.17	0.07	0.00	0.00	0.00	0.01

Average Histogram Tracking Performance using Cb-layer in YCbCr Colour Space

Histogram Intersection

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.19	0.00	0.00	0.23	0.00	0.01	0.00	0.00	0.01	0.00
whole_seq1_2	0.00	0.27	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
whole_seq1_3	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.12	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.01

Euclidean Intersection

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.21	0.13	0.15	0.33	0.23	0.01	0.01	0.01	0.03	0.02
whole_seq1_2	0.17	0.10	0.24	0.15	0.24	0.01	0.01	0.02	0.01	0.04
whole_seq1_3	0.14	0.09	0.19	0.22	0.20	0.00	0.00	0.01	0.01	0.01
whole_seq1_4	0.12	0.13	0.10	0.13	0.13	0.01	0.02	0.01	0.01	0.01
whole_seq1_5	0.17	0.17	0.15	0.33	0.17	0.01	0.01	0.01	0.03	0.01

Chi Squared 1

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.31	0.23	0.04	0.27	0.13	0.03	0.02	0.00	0.02	0.01
whole_seq1_2	0.12	0.15	0.10	0.24	0.12	0.01	0.02	0.01	0.02	0.01
whole_seq1_3	0.15	0.07	0.09	0.15	0.12	0.01	0.00	0.00	0.01	0.00
whole_seq1_4	0.17	0.10	0.20	0.22	0.15	0.01	0.01	0.01	0.03	0.01
whole_seq1_5	0.06	0.02	0.26	0.07	0.22	0.00	0.00	0.02	0.00	0.01

Chi Squared 2

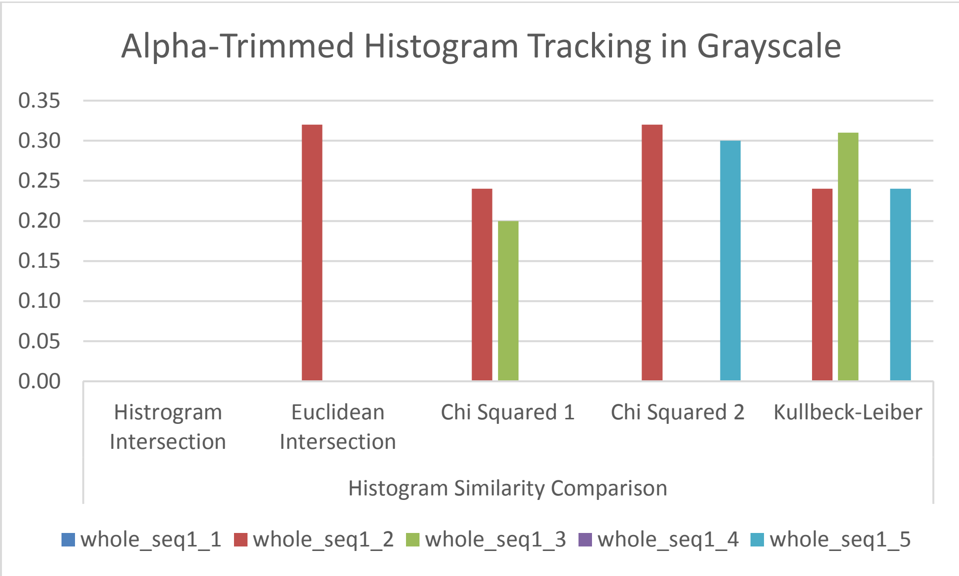
Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.35	0.15	0.13	0.40	0.13	0.03	0.02	0.01	0.04	0.01
whole_seq1_2	0.12	0.10	0.05	0.15	0.07	0.01	0.01	0.00	0.02	0.00
whole_seq1_3	0.14	0.00	0.18	0.16	0.16	0.01	0.00	0.01	0.01	0.01
whole_seq1_4	0.13	0.15	0.08	0.13	0.08	0.01	0.02	0.00	0.01	0.00
whole_seq1_5	0.09	0.00	0.31	0.09	0.33	0.00	0.00	0.02	0.00	0.02

Kullbeck Leibler

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.27	0.25	0.08	0.29	0.15	0.02	0.02	0.00	0.02	0.01
whole_seq1_2	0.17	0.17	0.15	0.27	0.10	0.02	0.02	0.01	0.03	0.01
whole_seq1_3	0.12	0.03	0.11	0.20	0.12	0.01	0.00	0.01	0.02	0.00
whole_seq1_4	0.13	0.10	0.23	0.23	0.17	0.01	0.01	0.01	0.03	0.01
whole_seq1_5	0.07	0.00	0.30	0.07	0.28	0.01	0.00	0.02	0.00	0.02

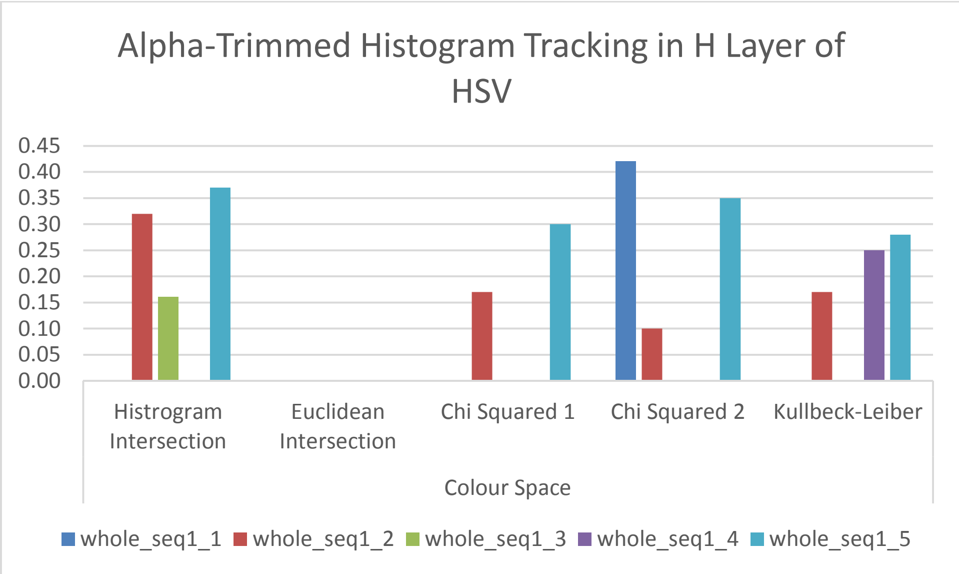
Average Histogram Tracking Performance using Cr-layer in YCbCr Colour Space

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.32	0.24	0.32	0.24
whole_seq1_3	0.00	0.00	0.20	0.00	0.31
whole_seq1_4	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.30	0.24



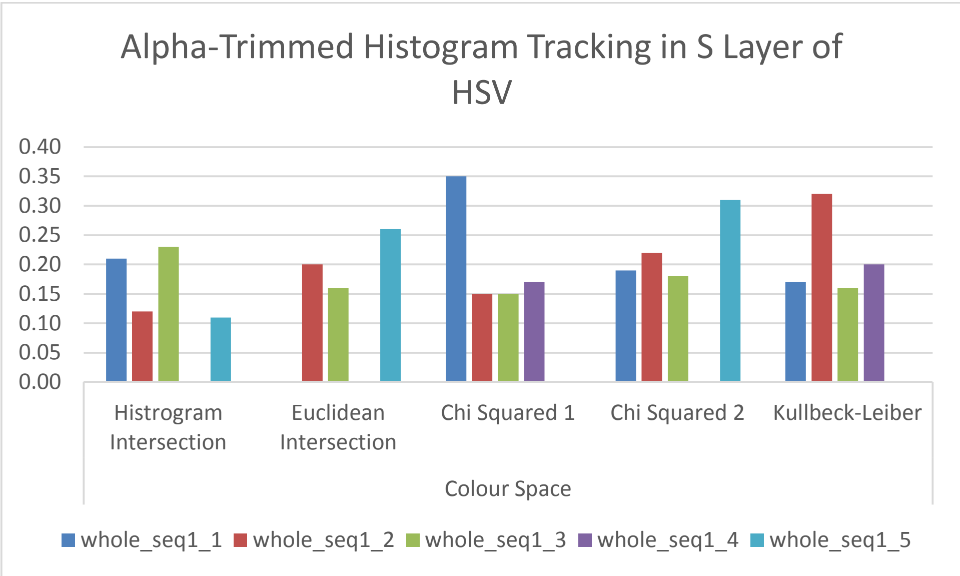
Summary of Alpha Trimmed Histogram Tracking in Grayscale Colour Space

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.00	0.00	0.00	0.42	0.00
whole_seq1_2	0.32	0.00	0.17	0.10	0.17
whole_seq1_3	0.16	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.00	0.25
whole_seq1_5	0.37	0.00	0.30	0.35	0.28



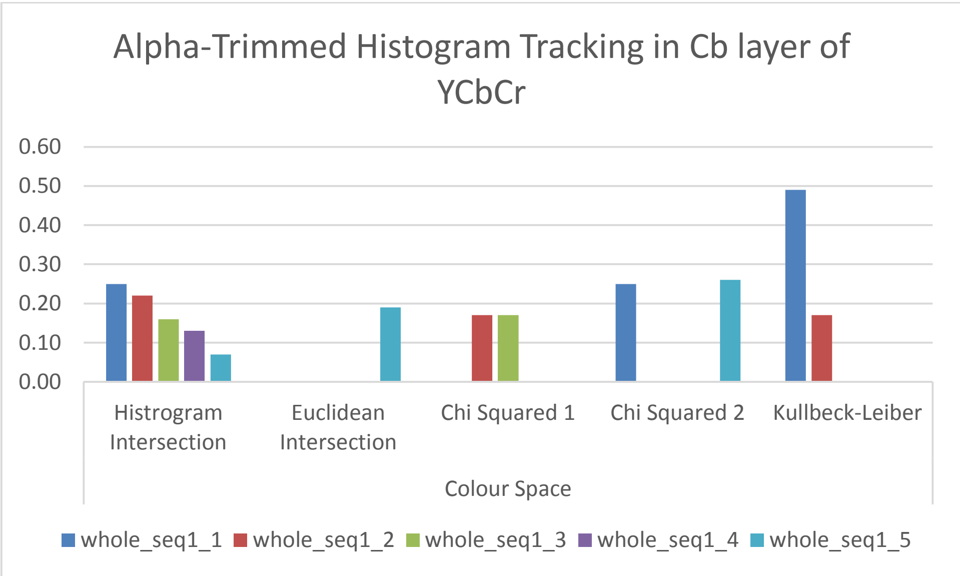
Summary of Alpha Trimmed Histogram Tracking in H Layer of HSV Colour Space

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.21	0.00	0.35	0.19	0.17
whole_seq1_2	0.12	0.20	0.15	0.22	0.32
whole_seq1_3	0.23	0.16	0.15	0.18	0.16
whole_seq1_4	0.00	0.00	0.17	0.00	0.20
whole_seq1_5	0.11	0.26	0.00	0.31	0.00



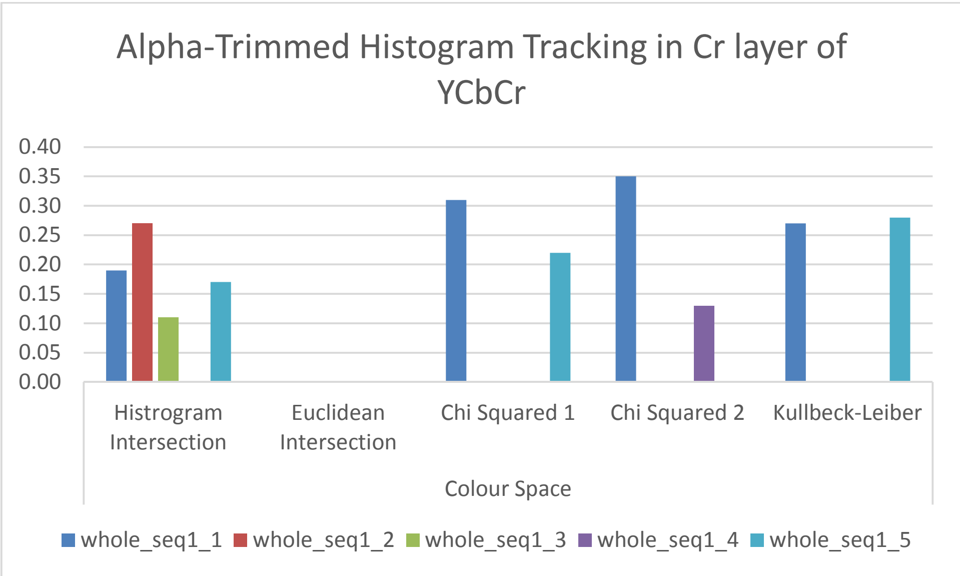
Summary of Alpha Trimmed Histogram Tracking in S Layer of HSV Colour Space

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.25	0.00	0.00	0.25	0.49
whole_seq1_2	0.22	0.00	0.17	0.00	0.17
whole_seq1_3	0.16	0.00	0.17	0.00	0.00
whole_seq1_4	0.13	0.00	0.00	0.00	0.00
whole_seq1_5	0.07	0.19	0.00	0.26	0.00



Summary of Alpha-Trimmed Histogram Tracking in Cb Layer of YCbCr Colour Space

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.19	0.00	0.31	0.35	0.27
whole_seq1_2	0.27	0.00	0.00	0.00	0.00
whole_seq1_3	0.11	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.13	0.00
whole_seq1_5	0.17	0.00	0.22	0.00	0.28



Summary of Alpha-Trimmed Histogram Tracking in Cr Layer of YCbCr Colour Space

Appendix 5

This appendix contains the tabulated data for composite image histogram tracking performance using different layers in different colour spaces.

Histogram Intersection

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.04	0.00	0.01	0.00	0.50	0.00	0.00	0.00	0.00	0.09
whole_seq1_2	0.00	0.22	0.18	0.00	0.00	0.00	0.02	0.01	0.00	0.00
whole_seq1_3	0.00	0.29	0.11	0.00	0.00	0.00	0.02	0.01	0.00	0.00
whole_seq1_4	0.00	0.12	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_5	0.02	0.00	0.09	0.13	0.02	0.00	0.00	0.00	0.01	0.00

Euclidean Intersection

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.15	0.14	0.10	0.11	0.00	0.01	0.01	0.00	0.01
whole_seq1_3	0.00	0.12	0.14	0.10	0.09	0.00	0.01	0.01	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.07	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00

Chi Squared 1

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.37	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00
whole_seq1_3	0.00	0.24	0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00
whole_seq1_4	0.00	0.12	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Chi Squared 2

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_2	0.00	0.32	0.00	0.23	0.04	0.00	0.03	0.00	0.03	0.00
whole_seq1_3	0.00	0.15	0.09	0.03	0.11	0.00	0.01	0.01	0.00	0.01
whole_seq1_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Kullbeck Leibler

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.22	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_3	0.00	0.17	0.03	0.00	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_4	0.00	0.12	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_5	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Composite Histogram Tracking Performance using Grayscale Colour Space

Histogram Intersection

	TDR					TC				
	Input Video					Input Video				
Reference Histogram	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.17	0.27	0.00	0.20	0.02	0.01	0.02	0.00	0.01	0.00
whole_seq1_2	0.00	0.24	0.12	0.15	0.28	0.00	0.01	0.01	0.01	0.01
whole_seq1_3	0.00	0.20	0.24	0.12	0.07	0.00	0.01	0.01	0.00	0.00
whole_seq1_4	0.00	0.22	0.01	0.23	0.04	0.00	0.01	0.00	0.01	0.00
whole_seq1_5	0.00	0.15	0.16	0.37	0.24	0.00	0.01	0.01	0.02	0.01

Euclidean Intersection

	TDR					TC				
	Input Video					Input Video				
Reference Histogram	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.19	0.20	0.00	0.25	0.04	0.01	0.01	0.00	0.03	0.00
whole_seq1_2	0.00	0.20	0.00	0.03	0.06	0.00	0.02	0.00	0.00	0.00
whole_seq1_3	0.00	0.27	0.18	0.18	0.26	0.00	0.05	0.01	0.02	0.02
whole_seq1_4	0.00	0.15	0.00	0.23	0.20	0.00	0.01	0.00	0.02	0.01
whole_seq1_5	0.00	0.12	0.14	0.27	0.20	0.00	0.01	0.01	0.04	0.02

Chi Squared 1

	TDR					TC				
	Input Video					Input Video				
Reference Histogram	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.23	0.54	0.01	0.17	0.00	0.02	0.09	0.00	0.01	0.00
whole_seq1_2	0.00	0.20	0.00	0.17	0.00	0.00	0.01	0.00	0.01	0.00
whole_seq1_3	0.00	0.00	0.14	0.00	0.11	0.00	0.00	0.01	0.00	0.01
whole_seq1_4	0.00	0.07	0.00	0.22	0.00	0.00	0.00	0.00	0.02	0.00
whole_seq1_5	0.00	0.32	0.07	0.42	0.22	0.00	0.04	0.00	0.03	0.01

Chi Squared 2

	TDR					TC				
	Input Video					Input Video				
Reference Histogram	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.27	0.12	0.00	0.15	0.00	0.02	0.01	0.00	0.01	0.00
whole_seq1_2	0.00	0.17	0.04	0.33	0.02	0.00	0.01	0.00	0.02	0.00
whole_seq1_3	0.15	0.20	0.24	0.08	0.44	0.01	0.01	0.02	0.00	0.04
whole_seq1_4	0.00	0.15	0.05	0.18	0.07	0.00	0.01	0.00	0.01	0.00
whole_seq1_5	0.08	0.00	0.09	0.00	0.15	0.01	0.00	0.01	0.00	0.01

Kullbeck Leibler

	TDR					TC				
	Input Video					Input Video				
Reference Histogram	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.21	0.05	0.00	0.05	0.00	0.01	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.34	0.00	0.25	0.00	0.00	0.03	0.00	0.01	0.00
whole_seq1_3	0.00	0.37	0.30	0.10	0.19	0.00	0.04	0.02	0.00	0.01
whole_seq1_4	0.00	0.07	0.00	0.22	0.00	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.00	0.37	0.07	0.07	0.20	0.00	0.05	0.00	0.00	0.01

Composite Histogram Tracking Performance using H-layer in HSV Colour Space

Histogram Intersection

	TDR					TC				
Reference Histogram	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Euclidean Intersection

	TDR					TC				
Reference Histogram	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.00	0.29	0.05	0.05	0.07	0.00	0.05	0.00	0.00	0.00
whole_seq1_2	0.00	0.00	0.09	0.18	0.00	0.00	0.00	0.00	0.02	0.00
whole_seq1_3	0.10	0.00	0.08	0.00	0.26	0.01	0.00	0.00	0.00	0.02
whole_seq1_4	0.04	0.00	0.00	0.08	0.39	0.00	0.00	0.00	0.00	0.03
whole_seq1_5	0.31	0.00	0.03	0.00	0.19	0.02	0.00	0.00	0.00	0.01

Chi Squared 1

	TDR					TC				
Reference Histogram	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.00	0.07	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Chi Squared 2

	TDR					TC				
Reference Histogram	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.00	0.35	0.43	0.19	0.00	0.00	0.04	0.08	0.02
whole_seq1_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.10	0.08	0.08	0.00	0.00	0.01	0.00	0.01	0.00

Kullbeck Leibler

	TDR					TC				
Reference Histogram	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.04	0.00	0.18	0.00	0.33	0.00	0.00	0.01	0.00	0.03
whole_seq1_3	0.00	0.05	0.00	0.15	0.13	0.00	0.00	0.00	0.01	0.01
whole_seq1_4	0.13	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.03	0.10	0.13	0.00	0.00	0.00	0.00	0.01

Composite Histogram Tracking Performance using S-layer in HSV Colour Space

Histogram Intersection

	TDR					TC				
	Input Video					Input Video				
Reference Histogram	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.31	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.17	0.00	0.12	0.00	0.00	0.01	0.00	0.01	0.00
whole_seq1_3	0.02	0.10	0.19	0.02	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.00	0.00	0.00	0.05	0.24	0.00	0.00	0.00	0.00	0.01

Euclidean Intersection

	TDR					TC				
	Input Video					Input Video				
Reference Histogram	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.29	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.12	0.00	0.13	0.17	0.00	0.01	0.00	0.01	0.02
whole_seq1_3	0.00	0.07	0.16	0.60	0.30	0.00	0.01	0.01	0.13	0.03
whole_seq1_4	0.04	0.12	0.41	0.13	0.17	0.00	0.01	0.03	0.01	0.01
whole_seq1_5	0.00	0.20	0.14	0.03	0.15	0.00	0.02	0.01	0.00	0.02

Chi Squared 1

	TDR					TC				
	Input Video					Input Video				
Reference Histogram	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.17	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.24	0.00	0.67	0.00	0.00	0.05	0.00	0.45	0.00
whole_seq1_3	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.05	0.00	0.25	0.00	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.00	0.00	0.00	0.17	0.28	0.00	0.00	0.00	0.01	0.02

Chi Squared 2

	TDR					TC				
	Input Video					Input Video				
Reference Histogram	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.27	0.00	0.00	0.17	0.00	0.02	0.00	0.00	0.01	0.00
whole_seq1_2	0.00	0.27	0.00	0.20	0.00	0.00	0.03	0.00	0.01	0.00
whole_seq1_3	0.00	0.02	0.09	0.07	0.06	0.00	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.05	0.00	0.18	0.00	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.00	0.10	0.05	0.03	0.28	0.00	0.01	0.00	0.00	0.02

Kullbeck Leibler

	TDR					TC				
	Input Video					Input Video				
Reference Histogram	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.29	0.00	0.00	0.00	0.28	0.02	0.00	0.00	0.00	0.03
whole_seq1_2	0.00	0.29	0.00	0.18	0.00	0.00	0.02	0.00	0.02	0.00
whole_seq1_3	0.00	0.00	0.08	0.00	0.13	0.00	0.00	0.00	0.00	0.01
whole_seq1_4	0.00	0.02	0.00	0.18	0.00	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.00	0.00	0.04	0.00	0.22	0.00	0.00	0.00	0.00	0.01

Composite Histogram Tracking Performance using Cb-layer in YCbCr Colour Space

Histogram Intersection

	TDR					TC				
Reference Histogram	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.21	0.00	0.12	0.00	0.19	0.01	0.00	0.01	0.00	0.01
whole_seq1_2	0.00	0.27	0.00	0.27	0.00	0.00	0.02	0.00	0.01	0.00
whole_seq1_3	0.00	0.00	0.20	0.00	0.09	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.07	0.00	0.17	0.00	0.00	0.00	0.00	0.01	0.00
whole_seq1_5	0.00	0.00	0.00	0.05	0.30	0.00	0.00	0.00	0.00	0.01

Euclidean Intersection

	TDR					TC				
Reference Histogram	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.23	0.00	0.05	0.00	0.28	0.03	0.00	0.00	0.00	0.03
whole_seq1_2	0.00	0.12	0.00	0.35	0.28	0.00	0.01	0.00	0.04	0.07
whole_seq1_3	0.00	0.41	0.11	0.67	0.19	0.00	0.12	0.01	0.41	0.03
whole_seq1_4	0.00	0.27	0.00	0.18	0.19	0.00	0.02	0.00	0.01	0.02
whole_seq1_5	0.00	0.15	0.27	0.02	0.15	0.00	0.01	0.02	0.00	0.01

Chi Squared 1

	TDR					TC				
Reference Histogram	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.15	0.10	0.00	0.20	0.00	0.01	0.01	0.00	0.01	0.00
whole_seq1_2	0.00	0.12	0.00	0.50	0.00	0.00	0.01	0.00	0.11	0.00
whole_seq1_3	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq1_4	0.00	0.29	0.00	0.08	0.00	0.00	0.03	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.38	0.00	0.17	0.00	0.00	0.03	0.00	0.01

Chi Squared 2

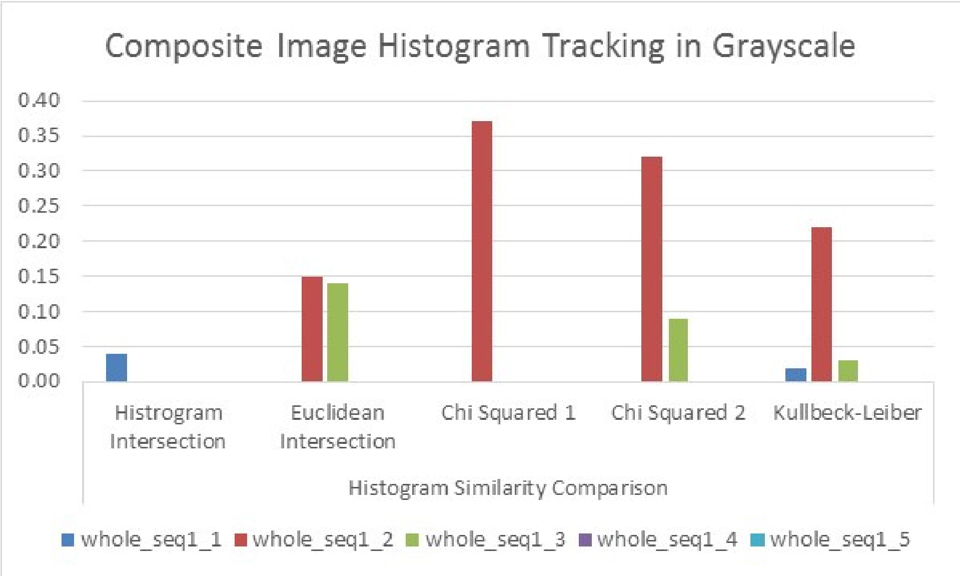
	TDR					TC				
Reference Histogram	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.17	0.02	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.22	0.00	0.12	0.00	0.00	0.01	0.00	0.01	0.00
whole_seq1_3	0.00	0.17	0.16	0.05	0.31	0.00	0.02	0.01	0.00	0.02
whole_seq1_4	0.06	0.20	0.00	0.18	0.00	0.00	0.02	0.00	0.01	0.00
whole_seq1_5	0.00	0.10	0.36	0.02	0.17	0.00	0.01	0.02	0.00	0.01

Kullbeck Leibler

	TDR					TC				
Reference Histogram	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.23	0.10	0.26	0.08	0.17	0.01	0.00	0.01	0.00	0.01
whole_seq1_2	0.00	0.12	0.00	0.03	0.00	0.00	0.01	0.00	0.00	0.00
whole_seq1_3	0.00	0.00	0.20	0.00	0.22	0.00	0.00	0.01	0.00	0.02
whole_seq1_4	0.00	0.32	0.00	0.07	0.00	0.00	0.03	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.27	0.00	0.20	0.00	0.00	0.01	0.00	0.02

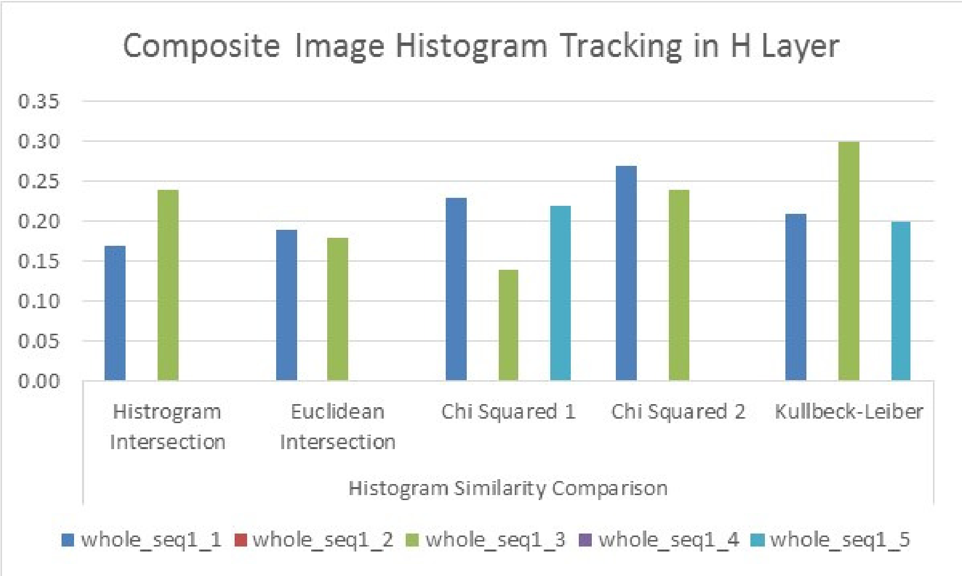
Composite Histogram Tracking Performance using Cr-layer in YCbCr Colour Space

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.04	0.00	0.00	0.00	0.02
whole_seq1_2	0.00	0.15	0.37	0.32	0.22
whole_seq1_3	0.00	0.14	0.00	0.09	0.03
whole_seq1_4	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.00



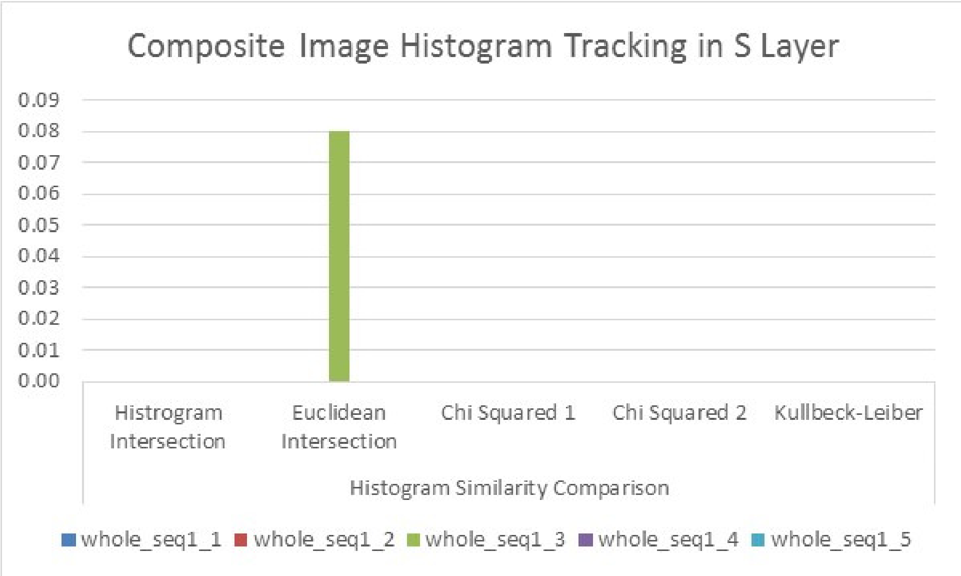
Summary of Composite Image Histogram Tracking in Grayscale Colour Space

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.17	0.19	0.23	0.27	0.21
whole_seq1_2	0.00	0.00	0.00	0.00	0.00
whole_seq1_3	0.24	0.18	0.14	0.24	0.30
whole_seq1_4	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.22	0.00	0.20



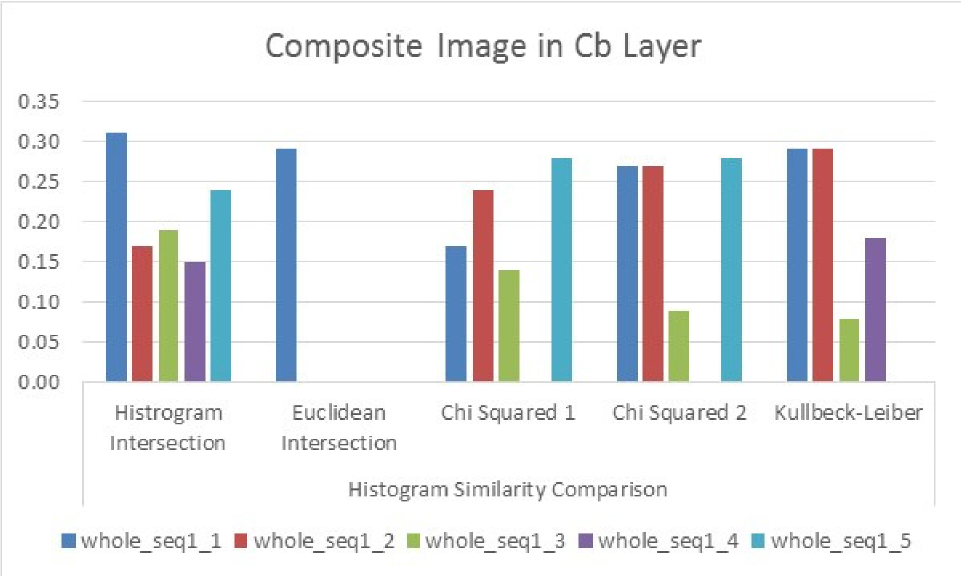
Summary of Composite Image Histogram Tracking in H Layer of HSV Colour Space

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.00	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.00	0.00	0.00	0.00
whole_seq1_3	0.00	0.08	0.00	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.00	0.00
whole_seq1_5	0.00	0.00	0.00	0.00	0.00



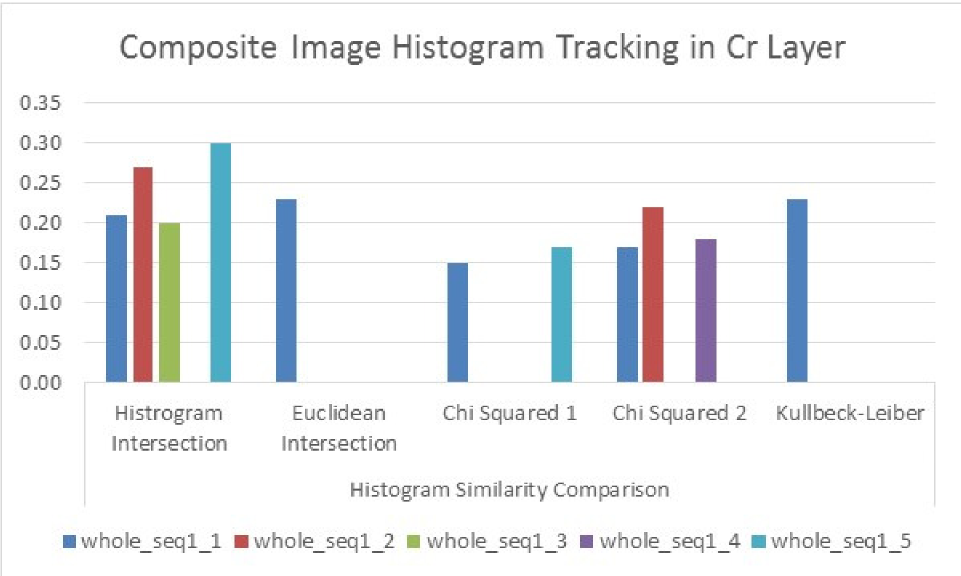
Summary of Composite Image Histogram Tracking in S Layer of HSV Colour Space

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.31	0.29	0.17	0.27	0.29
whole_seq1_2	0.17	0.00	0.24	0.27	0.29
whole_seq1_3	0.19	0.00	0.14	0.09	0.08
whole_seq1_4	0.15	0.00	0.00	0.00	0.18
whole_seq1_5	0.24	0.00	0.28	0.28	0.00



Summary of Composite Image Histogram Tracking in Cb Layer of YCbCr Colour Space

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.21	0.23	0.15	0.17	0.23
whole_seq1_2	0.27	0.00	0.00	0.22	0.00
whole_seq1_3	0.20	0.00	0.00	0.00	0.00
whole_seq1_4	0.00	0.00	0.00	0.18	0.00
whole_seq1_5	0.30	0.00	0.17	0.00	0.00



Summary of Composite Image Histogram Tracking in Cr Layer of YCbCr Colour Space

Appendix 6

This appendix contains the tabulated data for segmented image histogram tracking performance using different layers in different colour spaces.

Histogram Intersection

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.65	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.71	0.32	0.57	0.07	0.00	0.12	0.05	0.10	0.00
whole_seq1_3	0.02	0.07	0.57	0.03	0.72	0.00	0.01	0.06	0.00	0.03
whole_seq1_4	0.00	0.56	0.05	0.57	0.65	0.00	0.06	0.00	0.08	0.10
whole_seq1_5	0.00	0.17	0.15	0.20	0.63	0.00	0.01	0.01	0.01	0.08

Euclidean Intersection

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.69	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00
whole_seq1_2	0.06	0.73	0.24	0.30	0.11	0.00	0.11	0.06	0.02	0.01
whole_seq1_3	0.00	0.22	0.53	0.18	0.20	0.00	0.02	0.04	0.01	0.01
whole_seq1_4	0.17	0.37	0.10	0.62	0.35	0.02	0.04	0.00	0.07	0.06
whole_seq1_5	0.00	0.29	0.49	0.37	0.56	0.00	0.04	0.04	0.03	0.04

Chi Squared 1

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.73	0.00	0.00	0.00	0.00	0.18	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.66	0.09	0.25	0.02	0.00	0.10	0.01	0.02	0.00
whole_seq1_3	0.00	0.10	0.49	0.00	0.39	0.00	0.01	0.03	0.00	0.03
whole_seq1_4	0.00	0.20	0.00	0.53	0.37	0.00	0.01	0.00	0.10	0.10
whole_seq1_5	0.00	0.00	0.27	0.07	0.59	0.00	0.00	0.02	0.00	0.07

Chi Squared 2

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.60	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.73	0.23	0.42	0.00	0.00	0.13	0.02	0.03	0.00
whole_seq1_3	0.00	0.02	0.49	0.00	0.13	0.00	0.00	0.03	0.00	0.01
whole_seq1_4	0.00	0.27	0.00	0.63	0.52	0.00	0.03	0.00	0.09	0.08
whole_seq1_5	0.00	0.02	0.20	0.10	0.65	0.00	0.00	0.01	0.01	0.07

Kullbeck Leibler

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.56	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.68	0.15	0.42	0.00	0.00	0.13	0.01	0.04	0.00
whole_seq1_3	0.00	0.05	0.55	0.00	0.13	0.00	0.00	0.04	0.00	0.01
whole_seq1_4	0.00	0.24	0.00	0.65	0.35	0.00	0.02	0.00	0.11	0.08
whole_seq1_5	0.00	0.00	0.31	0.03	0.61	0.00	0.00	0.02	0.00	0.06

Segmented Histogram Tracking Performance using Grayscale Colour Space

Histogram Intersection

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.63	0.37	0.28	0.23	0.19	0.12	0.03	0.01	0.02	0.01
whole_seq1_2	0.15	0.63	0.04	0.47	0.15	0.01	0.09	0.00	0.04	0.01
whole_seq1_3	0.17	0.29	0.49	0.12	0.43	0.01	0.02	0.03	0.00	0.03
whole_seq1_4	0.08	0.29	0.03	0.62	0.06	0.01	0.02	0.00	0.08	0.00
whole_seq1_5	0.31	0.22	0.37	0.18	0.57	0.02	0.01	0.02	0.01	0.05

Euclidean Intersection

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.73	0.05	0.00	0.08	0.00	0.15	0.00	0.00	0.00	0.00
whole_seq1_2	0.38	0.56	0.03	0.42	0.20	0.03	0.10	0.00	0.03	0.04
whole_seq1_3	0.00	0.00	0.38	0.00	0.26	0.00	0.00	0.02	0.00	0.01
whole_seq1_4	0.42	0.42	0.08	0.63	0.13	0.04	0.04	0.00	0.14	0.01
whole_seq1_5	0.00	0.22	0.62	0.15	0.65	0.00	0.02	0.07	0.01	0.05

Chi Squared 1

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.69	0.02	0.00	0.03	0.00	0.09	0.00	0.00	0.00	0.00
whole_seq1_2	0.13	0.71	0.00	0.60	0.00	0.01	0.11	0.00	0.08	0.00
whole_seq1_3	0.00	0.00	0.47	0.00	0.20	0.00	0.00	0.03	0.00	0.02
whole_seq1_4	0.13	0.39	0.00	0.57	0.00	0.01	0.03	0.00	0.06	0.00
whole_seq1_5	0.00	0.00	0.20	0.00	0.69	0.00	0.00	0.01	0.00	0.12

Chi Squared 2

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.63	0.00	0.00	0.02	0.00	0.08	0.00	0.00	0.00	0.00
whole_seq1_2	0.19	0.63	0.04	0.50	0.02	0.01	0.08	0.00	0.03	0.00
whole_seq1_3	0.00	0.00	0.47	0.00	0.11	0.00	0.00	0.03	0.00	0.01
whole_seq1_4	0.17	0.46	0.01	0.63	0.00	0.02	0.04	0.00	0.09	0.00
whole_seq1_5	0.00	0.00	0.14	0.00	0.59	0.00	0.00	0.01	0.00	0.06

Kullbeck Leibler

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.67	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00
whole_seq1_2	0.33	0.73	0.04	0.72	0.00	0.03	0.17	0.00	0.07	0.00
whole_seq1_3	0.00	0.00	0.38	0.00	0.33	0.00	0.00	0.02	0.00	0.06
whole_seq1_4	0.23	0.41	0.00	0.48	0.00	0.01	0.03	0.00	0.06	0.00
whole_seq1_5	0.00	0.00	0.16	0.00	0.61	0.00	0.00	0.01	0.00	0.08

Segmented Histogram Tracking Performance using H-layer in HSV Colour Space

Histogram Intersection

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.58	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.71	0.00	0.48	0.00	0.00	0.12	0.00	0.07	0.00
whole_seq1_3	0.00	0.00	0.50	0.00	0.17	0.00	0.00	0.04	0.00	0.02
whole_seq1_4	0.00	0.51	0.00	0.62	0.00	0.00	0.04	0.00	0.08	0.00
whole_seq1_5	0.04	0.00	0.19	0.00	0.19	0.00	0.00	0.01	0.00	0.02

Euclidean Intersection

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.65	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00
whole_seq1_2	0.08	0.66	0.16	0.40	0.04	0.01	0.14	0.01	0.02	0.00
whole_seq1_3	0.00	0.02	0.50	0.02	0.43	0.00	0.00	0.03	0.00	0.02
whole_seq1_4	0.40	0.32	0.01	0.63	0.02	0.04	0.03	0.00	0.09	0.00
whole_seq1_5	0.00	0.02	0.50	0.03	0.61	0.00	0.00	0.04	0.00	0.05

Chi Squared 1

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.60	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.71	0.00	0.33	0.00	0.00	0.10	0.00	0.03	0.00
whole_seq1_3	0.00	0.00	0.54	0.00	0.20	0.00	0.00	0.04	0.00	0.02
whole_seq1_4	0.00	0.39	0.00	0.57	0.00	0.00	0.03	0.00	0.08	0.00
whole_seq1_5	0.00	0.00	0.11	0.00	0.63	0.00	0.00	0.00	0.00	0.11

Chi Squared 2

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.65	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.61	0.00	0.30	0.00	0.00	0.09	0.00	0.02	0.00
whole_seq1_3	0.00	0.00	0.54	0.00	0.02	0.00	0.00	0.05	0.00	0.00
whole_seq1_4	0.00	0.34	0.00	0.57	0.00	0.00	0.02	0.00	0.05	0.00
whole_seq1_5	0.00	0.00	0.22	0.00	0.61	0.00	0.00	0.02	0.00	0.09

Kullbeck Leibler

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.60	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00
whole_seq1_2	0.00	0.66	0.00	0.33	0.00	0.00	0.07	0.00	0.03	0.00
whole_seq1_3	0.00	0.00	0.54	0.00	0.06	0.00	0.00	0.04	0.00	0.00
whole_seq1_4	0.00	0.32	0.00	0.58	0.00	0.00	0.02	0.00	0.04	0.00
whole_seq1_5	0.00	0.00	0.07	0.00	0.61	0.00	0.00	0.00	0.00	0.05

Segmented Histogram Tracking Performance using S-layer in HSV Colour Space

Histogram Intersection

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.92	0.12	0.15	0.05	0.20	0.84	0.01	0.01	0.00	0.01
whole_seq1_2	0.38	0.98	0.53	0.92	0.96	0.03	0.51	0.04	0.27	0.48
whole_seq1_3	0.00	0.37	0.54	0.30	0.33	0.00	0.03	0.04	0.03	0.04
whole_seq1_4	0.15	0.24	0.23	0.25	0.30	0.01	0.02	0.01	0.01	0.02
whole_seq1_5	0.35	0.22	0.39	0.15	0.30	0.03	0.02	0.02	0.01	0.02

Euclidean Intersection

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.67	0.00	0.18	0.00	0.00	0.12	0.00	0.01	0.00	0.00
whole_seq1_2	0.17	0.71	0.10	0.65	0.22	0.01	0.28	0.00	0.10	0.05
whole_seq1_3	0.06	0.56	0.50	0.43	0.37	0.00	0.10	0.03	0.05	0.03
whole_seq1_4	0.19	0.39	0.23	0.67	0.09	0.02	0.04	0.01	0.09	0.01
whole_seq1_5	0.00	0.24	0.53	0.13	0.52	0.00	0.03	0.08	0.01	0.06

Chi Squared 1

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.85	0.12	0.01	0.00	0.00	0.64	0.01	0.00	0.00	0.00
whole_seq1_2	0.23	0.78	0.07	0.83	0.00	0.02	0.36	0.00	0.62	0.00
whole_seq1_3	0.00	0.10	0.54	0.03	0.06	0.00	0.01	0.05	0.00	0.00
whole_seq1_4	0.27	0.44	0.04	0.63	0.00	0.02	0.04	0.00	0.07	0.00
whole_seq1_5	0.00	0.00	0.11	0.00	0.70	0.00	0.00	0.00	0.00	0.07

Chi Squared 2

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.73	0.07	0.00	0.13	0.00	0.14	0.00	0.00	0.01	0.00
whole_seq1_2	0.15	0.71	0.04	0.52	0.00	0.01	0.31	0.00	0.07	0.00
whole_seq1_3	0.04	0.05	0.50	0.02	0.15	0.00	0.00	0.04	0.00	0.01
whole_seq1_4	0.19	0.22	0.50	0.32	0.00	0.01	0.02	0.04	0.02	0.00
whole_seq1_5	0.10	0.00	0.16	0.00	0.65	0.01	0.00	0.01	0.00	0.08

Kullbeck Leibler

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.81	0.05	0.00	0.05	0.00	0.57	0.00	0.00	0.00	0.00
whole_seq1_2	0.15	0.71	0.14	0.80	0.00	0.01	0.27	0.01	0.33	0.00
whole_seq1_3	0.06	0.07	0.53	0.02	0.09	0.00	0.00	0.04	0.00	0.00
whole_seq1_4	0.15	0.44	0.19	0.60	0.00	0.01	0.07	0.01	0.04	0.00
whole_seq1_5	0.00	0.02	0.23	0.00	0.61	0.00	0.00	0.01	0.00	0.05

Segmented Histogram Tracking Performance using Cb-layer in YCbCr Colour Space

Histogram Intersection

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.60	0.02	0.04	0.05	0.13	0.10	0.00	0.00	0.00	0.01
whole_seq1_2	0.02	0.73	0.35	0.85	0.83	0.00	0.12	0.02	0.16	0.22
whole_seq1_3	0.06	0.00	0.26	0.00	0.41	0.00	0.00	0.01	0.00	0.03
whole_seq1_4	0.17	0.17	0.00	0.20	0.06	0.02	0.01	0.00	0.01	0.00
whole_seq1_5	0.13	0.00	0.32	0.03	0.54	0.01	0.00	0.01	0.00	0.06

Euclidean Intersection

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.65	0.00	0.00	0.10	0.00	0.09	0.00	0.00	0.01	0.00
whole_seq1_2	0.46	0.68	0.10	0.68	0.19	0.03	0.10	0.01	0.01	0.02
whole_seq1_3	0.00	0.39	0.46	0.35	0.22	0.00	0.06	0.02	0.03	0.01
whole_seq1_4	0.08	0.42	0.19	0.52	0.11	0.01	0.04	0.01	0.05	0.01
whole_seq1_5	0.04	0.98	0.97	1.00	0.78	0.00	0.95	0.49	1.00	0.25

Chi Squared 1

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.58	0.02	0.00	0.08	0.00	0.09	0.00	0.00	0.00	0.00
whole_seq1_2	0.19	0.66	0.00	0.60	0.00	0.01	0.10	0.00	0.05	0.00
whole_seq1_3	0.00	0.00	0.46	0.00	0.09	0.00	0.00	0.03	0.00	0.00
whole_seq1_4	0.10	0.24	0.00	0.60	0.02	0.01	0.02	0.00	0.07	0.00
whole_seq1_5	0.00	0.00	0.16	0.00	0.59	0.00	0.00	0.01	0.00	0.05

Chi Squared 2

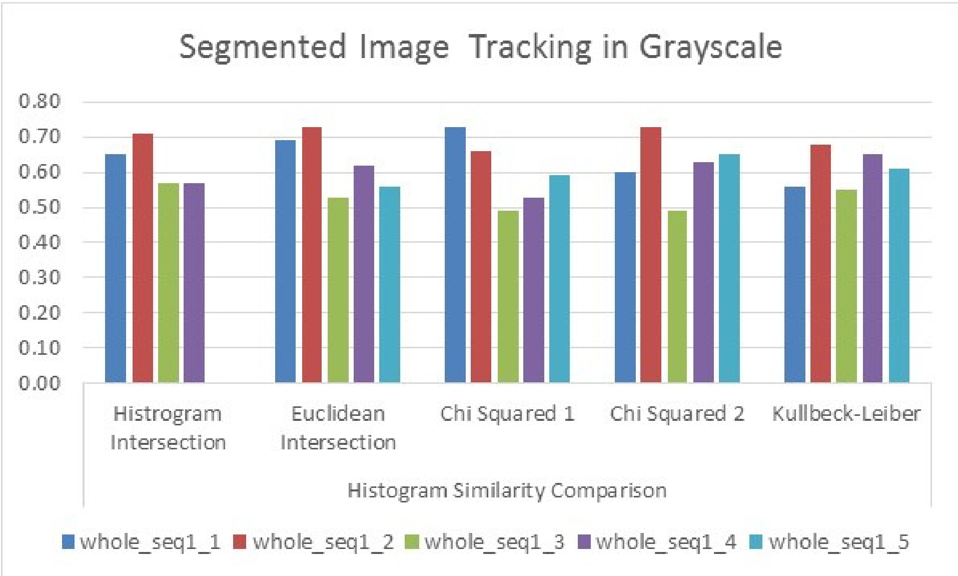
Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.65	0.05	0.00	0.08	0.00	0.12	0.00	0.00	0.00	0.00
whole_seq1_2	0.06	0.80	0.01	0.50	0.00	0.00	0.20	0.00	0.03	0.00
whole_seq1_3	0.00	0.00	0.50	0.00	0.15	0.00	0.00	0.03	0.00	0.01
whole_seq1_4	0.15	0.56	0.04	0.58	0.00	0.01	0.07	0.00	0.05	0.00
whole_seq1_5	0.04	0.00	0.14	0.00	0.61	0.00	0.00	0.00	0.00	0.06

Kullbeck Leibler

Reference Histogram	TDR					TC				
	Input Video					Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.65	0.12	0.00	0.08	0.00	0.09	0.01	0.00	0.00	0.00
whole_seq1_2	0.17	0.68	0.04	0.57	0.00	0.01	0.11	0.00	0.05	0.00
whole_seq1_3	0.00	0.00	0.49	0.00	0.07	0.00	0.00	0.06	0.00	0.00
whole_seq1_4	0.13	0.37	0.03	0.65	0.00	0.01	0.03	0.00	0.13	0.00
whole_seq1_5	0.00	0.00	0.36	0.00	0.59	0.00	0.00	0.02	0.00	0.06

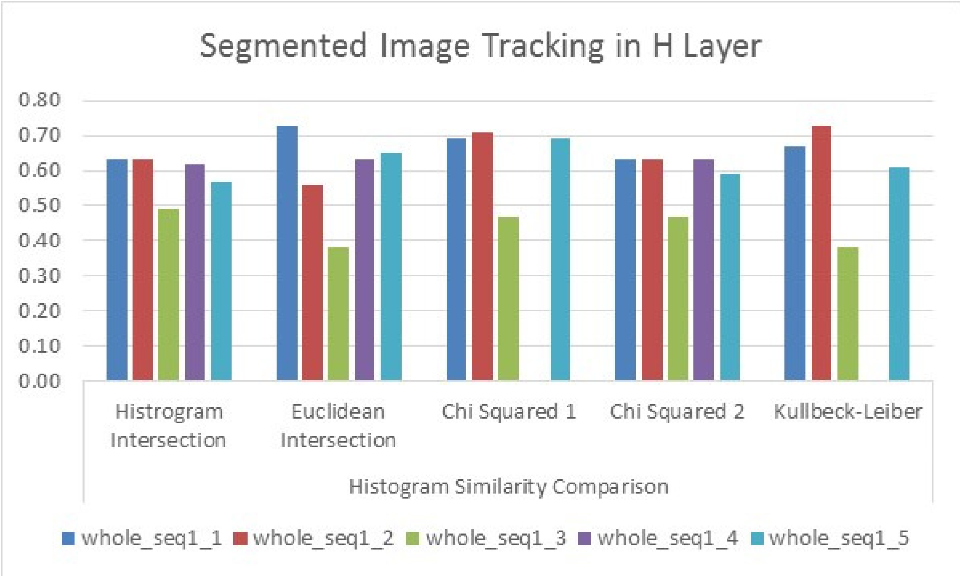
Segmented Histogram Tracking Performance using Cr-layer in YCbCr Colour Space

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.65	0.69	0.73	0.60	0.56
whole_seq1_2	0.71	0.73	0.66	0.73	0.68
whole_seq1_3	0.57	0.53	0.49	0.49	0.55
whole_seq1_4	0.57	0.62	0.53	0.63	0.65
whole_seq1_5	0.00	0.56	0.59	0.65	0.61



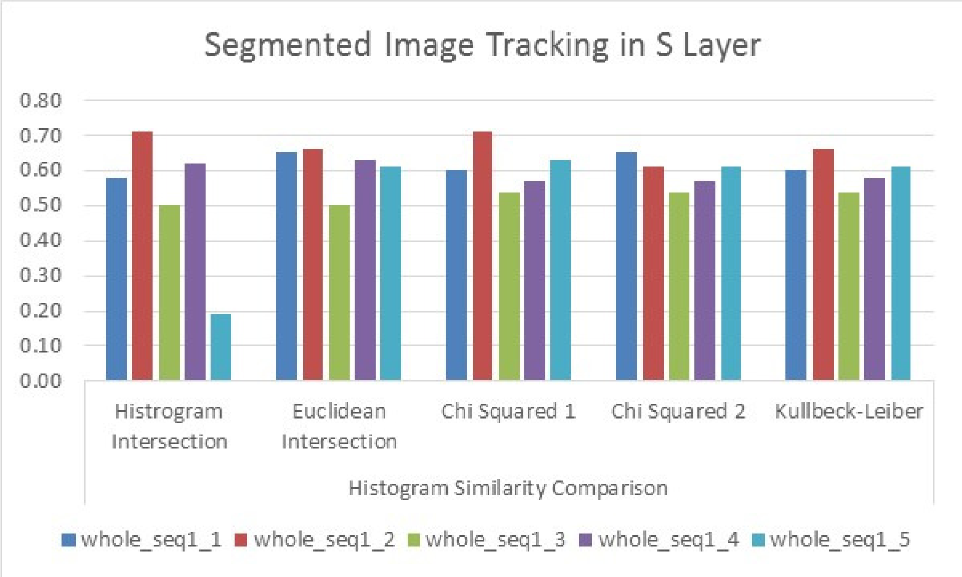
Summary of Segmented Image Histogram Tracking in Grayscale

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.63	0.73	0.69	0.63	0.67
whole_seq1_2	0.63	0.56	0.71	0.63	0.73
whole_seq1_3	0.49	0.38	0.47	0.47	0.38
whole_seq1_4	0.62	0.63	0.00	0.63	0.00
whole_seq1_5	0.57	0.65	0.69	0.59	0.61



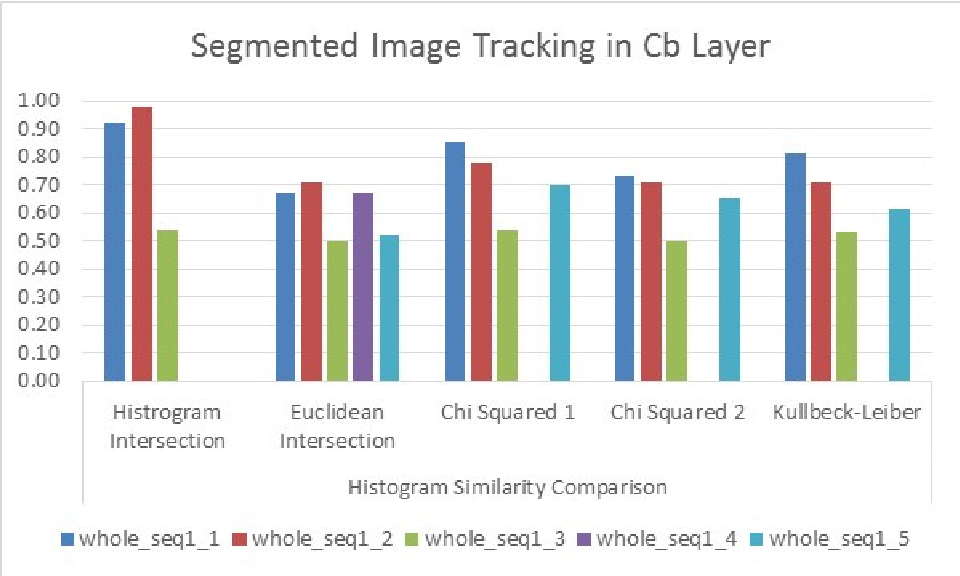
Summary of Segmented Image Histogram Tracking in H layer of HSV Colour Space

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.58	0.65	0.60	0.65	0.60
whole_seq1_2	0.71	0.66	0.71	0.61	0.66
whole_seq1_3	0.50	0.50	0.54	0.54	0.54
whole_seq1_4	0.62	0.63	0.57	0.57	0.58
whole_seq1_5	0.19	0.61	0.63	0.61	0.61



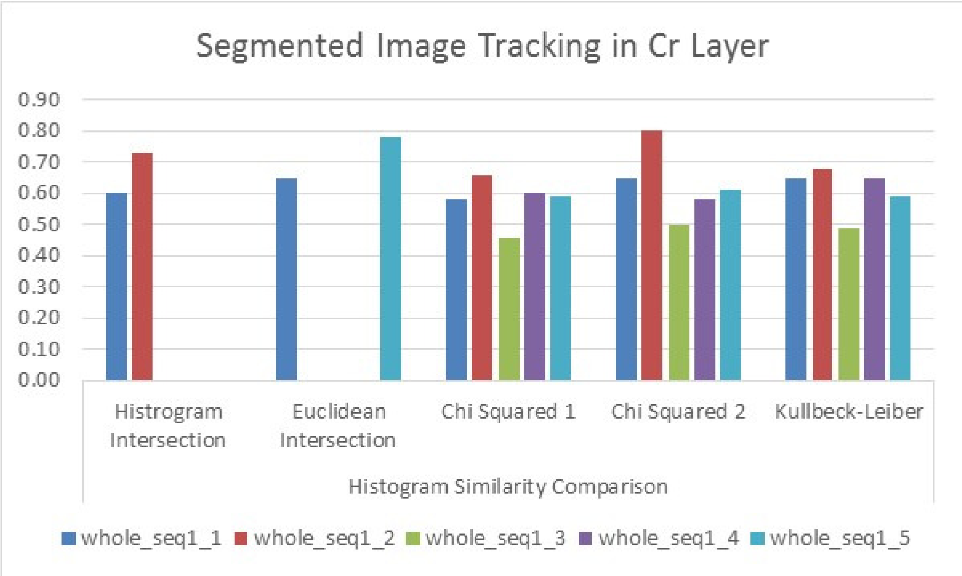
Summary of Segmented Image Histogram Tracking in S Layer of HSV Colour Space

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.92	0.67	0.85	0.73	0.81
whole_seq1_2	0.98	0.71	0.78	0.71	0.71
whole_seq1_3	0.54	0.50	0.54	0.50	0.53
whole_seq1_4	0.00	0.67	0.00	0.00	0.00
whole_seq1_5	0.00	0.52	0.70	0.65	0.61



Summary of Segmented Image Histogram Tracking in Cb Layer of YCbCr Colour Space

Input Video	Histogram Similarity Comparison				
	Histogram Intersection	Euclidean Intersection	Chi Squared 1	Chi Squared 2	Kullbeck-Leibler
whole_seq1_1	0.60	0.65	0.58	0.65	0.65
whole_seq1_2	0.73	0.00	0.66	0.80	0.68
whole_seq1_3	0.00	0.00	0.46	0.50	0.49
whole_seq1_4	0.00	0.00	0.60	0.58	0.65
whole_seq1_5	0.00	0.78	0.59	0.61	0.59



Summary of Segmented Image Histogram Tracking in Cr Layer of YCbCr Colour Space

Appendix 7

This appendix contains the tabulated data for combinational vector tracking performance using on each colour information layer from the YCbCr colour space.

Tracking Performance in Cr colour layer

Reference Histogram	TDR Input Video whole_seq1_1	TDR whole_seq1_2	TDR whole_seq1_3	TDR whole_seq1_4	TDR whole_seq1_5	TC Input Video whole_seq1_1	TC whole_seq1_2	TC whole_seq1_3	TC whole_seq1_4	TC whole_seq1_5
whole_seq1_1	0.58	0.23	0.04	0.40	0.08	0.09	0.03	0.00	0.06	0.01
whole_seq1_2	0.78	0.78	0.29	0.71	0.00	0.42	0.42	0.05	0.38	0.00
whole_seq1_3	1.00	0.74	0.93	1.00	0.09	1.00	0.18	0.59	1.00	0.00
whole_seq1_4	0.95	0.80	0.65	0.95	0.00	0.54	0.29	0.25	0.36	0.00
whole_seq1_5	0.94	0.76	0.81	1.00	0.06	0.37	0.33	0.34	1.00	0.00

Tracking Performance in Cb colour layer

Reference Histogram	TDR Input Video whole_seq1_1	TDR whole_seq1_2	TDR whole_seq1_3	TDR whole_seq1_4	TDR whole_seq1_5	TC Input Video whole_seq1_1	TC whole_seq1_2	TC whole_seq1_3	TC whole_seq1_4	TC whole_seq1_5
whole_seq1_1	0.65	0.10	0.42	0.81	0.00	0.09	0.01	0.08	0.27	0.00
whole_seq1_2	0.34	0.34	0.10	0.20	0.05	0.03	0.02	0.01	0.02	0.00
whole_seq1_3	0.77	0.53	0.18	0.91	0.05	0.15	0.04	0.03	0.48	0.00
whole_seq1_4	0.72	0.75	0.00	0.78	0.00	0.22	0.37	0.00	0.24	0.00
whole_seq1_5	1.00	0.09	0.22	1.00	0.06	1.00	0.00	0.05	1.00	0.00

All Vector Combination Tracking Performance

Appendix 8

This appendix contains the tabulated data for maximum vector tracking performance using Cb and Cr layers in YCbCr colour space.

Tracking Performance in Cr colour layer

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.69	0.00	0.42	0.00	0.04	0.21	0.00	0.08	0.00	0.00
whole_seq1_2	0.05	0.44	0.10	0.00	0.02	0.00	0.08	0.01	0.00	0.00
whole_seq1_3	0.30	0.07	0.92	0.00	0.12	0.02	0.00	0.34	0.00	0.00
whole_seq1_4	0.60	0.20	0.30	0.00	0.15	0.10	0.01	0.02	0.00	0.01
whole_seq1_5	0.61	0.00	0.37	0.37	0.98	0.17	0.00	0.04	0.12	0.83

Tracking Performance in Cb colour layer

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.83	1.00	0.00	0.10	0.00	0.31	1.00	0.00	0.01	0.00
whole_seq1_2	0.00	0.41	0.10	0.05	0.00	0.00	0.04	0.01	0.00	0.00
whole_seq1_3	0.00	0.78	0.50	0.04	0.15	0.00	0.39	0.09	0.00	0.01
whole_seq1_4	0.05	0.85	0.32	0.82	0.10	0.00	0.57	0.03	0.57	0.01
whole_seq1_5	0.00	0.98	0.02	0.48	0.83	0.00	0.96	0.00	0.18	0.28

Absolute Parameter Tracking Performance in Cr colour layer

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.94	0.40	0.06	0.35	0.04	0.77	0.07	0.00	0.04	0.00
whole_seq1_2	0.59	0.71	0.46	0.00	0.07	0.10	0.31	0.08	0.00	0.00
whole_seq1_3	0.51	0.41	0.91	0.00	0.19	0.05	0.03	0.32	0.00	0.01
whole_seq1_4	0.43	0.83	0.35	0.02	0.20	0.05	0.26	0.03	0.00	0.01
whole_seq1_5	0.15	0.13	0.30	0.39	0.94	0.01	0.01	0.04	0.12	0.43

Absolute Parameter Tracking Performance in Cb colour layer

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.98	1.00	0.08	0.13	0.00	0.52	1.00	0.01	0.01	0.00
whole_seq1_2	0.02	0.41	0.15	0.12	0.02	0.00	0.04	0.01	0.01	0.00
whole_seq1_3	0.01	0.66	0.55	0.30	0.22	0.00	0.26	0.15	0.05	0.01
whole_seq1_4	0.15	0.78	0.28	0.52	0.30	0.01	0.57	0.02	0.06	0.02
whole_seq1_5	0.00	0.94	0.00	0.46	1.00	0.00	0.89	0.00	0.22	1.00

Maximum Vector Tracking Performance in individual Cb and Cr colour layer

Tracking Performance in Cr colour layer

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.81	0.00	0.42	0.63	0.04	0.25	0.00	0.08	0.13	0.00
whole_seq1_2	0.15	0.51	0.10	0.27	0.02	0.02	0.11	0.01	0.03	0.00
whole_seq1_3	0.30	0.07	0.92	0.28	0.12	0.02	0.00	0.34	0.01	0.00
whole_seq1_4	0.62	0.18	0.35	0.88	0.15	0.11	0.01	0.02	0.36	0.01
whole_seq1_5	0.61	0.00	0.52	0.78	0.98	0.17	0.00	0.07	0.29	0.83

Tracking Performance in Cb colour layer

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.90	1.00	0.19	0.27	0.04	0.31	1.00	0.01	0.03	0.00
whole_seq1_2	0.00	0.90	0.27	0.32	0.02	0.00	0.29	0.05	0.04	0.00
whole_seq1_3	0.00	0.51	0.93	0.14	0.19	0.00	0.11	0.54	0.01	0.01
whole_seq1_4	0.12	0.92	0.32	0.90	0.17	0.01	0.70	0.03	0.70	0.01
whole_seq1_5	0.02	0.91	0.19	0.69	0.91	0.00	0.51	0.02	0.23	0.32

Absolute Parameter Tracking Performance in Cr colour layer

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.94	0.48	0.06	0.69	0.04	0.77	0.08	0.00	0.13	0.00
whole_seq1_2	0.73	0.95	0.54	0.66	0.17	0.15	0.86	0.08	0.21	0.01
whole_seq1_3	0.51	0.41	0.91	0.42	0.19	0.05	0.03	0.32	0.03	0.01
whole_seq1_4	0.45	0.83	0.38	0.95	0.25	0.06	0.26	0.03	0.50	0.01
whole_seq1_5	0.17	0.13	0.44	0.83	0.94	0.01	0.01	0.06	0.28	0.43

Absolute Parameter Tracking Performance in Cb colour layer

Reference Histogram	TDR Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5	TC Input Video whole_seq1_1	whole_seq1_2	whole_seq1_3	whole_seq1_4	whole_seq1_5
whole_seq1_1	0.98	1.00	0.31	0.19	0.06	0.52	1.00	0.04	0.02	0.01
whole_seq1_2	0.24	0.90	0.51	0.78	0.39	0.02	0.29	0.08	0.19	0.06
whole_seq1_3	0.01	0.49	0.93	0.45	0.28	0.00	0.11	0.35	0.05	0.02
whole_seq1_4	0.22	0.82	0.42	0.60	0.37	0.01	0.57	0.03	0.09	0.03
whole_seq1_5	0.02	0.56	0.24	0.74	1.00	0.00	0.10	0.03	0.24	1.00

Maximum Vector Tracking Performance in individual Cb and Cr colour layer using Vector parameter alone

Tracking Performance in CbCr colour layer

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.38	0.00	0.08	0.50	0.00	0.11	0.00	0.01	0.08	0.00
whole_seq1_2	0.00	0.88	0.12	0.24	0.02	0.00	0.45	0.01	0.02	0.00
whole_seq1_3	0.00	0.15	0.82	0.46	0.03	0.00	0.01	0.12	0.05	0.00
whole_seq1_4	0.02	0.32	0.35	0.90	0.00	0.00	0.05	0.08	0.65	0.00
whole_seq1_5	0.04	0.61	0.96	0.54	0.20	0.00	0.11	0.83	0.10	0.04

Absolute Parameter Tracking Performance in CbCr colour layer

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.90	0.67	0.40	0.63	0.00	0.24	0.12	0.07	0.11	0.00
whole_seq1_2	0.10	1.00	0.44	0.49	0.07	0.01	1.00	0.08	0.09	0.00
whole_seq1_3	0.01	0.64	0.88	0.31	0.03	0.00	0.09	0.28	0.02	0.00
whole_seq1_4	0.02	0.83	0.10	0.90	0.00	0.00	0.41	0.01	0.65	0.00
whole_seq1_5	0.04	0.44	0.07	0.50	0.24	0.00	0.09	0.00	0.07	0.05

Tracking Performance in CbCr colour layer using Vector and Angle only

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.90	0.00	0.73	0.54	0.02	0.34	0.00	0.19	0.08	0.00
whole_seq1_2	0.00	0.93	0.22	0.29	0.05	0.00	0.52	0.01	0.03	0.00
whole_seq1_3	0.03	0.15	0.91	0.53	0.23	0.00	0.01	0.16	0.05	0.01
whole_seq1_4	0.13	0.38	0.85	0.90	0.50	0.01	0.05	0.15	0.65	0.19
whole_seq1_5	0.04	0.61	0.96	0.54	0.94	0.00	0.11	0.83	0.10	0.54

Absolute Parameter Tracking Performance in CbCr colour layer using Vector and Angle only

Reference Histogram	TDR Input Video					TC Input Video				
	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5	whole_ seq1_1	whole_ seq1_2	whole_ seq1_3	whole_ seq1_4	whole_ seq1_5
whole_seq1_1	0.90	0.67	0.40	0.65	0.02	0.24	0.12	0.07	0.11	0.00
whole_seq1_2	0.12	1.00	0.61	0.54	0.34	0.01	1.00	0.11	0.11	0.02
whole_seq1_3	0.01	0.64	0.91	0.39	0.32	0.00	0.09	0.29	0.03	0.03
whole_seq1_4	0.15	0.87	0.62	0.90	0.70	0.01	0.41	0.07	0.65	0.21
whole_seq1_5	0.04	0.44	0.44	0.50	0.93	0.00	0.09	0.07	0.07	0.48

Maximum Vector Tracking Performance in CbCr colour layer

Appendix 9

This appendix contains the tabulated data for maximum vector tracking performance across multiple cameras using CbCr layers in YCbCr colour space.

Camera 1

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.70	0.44	0.36	0.17	0.60	0.06	0.73	0.33	0.20
whole_seq2_2	0.87	0.97	0.87	0.38	0.69	0.11	0.90	0.84	0.34
whole_seq2_3	0.95	0.87	0.96	0.68	0.83	0.60	0.94	0.86	0.57
whole_seq2_4	1.00	1.00	1.00	0.92	0.91	0.29	0.96	0.96	0.53
whole_seq2_5	0.17	0.13	0.13	0.03	0.26	0.00	0.23	0.08	0.06
whole_seq2_6	0.78	0.92	0.77	0.56	0.62	0.55	0.83	0.84	0.23
whole_seq2_7	0.65	0.61	0.57	0.23	0.71	0.04	0.80	0.42	0.10
whole_seq2_8	0.89	0.97	0.92	0.44	0.73	0.21	0.92	0.97	0.48
whole_seq2_9	0.97	1.00	0.99	0.64	0.89	0.35	0.88	0.97	0.96

Camera 2

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.66	0.44	0.41	0.27	0.00	0.08	0.58	0.38	0.27
whole_seq2_2	0.70	0.84	0.84	0.33	0.00	0.21	0.86	0.57	0.31
whole_seq2_3	0.93	0.96	0.91	0.65	0.00	0.30	0.82	0.88	0.39
whole_seq2_4	0.93	1.00	0.86	0.59	0.00	0.33	1.00	0.90	0.50
whole_seq2_5	0.40	0.13	0.12	0.60	0.00	0.00	0.40	0.10	0.05
whole_seq2_6	0.75	0.98	0.89	0.49	0.00	0.30	0.85	0.81	0.43
whole_seq2_7	0.65	0.63	0.51	0.21	0.00	0.07	0.88	0.37	0.18
whole_seq2_8	0.93	0.86	0.95	0.63	0.00	0.36	0.93	0.93	0.51
whole_seq2_9	0.86	0.99	0.94	0.38	0.00	0.17	0.90	0.92	0.74

Camera 3

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.78	0.18	0.18	0.10	0.00	0.01	0.76	0.14	0.11
whole_seq2_2	0.91	0.89	0.86	0.39	0.00	0.29	0.98	0.82	0.46
whole_seq2_3	0.73	0.94	0.86	0.41	0.00	0.30	0.95	0.82	0.25
whole_seq2_4	0.93	0.97	0.95	0.73	0.00	0.14	0.98	0.92	0.55
whole_seq2_5	0.88	0.57	0.55	0.25	0.00	0.16	0.90	0.40	0.24
whole_seq2_6	0.65	0.92	0.75	0.21	0.00	0.00	0.86	0.77	0.34
whole_seq2_7	0.68	0.66	0.66	0.23	0.00	0.10	0.73	0.32	0.29
whole_seq2_8	0.80	0.92	0.87	0.38	0.00	0.20	0.92	0.80	0.34
whole_seq2_9	0.80	0.84	0.99	0.49	0.00	0.45	0.83	0.83	0.60

Maximum Vector Tracking Performance in CbCr with 6 colour segmentation and raw data

Camera 1

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.76	0.06	0.00	0.03	0.00	0.01	0.69	0.06	0.42
whole_seq2_2	0.02	0.90	0.46	0.33	0.00	0.29	0.28	0.92	0.66
whole_seq2_3	0.12	0.64	0.92	0.45	0.00	0.54	0.31	1.00	0.80
whole_seq2_4	0.25	0.92	0.65	0.93	0.00	0.02	0.42	1.00	0.99
whole_seq2_5	0.44	0.00	0.00	0.00	0.00	0.00	0.40	0.00	0.24
whole_seq2_6	0.03	0.82	0.62	0.09	0.00	0.91	0.27	0.87	0.42
whole_seq2_7	0.62	0.12	0.00	0.03	0.00	0.00	0.83	0.20	0.29
whole_seq2_8	0.03	0.61	0.23	0.26	0.00	0.14	0.33	0.97	0.62
whole_seq2_9	0.16	0.78	0.65	0.32	0.00	0.30	0.53	0.92	0.92

Camera 2

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.70	0.00	0.00	0.00	0.00	0.00	0.67	0.00	0.31
whole_seq2_2	0.06	0.60	0.53	0.37	0.00	0.19	0.59	0.91	0.56
whole_seq2_3	0.04	0.73	0.69	0.24	0.00	0.49	0.46	1.00	0.57
whole_seq2_4	0.17	0.93	0.27	0.58	0.00	0.13	0.36	0.96	0.74
whole_seq2_5	0.39	0.01	0.00	0.00	0.00	0.00	0.40	0.02	0.25
whole_seq2_6	0.01	0.78	0.24	0.11	0.00	0.55	0.46	0.90	0.55
whole_seq2_7	0.65	0.09	0.00	0.02	0.00	0.02	0.82	0.07	0.33
whole_seq2_8	0.12	0.85	0.07	0.41	0.00	0.14	0.56	1.00	0.88
whole_seq2_9	0.06	0.72	0.38	0.12	0.00	0.37	0.56	1.00	0.78

Camera 3

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.78	0.04	0.00	0.03	0.00	0.00	0.79	0.05	0.49
whole_seq2_2	0.02	0.78	0.50	0.40	0.00	0.26	0.36	0.95	0.73
whole_seq2_3	0.03	0.67	0.58	0.22	0.00	0.39	0.28	1.00	0.46
whole_seq2_4	0.13	0.90	0.48	0.57	0.00	0.05	0.15	1.00	0.83
whole_seq2_5	0.76	0.07	0.00	0.06	0.00	0.00	0.82	0.06	0.67
whole_seq2_6	0.00	0.79	0.35	0.08	0.00	0.26	0.40	0.64	0.39
whole_seq2_7	0.55	0.19	0.02	0.03	0.00	0.05	0.76	0.13	0.52
whole_seq2_8	0.20	0.66	0.10	0.31	0.00	0.21	0.34	0.75	0.44
whole_seq2_9	0.04	0.67	0.37	0.07	0.00	0.31	0.23	0.95	0.60

Maximum Vector Tracking Performance in CbCr with 5 colour segmentation and raw data

Camera 1

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.21	0.00	0.01	0.00	0.00	0.02	0.06	0.12	0.01
whole_seq2_2	0.55	0.44	0.40	0.10	0.00	0.10	0.16	0.66	0.23
whole_seq2_3	0.93	0.70	0.94	0.21	0.00	0.52	0.45	0.88	0.58
whole_seq2_4	0.58	0.46	0.08	0.84	0.00	0.65	0.05	0.96	0.27
whole_seq2_5	0.06	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.00
whole_seq2_6	0.68	0.14	0.81	0.06	0.00	0.51	0.12	0.83	0.26
whole_seq2_7	0.25	0.00	0.07	0.00	0.00	0.04	0.35	0.29	0.06
whole_seq2_8	0.55	0.26	0.59	0.15	0.00	0.12	0.15	0.70	0.47
whole_seq2_9	0.59	0.47	0.58	0.38	0.00	0.47	0.24	0.81	0.88

Camera 2

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.33	0.10	0.13	0.00	0.00	0.10	0.12	0.25	0.14
whole_seq2_2	0.63	0.41	0.86	0.03	0.00	0.01	0.29	0.31	0.26
whole_seq2_3	0.72	0.39	0.93	0.09	0.00	0.62	0.19	0.82	0.28
whole_seq2_4	0.63	0.23	0.01	0.16	0.00	0.51	0.04	0.98	0.20
whole_seq2_5	0.24	0.00	0.00	0.00	0.00	0.00	0.15	0.15	0.00
whole_seq2_6	0.56	0.05	0.81	0.03	0.00	0.26	0.04	0.30	0.34
whole_seq2_7	0.72	0.11	0.12	0.04	0.00	0.04	0.42	0.39	0.12
whole_seq2_8	0.73	0.25	0.47	0.05	0.00	0.34	0.37	0.59	0.54
whole_seq2_9	0.45	0.27	0.83	0.05	0.00	0.31	0.36	0.55	0.83

Camera 3

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.55	0.03	0.02	0.02	0.00	0.05	0.39	0.39	0.07
whole_seq2_2	0.58	0.38	0.46	0.22	0.00	0.40	0.14	0.61	0.28
whole_seq2_3	0.84	0.23	0.68	0.05	0.00	0.19	0.32	0.38	0.20
whole_seq2_4	0.74	0.28	0.09	0.23	0.00	0.01	0.21	0.76	0.27
whole_seq2_5	0.75	0.06	0.08	0.00	0.00	0.04	0.30	0.27	0.15
whole_seq2_6	0.49	0.08	0.69	0.01	0.00	0.18	0.09	0.19	0.21
whole_seq2_7	0.52	0.02	0.13	0.05	0.00	0.06	0.29	0.35	0.18
whole_seq2_8	0.52	0.10	0.52	0.00	0.00	0.08	0.31	0.46	0.23
whole_seq2_9	0.73	0.23	0.41	0.20	0.00	0.39	0.21	0.61	0.63

Maximum Vector Tracking Performance in CbCr with 4 colour segmentation and raw data

Camera 1

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.56	0.11	0.23	0.00	0.31	0.00	0.38	0.28	0.00
whole_seq2_2	0.25	0.51	0.25	0.09	0.02	0.00	0.06	0.08	0.07
whole_seq2_3	0.40	0.43	0.74	0.02	0.08	0.00	0.17	0.10	0.25
whole_seq2_4	0.22	0.13	0.07	0.20	0.00	0.02	0.00	0.23	0.02
whole_seq2_5	0.34	0.03	0.08	0.00	0.33	0.02	0.26	0.11	0.00
whole_seq2_6	0.06	0.03	0.00	0.00	0.06	0.00	0.05	0.03	0.00
whole_seq2_7	0.62	0.09	0.20	0.00	0.43	0.00	0.61	0.25	0.00
whole_seq2_8	0.11	0.05	0.02	0.00	0.05	0.00	0.50	0.82	0.00
whole_seq2_9	0.58	0.14	0.28	0.09	0.12	0.00	0.19	0.16	0.68

Camera 2

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.55	0.04	0.17	0.00	0.31	0.04	0.30	0.08	0.00
whole_seq2_2	0.23	0.24	0.40	0.06	0.03	0.00	0.01	0.10	0.04
whole_seq2_3	0.20	0.15	0.65	0.08	0.00	0.00	0.04	0.12	0.04
whole_seq2_4	0.19	0.09	0.04	0.17	0.00	0.00	0.00	0.18	0.00
whole_seq2_5	0.37	0.07	0.08	0.00	0.36	0.00	0.31	0.17	0.00
whole_seq2_6	0.05	0.03	0.04	0.00	0.03	0.00	0.04	0.06	0.00
whole_seq2_7	0.65	0.14	0.21	0.00	0.25	0.00	0.53	0.30	0.00
whole_seq2_8	0.24	0.00	0.19	0.00	0.02	0.00	0.19	0.03	0.00
whole_seq2_9	0.26	0.26	0.65	0.00	0.00	0.00	0.19	0.00	0.42

Camera 3

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.82	0.07	0.30	0.00	0.63	0.01	0.73	0.44	0.00
whole_seq2_2	0.34	0.23	0.17	0.07	0.01	0.00	0.01	0.06	0.05
whole_seq2_3	0.27	0.19	0.38	0.00	0.00	0.00	0.01	0.00	0.24
whole_seq2_4	0.28	0.30	0.25	0.10	0.02	0.00	0.00	0.23	0.03
whole_seq2_5	0.93	0.24	0.60	0.00	0.61	0.00	0.82	0.60	0.01
whole_seq2_6	0.08	0.04	0.16	0.03	0.00	0.00	0.00	0.00	0.00
whole_seq2_7	0.79	0.18	0.31	0.00	0.44	0.00	0.47	0.44	0.00
whole_seq2_8	0.18	0.05	0.05	0.03	0.00	0.00	0.00	0.00	0.00
whole_seq2_9	0.72	0.08	0.48	0.01	0.01	0.00	0.12	0.00	0.40

Maximum Vector Tracking Performance in CbCr with 3 colour segmentation and raw data

Camera 1

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.67	0.39	0.43	0.11	0.57	0.04	0.59	0.29	0.19
whole_seq2_2	0.92	0.89	0.53	0.37	0.80	0.00	0.31	0.68	0.28
whole_seq2_3	0.92	0.90	0.90	0.60	0.89	0.00	0.74	0.82	0.49
whole_seq2_4	0.96	0.95	0.63	0.87	0.97	0.05	0.33	0.97	0.44
whole_seq2_5	0.21	0.15	0.16	0.03	0.25	0.00	0.22	0.07	0.06
whole_seq2_6	0.84	0.82	0.69	0.57	0.64	0.08	0.09	0.75	0.22
whole_seq2_7	0.71	0.61	0.59	0.20	0.67	0.01	0.78	0.30	0.07
whole_seq2_8	0.85	0.88	0.56	0.48	0.76	0.00	0.41	0.88	0.42
whole_seq2_9	1.00	1.00	0.52	0.53	0.99	0.00	0.65	0.97	0.88

Camera 2

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.55	0.43	0.43	0.23	0.00	0.05	0.43	0.37	0.23
whole_seq2_2	0.84	0.81	0.61	0.37	0.00	0.10	0.63	0.51	0.30
whole_seq2_3	0.86	0.96	0.65	0.31	0.00	0.01	0.42	0.74	0.27
whole_seq2_4	1.00	0.87	0.60	0.66	0.00	0.31	0.31	0.89	0.46
whole_seq2_5	0.37	0.11	0.13	0.04	0.00	0.00	0.32	0.07	0.05
whole_seq2_6	0.86	0.90	0.71	0.45	0.00	0.28	0.14	0.74	0.41
whole_seq2_7	0.75	0.51	0.58	0.12	0.00	0.04	0.68	0.33	0.12
whole_seq2_8	0.92	0.93	0.63	0.53	0.00	0.02	0.69	0.78	0.41
whole_seq2_9	0.94	0.95	0.62	0.47	0.00	0.01	0.60	0.83	0.67

Camera 3

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.70	0.17	0.18	0.08	0.00	0.01	0.60	0.15	0.11
whole_seq2_2	0.97	0.90	0.54	0.44	0.00	0.01	0.43	0.60	0.27
whole_seq2_3	0.97	0.87	0.78	0.42	0.00	0.20	0.34	0.72	0.19
whole_seq2_4	0.98	0.97	0.63	0.68	0.00	0.02	0.49	0.89	0.40
whole_seq2_5	0.82	0.60	0.55	0.18	0.00	0.06	0.57	0.35	0.20
whole_seq2_6	0.86	0.70	0.48	0.23	0.00	0.08	0.13	0.66	0.32
whole_seq2_7	0.68	0.56	0.58	0.18	0.00	0.13	0.45	0.32	0.24
whole_seq2_8	0.90	0.84	0.72	0.39	0.00	0.11	0.38	0.72	0.30
whole_seq2_9	0.81	0.81	0.57	0.39	0.00	0.01	0.65	0.75	0.55

Maximum Vector Tracking Performance in CbCr with 6 colour segmentation and outlier data removed

Camera 1

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.76	0.06	0.00	0.00	0.00	0.00	0.69	0.69	0.06
whole_seq2_2	0.02	0.87	0.43	0.32	0.00	0.09	0.45	0.38	0.55
whole_seq2_3	0.12	0.70	0.89	0.39	0.00	0.06	0.77	0.74	0.77
whole_seq2_4	0.25	0.86	0.51	0.92	0.00	0.00	0.48	0.47	0.98
whole_seq2_5	0.44	0.00	0.00	0.00	0.00	0.00	0.35	0.35	0.00
whole_seq2_6	0.03	0.77	0.31	0.08	0.00	0.71	0.34	0.18	0.38
whole_seq2_7	0.62	0.09	0.00	0.01	0.00	0.00	0.78	0.78	0.09
whole_seq2_8	0.03	0.30	0.20	0.27	0.00	0.05	0.36	0.36	0.39
whole_seq2_9	0.16	0.73	0.28	0.31	0.00	0.18	0.64	0.55	0.89

Camera 2

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.70	0.00	0.00	0.00	0.00	0.00	0.60	0.60	0.00
whole_seq2_2	0.06	0.53	0.40	0.33	0.00	0.07	0.74	0.70	0.51
whole_seq2_3	0.04	0.88	0.61	0.23	0.00	0.15	0.73	0.64	0.54
whole_seq2_4	0.17	0.63	0.10	0.44	0.00	0.14	0.40	0.29	0.70
whole_seq2_5	0.39	0.00	0.00	0.00	0.00	0.00	0.37	0.37	0.00
whole_seq2_6	0.01	0.69	0.08	0.10	0.00	0.59	0.50	0.35	0.46
whole_seq2_7	0.65	0.12	0.00	0.18	0.00	0.02	0.79	0.72	0.05
whole_seq2_8	0.12	0.73	0.03	0.31	0.00	0.08	0.63	0.59	0.71
whole_seq2_9	0.06	0.64	0.36	0.10	0.00	0.26	0.87	0.77	0.78

Camera 3

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.82	0.02	0.00	0.02	0.00	0.00	0.77	0.77	0.04
whole_seq2_2	0.00	0.78	0.53	0.38	0.00	0.00	0.64	0.61	0.64
whole_seq2_3	0.05	0.63	0.57	0.23	0.00	0.14	0.56	0.52	0.38
whole_seq2_4	0.25	0.77	0.50	0.54	0.00	0.03	0.20	0.19	0.72
whole_seq2_5	0.81	0.06	0.00	0.06	0.00	0.00	0.79	0.79	0.07
whole_seq2_6	0.01	0.69	0.01	0.04	0.00	0.26	0.40	0.34	0.35
whole_seq2_7	0.61	0.13	0.02	0.03	0.00	0.02	0.74	0.73	0.08
whole_seq2_8	0.05	0.52	0.11	0.23	0.00	0.20	0.36	0.33	0.38
whole_seq2_9	0.81	0.81	0.57	0.39	0.00	0.01	0.65	0.75	0.55

Maximum Vector Tracking Performance in CbCr with 5 colour segmentation and outlier data removed

Camera 1

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.20	0.00	0.01	0.00	0.00	0.01	0.06	0.02	0.01
whole_seq2_2	0.83	0.52	0.36	0.10	0.00	0.10	0.16	0.28	0.18
whole_seq2_3	0.83	0.67	0.94	0.21	0.00	0.50	0.45	0.43	0.48
whole_seq2_4	0.59	0.45	0.03	0.82	0.00	0.24	0.05	0.91	0.33
whole_seq2_5	0.02	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00
whole_seq2_6	0.70	0.43	0.52	0.06	0.00	0.51	0.12	0.88	0.27
whole_seq2_7	0.17	0.00	0.03	0.00	0.00	0.04	0.33	0.03	0.03
whole_seq2_8	0.58	0.29	0.20	0.14	0.00	0.09	0.15	0.65	0.44
whole_seq2_9	0.58	0.47	0.35	0.35	0.00	0.38	0.24	0.70	0.84

Camera 2

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.19	0.11	0.01	0.00	0.00	0.03	0.11	0.11	0.13
whole_seq2_2	0.46	0.36	0.33	0.03	0.00	0.01	0.29	0.13	0.23
whole_seq2_3	0.64	0.32	0.62	0.09	0.00	0.38	0.19	0.23	0.16
whole_seq2_4	0.70	0.21	0.08	0.16	0.00	0.00	0.04	0.64	0.17
whole_seq2_5	0.17	0.00	0.03	0.00	0.00	0.00	0.14	0.01	0.00
whole_seq2_6	0.55	0.05	0.35	0.03	0.00	0.01	0.04	0.56	0.31
whole_seq2_7	0.40	0.07	0.02	0.04	0.00	0.04	0.39	0.04	0.07
whole_seq2_8	0.68	0.14	0.30	0.03	0.00	0.19	0.37	0.39	0.47
whole_seq2_9	0.36	0.26	0.38	0.05	0.00	0.22	0.36	0.27	0.77

Camera 3

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.44	0.04	0.02	0.02	0.00	0.04	0.39	0.05	0.06
whole_seq2_2	0.85	0.38	0.33	0.22	0.00	0.09	0.14	0.34	0.24
whole_seq2_3	0.73	0.22	0.62	0.04	0.00	0.19	0.32	0.23	0.22
whole_seq2_4	0.79	0.28	0.08	0.01	0.00	0.00	0.21	0.15	0.17
whole_seq2_5	0.46	0.02	0.03	0.00	0.00	0.02	0.29	0.06	0.12
whole_seq2_6	0.60	0.03	0.35	0.01	0.00	0.16	0.09	0.19	0.21
whole_seq2_7	0.32	0.02	0.02	0.05	0.00	0.00	0.26	0.13	0.13
whole_seq2_8	0.62	0.10	0.30	0.00	0.00	0.05	0.31	0.03	0.31
whole_seq2_9	0.60	0.20	0.37	0.20	0.00	0.12	0.23	0.44	0.57

Maximum Vector Tracking Performance in CbCr with 4 colour segmentation and outlier data removed

Camera 1

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.54	0.00	0.00	0.00	0.27	0.00	0.43	0.00	0.00
whole_seq2_2	0.02	0.31	0.23	0.08	0.01	0.00	0.02	0.00	0.07
whole_seq2_3	0.10	0.33	0.64	0.05	0.07	0.00	0.10	0.00	0.25
whole_seq2_4	0.00	0.74	0.07	0.19	0.00	0.00	0.00	0.02	0.02
whole_seq2_5	0.33	0.00	0.00	0.00	0.31	0.00	0.32	0.00	0.00
whole_seq2_6	0.06	0.00	0.00	0.00	0.03	0.00	0.06	0.00	0.00
whole_seq2_7	0.55	0.00	0.00	0.00	0.30	0.00	0.61	0.00	0.00
whole_seq2_8	0.06	0.00	0.00	0.00	0.03	0.00	0.06	0.00	0.00
whole_seq2_9	0.12	0.22	0.19	0.03	0.08	0.00	0.12	0.09	0.68

Camera 2

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.43	0.00	0.00	0.00	0.18	0.00	0.31	0.00	0.00
whole_seq2_2	0.03	0.27	0.34	0.06	0.03	0.00	0.00	0.00	0.04
whole_seq2_3	0.04	0.15	0.61	0.11	0.00	0.00	0.04	0.00	0.04
whole_seq2_4	0.00	0.12	0.04	0.14	0.00	0.00	0.00	0.00	0.00
whole_seq2_5	0.37	0.00	0.00	0.00	0.27	0.00	0.35	0.00	0.00
whole_seq2_6	0.03	0.03	0.03	0.00	0.03	0.00	0.03	0.00	0.00
whole_seq2_7	0.51	0.00	0.00	0.00	0.25	0.00	0.49	0.00	0.00
whole_seq2_8	0.02	0.00	0.00	0.00	0.02	0.00	0.02	0.00	0.00
whole_seq2_9	0.00	0.26	0.36	0.00	0.00	0.00	0.00	0.00	0.42

Camera 3

Input Video	Reference								
	whole_seq2_1	whole_seq2_2	whole_seq2_3	whole_seq2_4	whole_seq2_5	whole_seq2_6	whole_seq2_7	whole_seq2_8	whole_seq2_9
whole_seq2_1	0.79	0.00	0.00	0.00	0.57	0.00	0.76	0.00	0.00
whole_seq2_2	0.01	0.27	0.12	0.07	0.01	0.00	0.01	0.00	0.05
whole_seq2_3	0.00	0.33	0.37	0.00	0.00	0.00	0.00	0.00	0.24
whole_seq2_4	0.02	0.46	0.24	0.15	0.02	0.00	0.02	0.00	0.03
whole_seq2_5	0.91	0.00	0.00	0.00	0.46	0.00	0.85	0.00	0.00
whole_seq2_6	0.00	0.04	0.04	0.01	0.00	0.00	0.00	0.00	0.00
whole_seq2_7	0.63	0.00	0.00	0.00	0.31	0.00	0.52	0.00	0.00
whole_seq2_8	0.00	0.54	0.05	0.03	0.00	0.00	0.00	0.00	0.00
whole_seq2_9	0.07	0.08	0.19	0.01	0.00	0.00	0.05	0.00	0.40

Maximum Vector Tracking Performance in CbCr with 3 colour segmentation and outlier data removed

Appendix 10

This appendix contains the tabulated data for maximum vector tracking performance across multiple cameras using CbCr layers in YCbCr colour space under different scenarios.

Scenario 1

Cam1
TDR

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.92	0.43	0.92	0.41	0.86	0.00	0.71	0.18
whole_seq3_2	0.47	0.82	0.51	0.81	0.54	0.00	0.51	0.93

TC

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.37	0.03	0.20	0.03	0.15	0.00	0.08	0.01
whole_seq3_2	0.01	0.19	0.08	0.16	0.07	0.00	0.04	0.45

Cam2
TDR

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.97	0.17	0.90	0.13	0.65	0.00	0.38	0.21
whole_seq3_2	0.84	0.23	0.83	0.22	0.53	0.00	0.45	0.24

TC

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.43	0.01	0.53	0.01	0.09	0.00	0.01	0.01
whole_seq3_2	0.47	0.01	0.25	0.02	0.03	0.00	0.02	0.02

Cam3
TDR

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.48	0.23	0.68	0.21	0.31	0.00	0.39	0.32
whole_seq3_2	0.68	0.34	0.62	0.28	0.60	0.00	0.52	0.53

TC

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.06	0.02	0.07	0.01	0.01	0.00	0.02	0.02
whole_seq3_2	0.05	0.01	0.04	0.01	0.03	0.00	0.02	0.04

Scenario 2

Cam1
TDR

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.26	0.04	0.33	0.05	0.30	0.04	0.19	0.16
whole_seq3_2	0.81	0.03	0.04	0.06	0.57	0.00	0.46	0.14

TC

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.01	0.00	0.03	0.00	0.01	0.00	0.01	0.01
whole_seq3_2	0.11	0.00	0.01	0.03	0.05	0.00	0.04	0.01

Cam2
TDR

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.78	0.10	0.67	0.05	0.73	0.00	0.32	0.16
whole_seq3_2	0.82	0.20	0.12	0.23	0.57	0.00	0.28	0.14

TC

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
whole_seq3_2	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Cam3
TDR

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq3_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TC

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
whole_seq3_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Scenario 3

Cam1

TDR

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.46	0.25	0.39	0.05	0.28	0.00	0.12	0.00
whole_seq3_2	0.24	0.10	0.23	0.12	0.24	0.00	0.11	0.00

TC

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.00
whole_seq3_2	0.01	0.00	0.01	0.00	0.01	0.00	0.00	0.00

Cam2

TDR

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.69	0.13	0.41	0.10	0.37	0.00	0.03	0.13
whole_seq3_2	0.36	0.29	0.30	0.14	0.27	0.00	0.10	0.23

TC

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.02	0.00	0.01	0.00	0.01	0.00	0.00	0.00
whole_seq3_2	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.01

Cam3

TDR

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.55	0.00	0.44	0.00	0.67	0.00	0.71	0.00
whole_seq3_2	0.00	0.14	0.00	0.25	0.00	0.38	0.00	0.53

TC

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.05	0.00	0.03	0.00	0.05	0.00	0.06	0.00
whole_seq3_2	0.00	0.00	0.00	0.01	0.00	0.03	0.00	0.06

Scenario 4

Cam1

TDR

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.01	0.01	0.28	0.17	0.30	0.00	0.00	0.05
whole_seq3_2	0.84	0.13	0.23	0.40	0.36	0.00	0.04	0.22

TC

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.02	0.02	0.01	0.00	0.01	0.00	0.00	0.00
whole_seq3_2	0.02	0.02	0.01	0.02	0.01	0.00	0.00	0.00

Cam2

TDR

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.61	0.00	0.39	0.06	0.61	0.00	0.00	0.00
whole_seq3_2	0.00	0.18	0.00	0.13	0.00	0.00	0.01	0.00

TC

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.01	0.00	0.01	0.00	0.01	0.00	0.00	0.00
whole_seq3_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Cam3

TDR

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.55	0.33	0.36	0.11	0.56	0.33	0.43	0.44
whole_seq3_2	0.38	0.24	0.24	0.16	0.49	0.17	0.52	0.27

TC

Input Video Sequence	3 segment		4 segment		5 segment		6 segment	
	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2	whole_seq3_1	whole_seq3_2
whole_seq3_1	0.04	0.02	0.02	0.00	0.04	0.02	0.02	0.02
whole_seq3_2	0.01	0.01	0.01	0.00	0.02	0.01	0.05	0.01

Appendix 11

This appendix contains the first pages of conference and journal papers that were generated from work done on this research

Performance Evaluation Metrics for Video Tracking

Patrick Sebastian, Yap Vooi Voon[1] and Richard Comley[2]

Departments of Electrical and Electronic Eng., Universiti Teknologi PETRONAS, Tronoh, [1]Department of Electrical Eng., Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia, [2]School of Engineering and Information Science, Middlesex University, The Burroughs, Hendon, London NW4-4B, UK

Abstract

Various tracking methods have been developed to track objects with different degrees or levels of tracking ability. The ability or performance of each tracking method is dependent on the feature or data that is being used for tracking purpose. The ability of a tracking method can be measured by utilizing tracking metrics to give an indication of the tracking ability of an algorithm. This paper offers some insights into the issues and similarities of performance measurement reporting of video tracking algorithms and proposes a method in assessing the robustness of a video tracking algorithm. The proposed metric introduces another measure to measure the consistency of a tracking algorithm. The work presented in this paper shows that using only one metric to measure the tracking performance is inadequate. The proposed metric presented in this paper shows that the utilization of multiple metrics such as tracking success rate and tracking consistency or robustness would give a better indication of the tracking ability of a tracking algorithm used in video surveillance.

Keywords

Color space, Tracking metrics, Video surveillance, Video tracking.

1. Introduction

Current implementations of tracking and surveillance systems have provided valuable information and assistance in monitoring large areas. These systems typically utilize human operators to determine suspicious human behavior and to manually track people of interest over an extended area using an array of cameras. The limitation or difficulty for a human operator to constantly monitor cameras or to track people over an array of cameras had brought about the merging of technologies of cameras and computers. This in turn has lead to the growth of intelligence in the surveillance systems being developed. Along with these intelligent surveillance systems, tracking algorithms have been developed. Each tracking algorithm that has been developed is based on improving the performance of previous algorithms by utilizing different parameters or features [1]. Metrics are measures of the ability of an algorithm in processing signals ranging from still images, audio signals [2] to video signals. In the area of video surveillance, which covers the action of tracking, a target tracking algorithm would have to be developed. To determine the performance of a tracking algorithm, a performance metric needs to be able to determine its correctness or accuracy. In addition to determining correctness or accuracy, other metrics can also be used to characterize the performance such as tracking consistency (TC).

There have been various evaluation metrics that were developed to evaluate the performance of a tracking algorithm [3-8]. Most of the published sets of agreed metrics utilized ground truth, which will be explored in the next section. There are alternatives to using ground truth to evaluate the performance of a tracking algorithm and a number of these have been identified by Ellis [9]. The alternatives include a method that requires each target to carry a mobile global positioning receiver to record 4D trajectory information that could be correlated directly with results of the video dataset. Another method cited takes advantage of the capability of automatically identifying targets in overlapping multi-camera views [10-14] and a third method uses synthetic image sequences to assess the performance of the algorithm [15].

In general, the objective of each tracking method was to ensure the ability of tracking or maintaining an accurate track of selected objects or targets of interest. The performance of each algorithm is based on the features used for tracking the selected targets. Tracking metrics are measures that indicate the performance or the ability of the tracking algorithms to successfully track the selected targets. In looking at the metric developed by different groups, there are similarities in terms of the measurement of particular performance metrics. Also, not all metrics developed are able to completely determine the performance of a tracking algorithm and as such other supporting metrics that measure certain other characteristics are required, in

[redacted]

Sebastian, P., Y. Vooi Voon, and R. Comley, *Performance Evaluation Metrics for Video Tracking.* IETE Technical Review, 2011. **28**(6): p. 493-502.

[redacted]

Colour Space Effect on Tracking in Video Surveillance

Patrick Sebastian[1], Yap Vooi Voon[2], and Richard Comley[3]

[1]Electrical and Electronic Engineering Dept. Universiti Teknologi PETRONAS
Bandar Seri Iskandar 31750 Tronoh, Perak, Malaysia
[2]Electronic Engineering Dept. Universiti Tunku Abdul Rahman
Jalan Universiti Bandar Barat, 31900 Kampar, Perak, Malaysia
[3]Middlesex University The Burroughs,
Hendon, London NW4 4B, United Kingdom

Abstract: The utilization of video surveillance systems is becoming common and is expected to more widespread as societies become more complex and the population continues to grow. The implementation of these video surveillance systems has provided valuable information and assistance in monitoring large areas. These systems typically use human operators to determine human behavior and to manually track people or objects of interest over an array of cameras. With the application of computers with video surveillance, real time surveillance of large public areas, people and their activities has been made possible for monitoring and security. The main contribution of this paper is to show the effect of color space on tracking methods in video surveillance. Results from evaluations on different tracking methods have indicated that YCbCr and HSV color spaces have better tracking ability compared to grayscale and RGB color spaces. In addition, the results from evaluations have also indicated that data from selected layers in some color spaces can be used for the purpose of tracking namely the Cb and Cr layers from the YCbCr color space and the H layer from the HSV color space.

Keywords: Video, Tracking, Video Surveillance, Tracking Metrics, Color space

1. Introduction

The utilization of video surveillance systems is becoming common and is expected to more widespread as societies become more complex and the population continues to grow. In utilizing video surveillance in a large scale environment, it leads to problems in recognizing human behavior and tracking people over an array of cameras. The task of recognizing human behavior is difficult by itself and in adding the task of tracking a particular person or persons over an array of cameras makes the task even more difficult for a human operator.

The motivation of this paper is to show how different color spaces effects the tracking performance of a video surveillance system. The paper is divided into 6 sections. Section 2 covers the fundamental background information related to colour spaces of images. Section 3 covers recent and related research work in tracking methodologies. Section 4 covers the experimental setup for evaluating the effect on tracking performance on different colour spaces. Section 5 covers the results of the experiments done in determining tracking performance in different colour spaces.

2. Background

Current implementations of tracking and surveillance systems have provided valuable information and assistance in monitoring of large areas. These systems typically utilize human operators to determine human behavior and to manually track people of interest over an array of cameras in the surveillance system. The limitation or difficulty for a human operator to constantly monitor cameras or to track people over an array of cameras leads to the merging of technologies of cameras and computers. The merging of cameras and computers in monitoring

[redacted]

Sebastian, P., Y. Vooi Voon, and R. Comley, *Colour Space Effect on Tracking in Video Surveillance.* International Journal on Electrical Engineering and Informatics, 2010. **2**(4): p. 298-312.

[redacted]

Tracking Consistency Metric for Video Surveillance Tracking

Patrick Sebastian
Electrical and Electronics Engineering Department
Universiti Teknologi PETRONAS
31750 Tronoh, Perak, Malaysia
Patrick_sebastian@petronas.com.my ,

Yap Vooi Voon
Electrical and Electronics Engineering Department
Universiti Teknologi PETRONAS
31750 Tronoh, Perak, Malaysia
vooiv@petronas.com.my

Abstract – The aim of this paper is to propose an additional performance tracking metric. The performance of a tracking algorithm depends on the parameters that are being tracked ranging from reference points that cover edges and corners, to blob parameters such as size, trajectory and color. Based on different features used in tracking, the tracking metrics range from contingency tables, rates of correct tracks to error differences between tracked features and reference points. Among the tracking metrics that have been developed were tracker detection rate (TDR), object tracking error (OTE) and contingency table. These metrics give a measure of the ability of a tracking algorithm to correctly detect or track a target object. There has been no metric published or utilized to determine the performance of a tracking algorithm in the consistency of maintaining a correct track on the selected target. This paper proposes a method and metric to give a measure of determining the consistency of a tracking. The results of the proposed TC metric have results that range from 0.81 to 0.35 for TDR of 90%, 0.65 to 0.21 for TDR of 80%, 0.5 to 0.16 for TDR of 70%, 0.37 to 0.12 for TDR of 60% and 0.26 to 0.08 for TDR of 50%.

***Keywords**: metric, tracking, tracking metrics, tracking performance*

1. INTRODUCTION

Tracking action in video surveillance is used to track specific targets in a live video stream or a recorded video. A number of video tracking methods are based on algorithms developed to track selected targets. These tracking algorithms were developed to take advantage of the differences in color spaces and data available for use in tracking. The data used for tracking range from color regions and motion metrics such as trajectory and speed of the tracked target. Each tracking algorithm was developed to improve on the performance of the previous tracking algorithm. In order to determine the performance of different tracking algorithms, the tracking algorithms have been measured and compared based on tracking metrics. These metrics are used for the determination of tracking algorithm performance based on the robustness and correctness of the tracking algorithm, comparison between alternative algorithms and improvements of the incremental improvements to the tracking algorithms [1]. Performance metrics are also used as a measure for the reliability or robustness of the tracking algorithms that are being measured [2].

2. BACKGROUND

In this paper, the performance metrics used to determine the performance of tracking algorithms are based on different parameters or features. The features used for tracking range from specific feature or reference items such as edges, lines, corners [3], limbs [4] and heads [5] to parameters of blobs being tracked such as size [6], trajectory [7], velocity [8] and color [9]. The different features and parameters used for tracking would yield different type of results due to the type of information used to determine the performance of the tracking algorithm. The tracking metrics range from contingency tables to list true and false positives [1], error metrics [10], surveillance metrics such as Tracker Detection Rate (TDR) [2] and trajectory comparison [11].

The tracking metrics cited earlier are determined with reference to a reference point that is called the ground truth [1]. Ground truth points are independent and objective data points for reference to the tracked points in the video. Ground truths are determined by the criteria of the target object as being easily identifiable either by general identification point or a specific reference tracking point and classification information can be generated [1].

A. Metric Types

Based on different tracking methods, the performance metrics available can be classified into different categories based on the type of information used such as contingency table of observations [1], tracker detection rates and object tracking error [2], rate of misclassification [10], trajectory comparisons [11], color contrast, histogram and motion difference [12]. Each type of tracking metric give different results based on the resultant data. Each of these metrics

[redacted]

Sebastian, P. and Y. Vooi Voon. *Tracking Consistency Metric for Video Surveillance Tracking*. in *International Conference on Signal Processing Systems*. 2009. Singapore.

[redacted]

Tracking Using Normalized Cross Correlation and Color Space

Patrick Sebastian, Yap Vooi Voon
Electrical and Electronics Engineering Department
Universiti Teknologi PETRONAS
Patrick_sebastian@petronas.com.my, vooiv@petronas.com.my

Abstract

The aim of this paper is to describe the implementation a face tracking algorithm for video conferencing environment using the normalized cross correlation method. The reference image was selected based on human skin properties in the YCbCr color space. The results obtained on different color spaces showed that the YCbCr color space had higher tracker detection rates (TDR) compared to the RGB and grayscale color spaces. In addition to the TDR being used as a metric to determine the accuracy of a tracking method, another metric was formulated to quantify the accuracy of a tracking algorithm. The metric developed was object tracking variance (OTV) and object tracking standard deviation (OTStd). The OTV and OTStd are used to indicate the accuracy of tracking a reference point. Based on the results obtained, it indicates that the YCbCr color space has lower standard deviation compared to the RGB and grayscale color spaces. The results also indicate that the CbCr components of the YCbCr can be used for correct tracking without the Y component.

Keywords: face tracking, skin properties, normalized correlation, phase correlation, color spaces

1. Introduction

Object tracking in video surveillance is the action of detecting a pre-determined reference object within another image called a target image. In this work the target image is obtained from a webcam. The face tracking algorithm involves matching a reference image with the target image.

2. Approach and Methods

In this paper, the object tracking algorithm utilizes template matching approach to perform the tracking task. Prior to the object tracking process, the image from the webcam is captured and stored in order to process the image for tracking purposes.

Correlation was one of the methods utilized in the development of the object tracking algorithm. This method compares an unknown signal to a known signal and obtains a measure of similarity between the two signals.

2.1. Cross Correlation in Spatial Domain

Cross correlation (CC) is the correlation of two different signals and it is a standard approach to feature detection [1]. One application of cross correlation is in template matching, where the task is to find the best match of two different images [2]. In template matching a given pattern in one image is compared with a template containing the same pattern of another image. This paper proposes to use normalized cross-correlation (NCC) and color space to track a person's face in a video conferencing environment. The NCC methods have significant advantage over standard CC methods, in that the NCC methods are robust to different lighting conditions across an image and less sensitive to noise [3].

The correlation between two images can be described as

$$c(u,v) = f(x,y)t(x-u, y-v) \quad (1)$$

where $f(x,y)$ is the feature and the $t(x,y)$ is the sub-image or the feature within the image. Eq. (1) is a measure of the similarity between the image and the feature. The range of $c(u,v)$ is dependent on the size of the feature.

In this paper, a template is shifted into different positions, where at each position, intensities are multiplied and summed, producing a normalized cross correlation matrix, $\gamma(u,v)$. The NCC matrix $\gamma(u,v)$ can be described as

[redacted]

Sebastian, P. and Y. Vooi Voon. *Tracking Using Normalized Cross Correlation and Color Space*. in *Internation Conference on Intelligent and Advanced Systems (ICIAS 2007)*. 2007. Kuala Lumpur, Malaysia.

The Effect of Colour Space on Tracking Robustness

Patrick Sebastian[1], Yap Vooi Voon[2], Richard Comley[3]
[1][2]Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh, Perak, Malaysia

[3]Middlesex University
The Burroughs,
Hendon, London NW4 4BT

***Abstract*-This paper studies the effect of colour space on the performance of tracking algorithms. The colour spaces that were investigated were grayscale, RGB, YCbCr and HSV. The performance of a normalised cross correlation tracking algorithm was measured to determine robustness and accuracy in the different colour spaces. Track Detection Rate (TDR) and Object Tracking Standard Deviation (OTStd) were used to provide quantitative measures of tracking performance. The combined results indicate that the colour spaces of YCbCr and HSV give more accurate and more robust tracking results compared to grayscale and RGB images. The results also show that the information stored in the chrominance layers of CbCr in the YCbCr colour space and chromaticity layers HS in the HSV colour space, were sufficient for robust tracking. The TDR results range from 93.7% to 97.1% for grayscale and RGB, and 98% to 100% for the YCbCr and HSV colour spaces respectively. A similar trend in the OTStd was observed with a range of 17.0 pixels to 23.9 pixels for grayscale and RGB, and 7.56 pixels to 20.5 pixels for YCbCr and HSV.**

I. Introduction

Object tracking in video surveillance is the action of detecting a reference object in an input image. The act of tracking in video surveillance is becoming an important task especially in monitoring large scale environments such as public and security sensitive areas. The implementation of video surveillance systems has provided valuable information and assistance in monitoring such spaces. These systems typically use human operators to determine human behaviour and to manually track people or objects of interest over an array of cameras. With the application of computers to video surveillance, real time surveillance of large public areas, people and their activities, has been made possible for monitoring and security [1].

In the field of video surveillance, an object of interest would be identified and then monitored or tracked. People are typically the object of interest in video surveillance applications, for example when walking through a secluded or security sensitive area. There is now increasing interest in monitoring people in public areas, e.g. shopping malls. When tracking objects of interest in a wide or public area, additional parameters are required to improve performance such as colour of clothing [2], path and velocity of tracked object [3, 4] and modelling set colours for tracked persons [5]. To obtain robust tracking of a target, a number of tracking methods are typically employed in order to overcome problems such as occlusion [1, 5, 6] and noise in surveillance videos [7]. Various factors affect the tracking ability and robustness of any tracking algorithm when used in open, public spaces such as variations in illumination of an area being monitored, image type used and tracking methodology.

II. Background

A. Colour Space

Different image formats or types range from grayscale images to colour images. Grayscale images can be classified as intensity type images where the data that is used to represent the image is a measurement of the intensity or the amount of light [8]. The number of bits used for each pixel determines the number of brightness levels that the pixel will have. Apart from grayscale images, another type of image that can be used is a colour image. A colour image can be classified as a set of multiple layers of grayscale images where each layer of the image corresponds to a certain band in the visible light spectrum [9]. The information that is stored in each layer of the colour image is the brightness in a specific spectral band. The most commonly used spectral bands are red (R), green (G) and blue (B), the three primary colours in the visible range of the electromagnetic spectrum [10]. The RGB colour bands are chosen as they correspond to the absorption characteristics of the human eye [10].

While RGB may be the best representation for many applications (e.g. television) it suffers from a number of serious limitations when it comes to activities such as computer based surveillance systems. The main limitation of the RGB colour space is due to the fact that the luminance information that is embedded into each layer of the image [11]. Varying levels of brightness in an image causes RGB values to shift [12] and that introduces instability in the image [13]. The susceptibility of the RGB colour space to brightness levels indicates that each layer is equally affected and that the layers are correlated to each other [13, 14]. To overcome this problem, the RGB colour space can be normalized to obtain the

[redacted]

Sebastian, P., Y. Vooi Voon, and R. Comley. *The Effect of Color Space on Tracking Robustness*. in *3rd IEEE Conference on Industrial Electronics and Applications*. 2008. Singapore.

[redacted]

Parametric Tracking Of Multiple Segmented Regions

Patrick Sebastian
Electrical and Electronic Engineering Department
Faculty of Engineering
Universiti Teknologi PETRONAS
Perak, Malaysia
Patrick_sebastian@petronas.com.my

Yap Vooi Voon
Department of Electronic Engineering
Faculty of Engineering and Green Technology
Universiti Tunku Abdul Rahman
Perak, Malaysia
vvyap@utar.edu.my

Richard Comley
Department of Engineering and Information Science
Middlesex Universiti
r.comley@mdx.ac.uk

***Abstract*—This paper proposes a tracking method based on parameters obtained from multiple blobs. The multiple blobs are derived or obtained from segmenting a single blob into multiple blobs or multiple regions that have the same color information where these regions generally remain the same as the target moves. The target being tracked is a person or human walking through the camera view field where the number of regions are dependent on the clothing worn and the general build of the person being tracked. This would indicate that the head, limbs and torso of a person would be segmented into regions of interest. The parameters used in tracking a multiple region or blob target are the vectors between the regions of interest and the mean values of the regions of interest. In this paper, the vector used is derived from the top most region to the lowest region of the multi region target. In addition, the difference between the mean values of the regions is used as a means of tracking in addition to the vector between the regions of interest. The results obtained showed that correct target tracking had consistently higher tracking rate compared to incorrect tracking. Correct tracking rates had higher than 0.9 Track Detection Rate (TDR) rates as compared to incorrect tracking that ranged from 0 to 0.6.**

Keywords-blob, tracking, parametric tracking

I. INTRODUCTION

Video surveillance is the usage of video cameras in observing and monitoring either a small or large area. Typical applications of video surveillance would range from crime prevention, traffic monitoring to industrial monitoring. Video surveillance is typically used in observing or monitoring areas of interest, active monitoring or tracking of a target is an added feature in video surveillance systems. Tracking of targets would give the ability to specifically track selected targets. The tracking of targets would lead to the development of tracking algorithms. Tracking algorithms are based on specific information extracted from targets such as area[1], color of clothing [2], path and velocity of the tracked targets [3-4].

II. BACKGROUND

Tracking of targets are based on parameters or features which can range from specific features such as edges, line, corners [5], limbs [6] and heads [7] to parameters of blobs that are being tracked such as size [8], trajectory [4], velocity [9] and color [10]. Based on some of the features and parameters used for tracking targets, tracking of targets can range from simple target tracking methods to more complex target tracking methods [5].

A. Color Space

In tracking targets, color space has an effect of tracking specific targets. Typical color spaces used in video tracking are RGB, HSV and YCbCr. From the different color spaces used, RGB color space has not been the preferred color space for tracking targets. This is due to the brightness information that is embedded in the respective color layers of the RGB color space [11]. In addition to having the brightness information embedded into the color layers, each color layer is correlated to each other which would lead to wrong tracking [12]. From different studies, it was found that color spaces with chrominance information on a separate layer from brightness information such as HSV and YCbCr have better tracking capabilities [13-15]. The utilization of chrominance information such as HS and CbCr layers could provide tracking capability without the use of any brightness information [14, 16-17].

B. Blob Modelling

Blob detection is the first step in video surveillance where the foreground and backgrounds are distinguished from each other [18]. Based on the blob that is detected, various portions of the tracked target can be tracked based on different parameter such as blob size, motion, vector and centroid [5, 19]. The segmented blobs are targeted based on the entire blob such as a whole human blob [20] or portions of the blob such as the head [21] for purposes of tracking.

The target of interest in this case would be the tracking of whole humans which would lead to a utilization of tracking

[redacted]

Sebastian, P., Y. Vooi Voon, and R. Comley, *Parametric Tracking of Multiple Segmented Regions*, in *2012 IEEE Symposium on Computer Applications and Industrial Electronics - Signal & Image Processing*. 2012: Kota Kinabalu, Sabah.