5-2005

Compensation for automatic white balance correction with histogram equalization

David D. Kirtley
University of Texas-Pan American

COMPENSATION FOR AUTOMATIC WHITE BALANCE CORRECTION WITH HISTOGRAM EQUALIZATION

A Thesis

by

DAVID D. KIRTLEY

Submitted to the Graduate School of the
University of Texas – Pan American
In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2005

Major Subject: Computer Science

COMPENSATION FOR AUTOMATIC WHITE BALANCE CORRECTION WITH HISTOGRAM EQUALIZATION

A Thesis
by
DAVID D. KIRTLEY

Approved as to style and content by:

Ping-Sing Tsai, Ph. D.
Chair of Committee

Zhixiang Chen, Ph. D.
Committee Member

Peter A. Ng, Ph. D.
Committee Member

May 2005

ABSTRACT

Kirtley, David D. Compensation for Automatic White Balance Correction with Histogram Equalization, Master of Science (MS), May, 2005, 46 pp., 25 Figures, references, 23 titles.

Histogram equalization rather than hard scaling can be used as an effective technique to counter automatic white balance correction in video processing to facilitate motion detection in video sequences. Benefits of this method are less user interaction needed by not needing to preview the image to select a scaling area and reduction of the non-focused changes in the video caused by using a scaling area. Reduced interaction lends itself to data mining of video.

ACKNOWLEDGMENTS

The author would like to thank Dr. Andrew Griffis and Dr. Tinku Acharya of Avisere Inc., Tucson, AZ, for their gracious contribution of the video sequences used in this experiment and their useful discussion.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Motion detection and image recognition from computer video is based on changes in color and intensity of picture elements (pixels) of individual frames of a sequence of images taken over time. It would be much simpler to use alternative sensors that detect and measure information such as light intensity, motion, and distance that a computer program would be able to interpret but the wealth of raw data in pre-existing video and the inexpensive nature of consumer grade video capture equipment makes their use attractive.

Video capture equipment is designed to create a representation of a scene that can be perceived in a similar fashion as a natural scene perceived by a person with normal human vision under normal viewing conditions. Because the nature of vision is perceptual, the captured image does not necessarily match what would be perceived. A video taken of a static outdoor scene changes over time as the conditions change. The angle of the sun changes during the day. Clouds pass overhead. Light reflected from objects outside the scene illuminates the scene differently. When you introduce moving objects into the scene, they add to the complexity as they also create their own shadows, occlude portions of the scene, and reflect light onto other parts of the scene, which in turn creates another generation of shadows and reflections. The human vision system is adapted to compensate for these changes and deals with them by automatically compensating and adjusting the perception of the scene. Image capture equipment tries to adapt to these changes as well for the resulting image to match what a person would see [1, 2, 3]. When utilizing captured video for information extraction, it is necessary to determine whether the changes in the video are a result of something passing through the images over time or are a result of the video capture equipment changing the underlying scene. The problem that the research presented here addresses is the effects of automatic white balance adjustment. An example can be seen in Figure 1 where two individual frames of a sequence are shown together to illustrate the changes in color brought about by this adjustment.

Figure 1 Illustration of effects of automatic white balance adjustment.

The image on the left is the first image in the sequence. On the right is the same scene but with a vehicle whose passage has caused a readjustment of the colors of the image. Figure 2 shows the amount of change created by the automatic white balance. The graph is a representation of the video over time. The x-axis is the frame number which represents the time counted in frames. The y-axis is the Root Mean Squared (RMS) of the change in intensity of the pixels contained in the area of interest. A RMS of zero indicates no change from the original frame and as the value increases, it represents a more discernable change from the original frame with nothing present in the area of interest. Before the passing of the first vehicle, the level of the graph of the line representing the average change of the color level was near zero within area of interest. As the car came into the frame, the object which had been selected as the white point of the scene was the wall on the left of the frame. The white car then became the new white point as it entered the frame and the other colors were adjusted accordingly. In the portion of the graph between the cars, the average level was near 50 when it should have returned to the near zero level present before the second car passed through the frame. This creates a problem in that you cannot reliably set a threshold defining the division between two cars based on this color intensity level as the intensity level is composed of two components. The first component is the actual change caused by the object and the other component is caused by the automatic white balance of the camera. Further complicating the issue is the fact that the automatic white balance may or may not be a result of in proximity to the area of interest. It could be brought about by any object within the scene. The simplest solution would be to turn off the automatic white balance feature of the camera but this is not an option with the lower cost video equipment. The ability to turn off that feature is only found in the higher end of video equipment product lines. The commonly available consumer grade video capture equipment has at most the ability to switch modes to compensate for the differences between incandescent, fluorescent, and natural lighting. Making that adjustment would only change one type of automatic white balance adjustment for another. It would also require a greater amount of human interaction during capture operations and is of no help when dealing with pre-existing video.

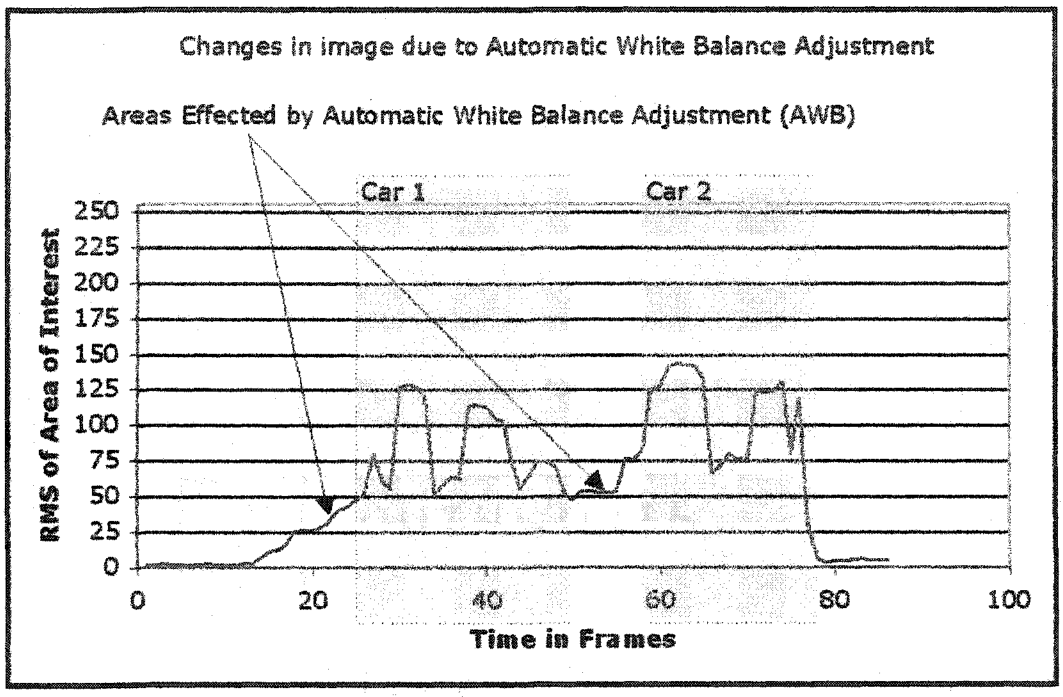


Figure 2 Example of changes in the color intensity created by Automatic White Balance

PREVIOUS RESEARCH

For the most part, previous work in this are has mostly avoided the issue altogether by using video that has been captured under controlled lighting conditions. There are many other difficult aspects to dealing with image and video data such as image recognition and motion detection [4] but the work presented here is has a goal of allowing other types of video sources which have been excluded due to the dynamic nature of their lighting. For short sequences over several minutes, one somewhat successful approach as been to take captured video, preview to select an area not occluded for the duration of the video and then apply some calculated scale factor to the images maintaining a constant value for that area through the sequence of images. While some readjustment is possible with this approach, is problematic for several reasons.

- It is time-intensive as the video must be previewed in its entirety before processing.
- The scaling factor cannot truly correct the problem as the transform function previously applied was not uniformly applied to the image as a whole.
- There might be no area within the image suitable for a scaling area.
- The selection of a scaling area creates a region within the image not suitable to be used as an area to detect motion.
- The selected region for scaling has to be isolated from the area of detection which minimizes its usefulness since the accuracy of the scaling is reduced as a function of it's distance from the area of interest.

The method proposed herein is a novel approach to processing video sequences to facilitate extracting motion data using the classical image processing technique of histogram equalization [5].

DEFINITION OF TERMS

Area of Interest (AOI): The area within the scene of a video image with potential for measurement of objects passing through the area over time.

Automatic White Balance (AWB): The means by which a video camera adjusts the captured image to allow a more natural appearance to the resulting digital image.

Histogram Equalization: The redistribution of intensity values in an image to allow an even distribution of values throughout the range available.

Pixel: Picture Element. A raster component of a single video image.

Scaling Region: An area selected as the region which should be of a constant value reference used as the basis for applying a transformation to the entire image.

CHAPTER II

REVIEW OF LITERATURE

The literature discussed in this section is focused on understanding the underlying principles and background information required to understand the impact of this effort. Processing this information and the techniques used draw from several fields of study including the psychophysics of color vision, the technology of video capture, and image processing techniques. The material presented here is an overview of the basic information needed to understand the principles and mechanisms this research utilizes.

Video images are stored in a computer as an array of integer values in triplets of the red, green, and blue components of the color [5]. They are constrained to values ranging from 0 to 255 for each component of a 24 bit color image. As the storage requirement for this type of image allocation is large, there are many image compression algorithms [6, 7, 8, 9, 10, 11, 12, 13, 14] for storing still and motion image data. This brings about a wealth of techniques to store the information with a reduction in size but there are two main strategies for storing this information. One technique is to maintain the original information in its entirety. This would be considered a lossless technique. The other one attempts to maintain just enough of the original information to reduce the size but still make it a fair representation of the original image. This type of technique is considered a lossy technique, as some of the information is lost and not recoverable.

HUMAN COLOR VISION

As shown in Figure 3, there are three parts to human vision: The physical focusing of light onto the retina of the eye, the reception of light by the photoreceptors in the retina, and the interpretation of the stimuli by the nervous system and brain. The following discussion is drawn from the literature of color appearance models [14].

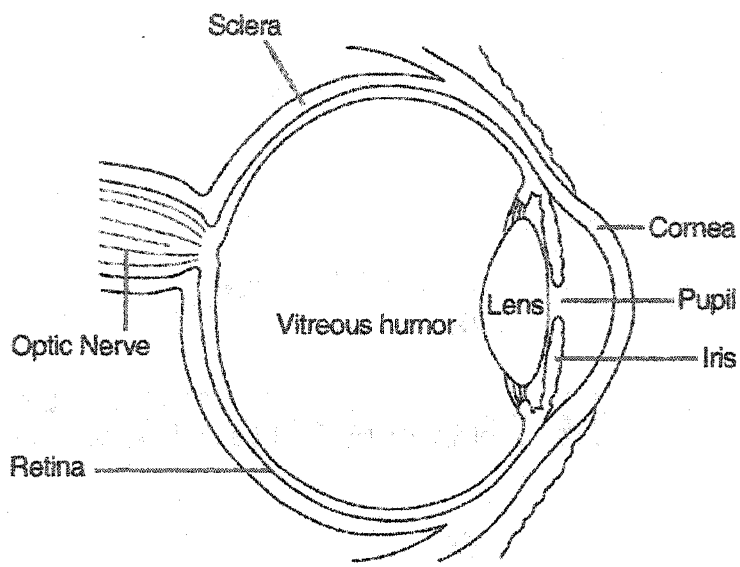


Figure 3 Diagram of internal anatomy of the human eye. (Courtesy of National Eye Institute, National Institutes of Health.)

The first portion is simple. The iris opens and closes to restrict the quantity of light entering the eye. The light passes through the lens and by the refractive nature of the lens and is focused on the retina. To bring objects into focus, the muscles surrounding the lens change its shape to alter its focal length.

The second part is more complex. The retina is composed of several types of photoreceptors named for their shape as rods and cones. The rods basically register light intensity over a broad spectrum of light wavelengths. The cones are more specialized. Each type of code will only register light of certain frequency ranges. They are of three types corresponding to their sensitivity to ranges of wavelengths of light associated with the colors red, blue and green. These rods and cones work together to detect light and characterize it as a nervous system signal that is then relayed to the brain for interpretation.

The third stage is mainly perceptual as far as color is concerned. The colors present in the image are transmitted to the brain and then interpreted to compensate for incident light, reflected light and the color of the object itself. The brain associates objects of known colors and then adjusts the perceived image to calculate the colors of other objects. It is through this mechanism, we can determine the color of an object under varied colors of light. There is a body of literature describing colorimetry, the description and specification of color, mostly building on the work of the Commission Internationale de l'Éclairage (CIE), which established a formal system to measure color in 1931. The exact definitions are not relevant to the work here but the manner of the measurement and comparison of color shows the extent of the problem. In order to do color matching, the specific viewing requirements are critical to evaluation of color. These requirements include lighting and time to acclimate the viewer to the viewing conditions.

VIDEO COLOR CAPTURE

The equipment used to capture video images in modern electronic systems mimics the way that the human eye detects light [15]. A lens and iris system, much the same as is present in the human eye, restricts and focuses the light on a detector which is analogous to the retina. Figure 4 shows a schematic diagram of a typical CCD camera.

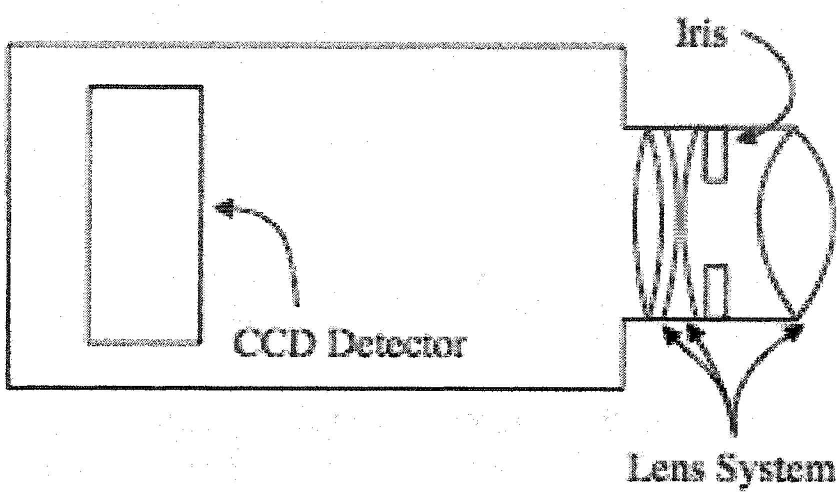


Figure 4 Schematic diagram of a CCD camera

An array of detectors on a silicon substrate are excited by incident light and the voltage generated by the light striking the substrate is measured to represent the number of photons of light striking the detector which is related to the light intensity. The red, blue, and green colors are isolated by either filtering or refracting the light into its individual red, green or blue components. The detectors are read in sequence to create an array of values representing the entire image. Varying types of filters and detectors are used depending on the nature of the scene. Most common detectors are the Charge Coupled Device (CCD) and Complementary Metal Oxide Semiconductor (CMOS). The resulting array of voltages is then passed on to circuitry which adjusts the image to make corrections in the image to approximately match the color that would be perceived by a human viewer of the same scene before final storage.

AUTOMATIC COLOR ADJUSTMENT

The signals sent from the retina to the brain are just raw data. For the color data to have any meaning, the color values mush be processed as a perceived image. This perceived image is then further subjected to adjustment by associating the colors to previously encountered colors for the same object. This compensation that we perform automatically is referred to as chromatic adaptation. An example of how this effects color perception would be a problem associated with traditional photography. When taking pictures indoors under incandescent lighting with photographic film intended to be used outdoors under natural light, the photographs appear yellow and do not match our memory of the scene. In actuality, the scene was yellow because the light of an incandescent bulb has more yellow than natural lighting but mentally we make the adjustment when we are under that type of light to correlate the colors that we see to the known colors in the scene.

The circuitry in modern video equipment has a functionality built in to perform an analogous function. This is the automatic white balance. Without the object memory referencing ability present in the human vision system to decide on the perceived color of an object, the only way that video circuitry can mimic this color adjustment is to base the adjustment on the color white as a representative of all colors of the spectrum. There are two main ways that this is accomplished. The first is to take the brightest color that is not totally saturated and closest to white, meaning that it is not outside the reliable detection range and that the three components of red, blue, and green are nearly the same value, and assume that it is white. The other is to assume that the average color of all things in the scene is equal to a uniform gray color. A transform function is then applied to the image or a portion of the image, resulting in a final image that approximates the way that a human viewer would perceive the same scene. Each manufacturer of video equipment creates their own proprietary transform function to match what they believe will appeal most to the consumers of their equipment, producing images which look natural and correct. Analyzing video adjusted as a result of these transform functions is a source of difficulty as these transform functions not easily reverse engineered. Many manufacturers compartmentalize the image into several sub-images and perform the adjustment regionally rather than apply the balancing to the image as a whole. The resultant color adjustment creates a problem for use other than the use that was the original equipment designer's intention. The video equipment was designed for the needs of a human viewer and not the needs of a computer. When analyzing the resulting video, the origins of any resulting change are indeterminate. It could be attributed to either the changing scene or the color adjustments made by the video circuitry or some combination.

DIGITAL VIDEO FORMATS AND HANDLING

In its simplest sense, the binary representation of still and motion video is a sequence of integer values representing the individual components of red, green, and blue colors. An example of that format would be a PNG (portable network graphic), which is either a text or binary file of the integer components of red, green and blue. Routinely, the storage is much more complicated for several other considerations. The red, green, and blue representation does not translate to other media such as print or film and therefore must be translated to another color space representation. Examples of this would be the CYMK (cyan, yellow, magenta, and black) or HSV (hue, saturation, and value) models. Other storage issues such as image size are dealt with by storing the images in either lossless or lossy formats such as GIF (graphics interchange file) or JPEG (a standard formed by the Joint Photographic Experts Group) for example [16]. These files must be decoded to perform any manipulation or interpretation of the included information.

Video image sequences are even more complex. They have the same size and media issues as still images. They also often contain audio stream information as well as the video information. To further exacerbate the problem, the standards by which the images are stored are set by standards but most of the encoding and decoding implementations (codecs) and libraries are proprietary and under patent, copyright, or protected as trade secrets. Many times, the only way to extract information about individual frames of a video sequence is to purchase a license or to use software which has reverse engineered the video codecs.

DESCRIPTION OF REMAINING CHAPTERS

The remainder of this document will discuss the methods used for performing this study, the results obtained, and the conclusions drawn from these results. The remainder will discuss the further application and the areas in which this should prove useful in aiding in the processing of image data from video.

CHAPTER III

SYSTEM ARCHITECTURE

The video used for this research was obtained from consumer grade video equipment. The sequences were taken of vehicular traffic moving through different scenes. The resulting video was separated into a sequence of individual frames using Transcode [17]. Transcode is a program that is designed to utilize different decoding and encoding algorithms to convert digital motion video into other formats. In this case it was used to take an input file of motion video and output to a sequence of individual frames of still images. Python [18] was used as a control language for the prototype application because of the availability of the associated imaging library and for the creation of the graphical user interface. The Python Imaging Library [19] was used to manipulate the images. This library has a wide array of functionality for processing images and collecting statistical information. The histogram equalization, grayscaling, and statistical measurement were done with this library. The interface for the program was created with the Tkinter [20] library, which is standard with the Python language. The library is widely available on many platforms and is based on the Tk graphical libraries originally created for the Tcl programming language. Graphing within the application was implemented with Gnuplot [21] as a convenience feature but is not reliable between different versions of Gnuplot for different platforms. The image compositing was done with ImageMagick [22] for comparison of images from different times in the video sequence. The tools in use are freely available and platform independent.

The resulting program as shown in Figure 5 presents the user with a graphical program to select the area of interest, the scaling area, and select from different processing options such as the histogram equalization and other filtering methods which are beyond the scope of this document. The other options are in current development and are not used in this time.

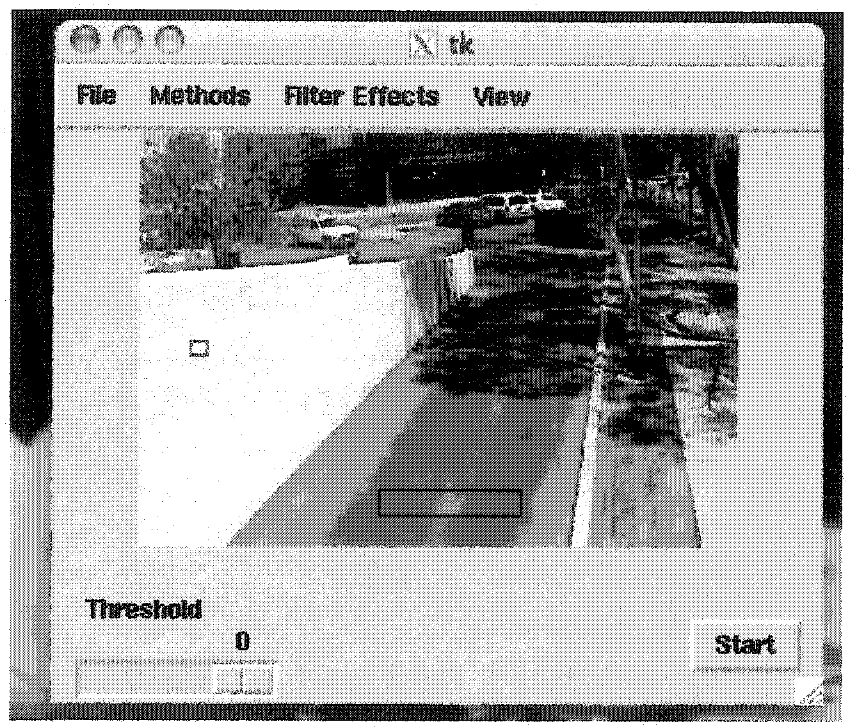


Figure 5 Screen shot of program.

The user can select the first image of the sequence and indicates the area of interest shown as a black rectangle and the scaling region shown as a red rectangle for the comparison of the effects. The results may be viewed in progress by selecting "show video" from the view menu. For generating the data used in this experiment, the user selects an output file name from the file menu to create the file for the values for each frame. The program will also generate a final graph of the data but it is dependant on the version of Gnuplot which is in transition on some platforms. It will work under the Windows version but not the Linux version at this time. It is only a convenience feature and not a supported feature. All of the source code and video sequences are available on the author's website [23].

In testing, the program was used on several machines with differing operating systems and processors. In all cases, the program yielded faster than real time processing of the video and would lend itself to real time data collection of live video and faster than real time data mining existing video sequences.

PROPOSED METHODOLOGY

Assuming the first images in all of the sequences were without any object in the area of interest (AOI). Both hard scaling and Histogram Equalization were performed with the same AOI to make a comparison of the efficacy of this technique. The general technique used in this experiment was:

- Use the first image in the sequence as a reference ($Image_0$).
- Select AOI and the scaling region.
- With all of the subsequent $Image_n$ in the sequence:
    - Create a grayscale version of $Image_n$ and perform the scaling and/or histogram equalization.
    - Create an abstract image that is the absolute value of the differences in $Image_0$ and $Image_n$. $|Image_0 - Image_n|$

- Calculate the Root Mean Square (RMS) of the intensity values of the AOI in the new image. Other metrics, such as sum of absolute difference, may be used for this step.

The resulting RMS values for the AOI were plotted in two sets of graphs for each image sequence, one with the value obtained with Histogram Equalization plotted against the raw image and the other with the value of a hard scaling against the same raw image data. The plot represents the amount of change in the area of interest as compared to the original image with nothing in the area of interest. The higher value represents more change from the original image. In an ideal case there would be a binary relationship between something being in the area of interest or not, but the individual variations in the image of the objects passing through the area of interest are not uniform. The different portions of the vehicles create a different amount of change. This is caused in variations in shape and material of the object. As an example, the graphs clearly show such components as the windshield due to the difference in its apparent color. This variation is unavoidable but can be limited in its effect by application of thresholds to distinguish a level at which the change can be considered positive indication of an object being in the area of interest.

CHAPTER IV

EXPERIMENTAL DATA AND RESULTS

The source of data for this project was sequences of video taken of different scenes of vehicular traffic under varying conditions. The original capture equipment for the first video file is unknown. The subsequent video files that were used to further test the process, were taken with a Panasonic digital camcorder and encoded into motion video files with a hardware MPEG encoder, in this case a Hauppauge PVR-250 video tuner and capture card but connecting the composite output of the digital camera to the composite input of the capture card. The transcoding program was then used to extract individual frames of these files as well.

The video used was representative of motion video obtained with normal video equipment and exhibited some of the usual problems associated with consumer grade equipment. These problems included slight camera instability and automatic focusing issues. The attempt was made to not seek out pristine video without any flaws.

The video used was 320x240 SIF encoded as NTSC at 30 frames per second. The first video was originally encoded to an AVI file and the later video sequences were encoded with MPEG2 encoding. For the evaluation of the technique, the images were converted to grayscale during this exercise.

VIDEO SEQUENCE 1

Video sequence 1 is the main sequence used for testing and development of the test programs. It consists of a scene that is an entrance lane to a parking garage. Along the left side of the scene is a white wall that dominates the scene. Figure 6 shows the graph of the raw video where the x-axis is the time in frames and the y-axis is the amount of change in value of the RMS of the area of interest (as the white rectangle box shown in images of Figure 7).

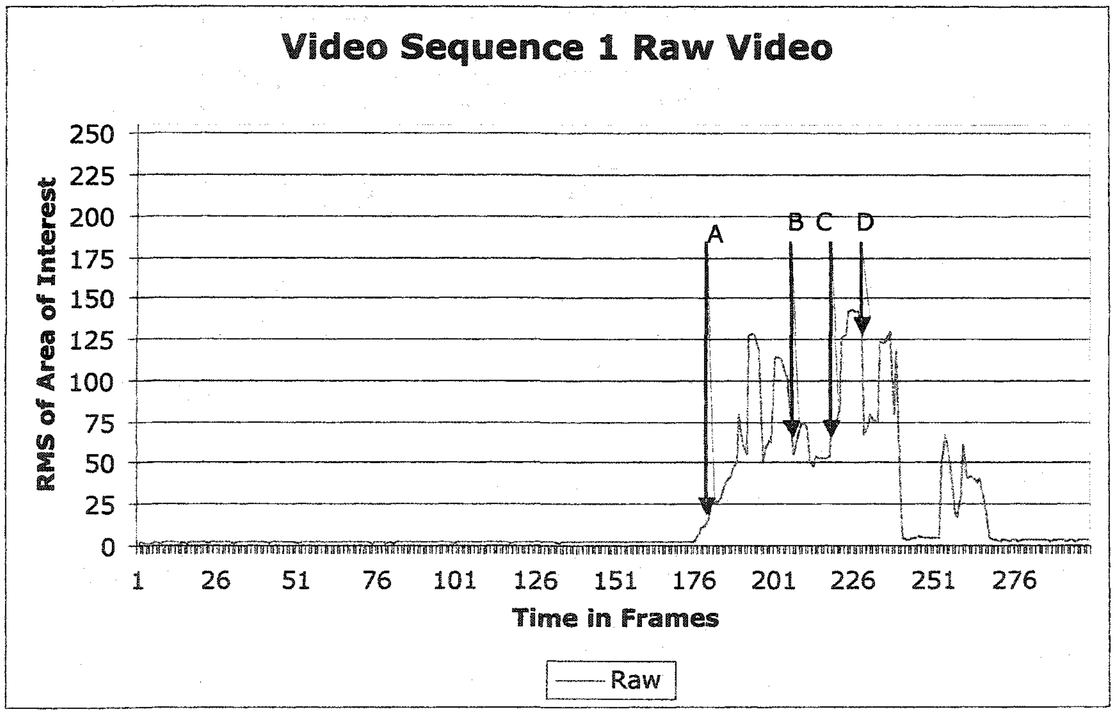


Figure 6 Raw Video Data

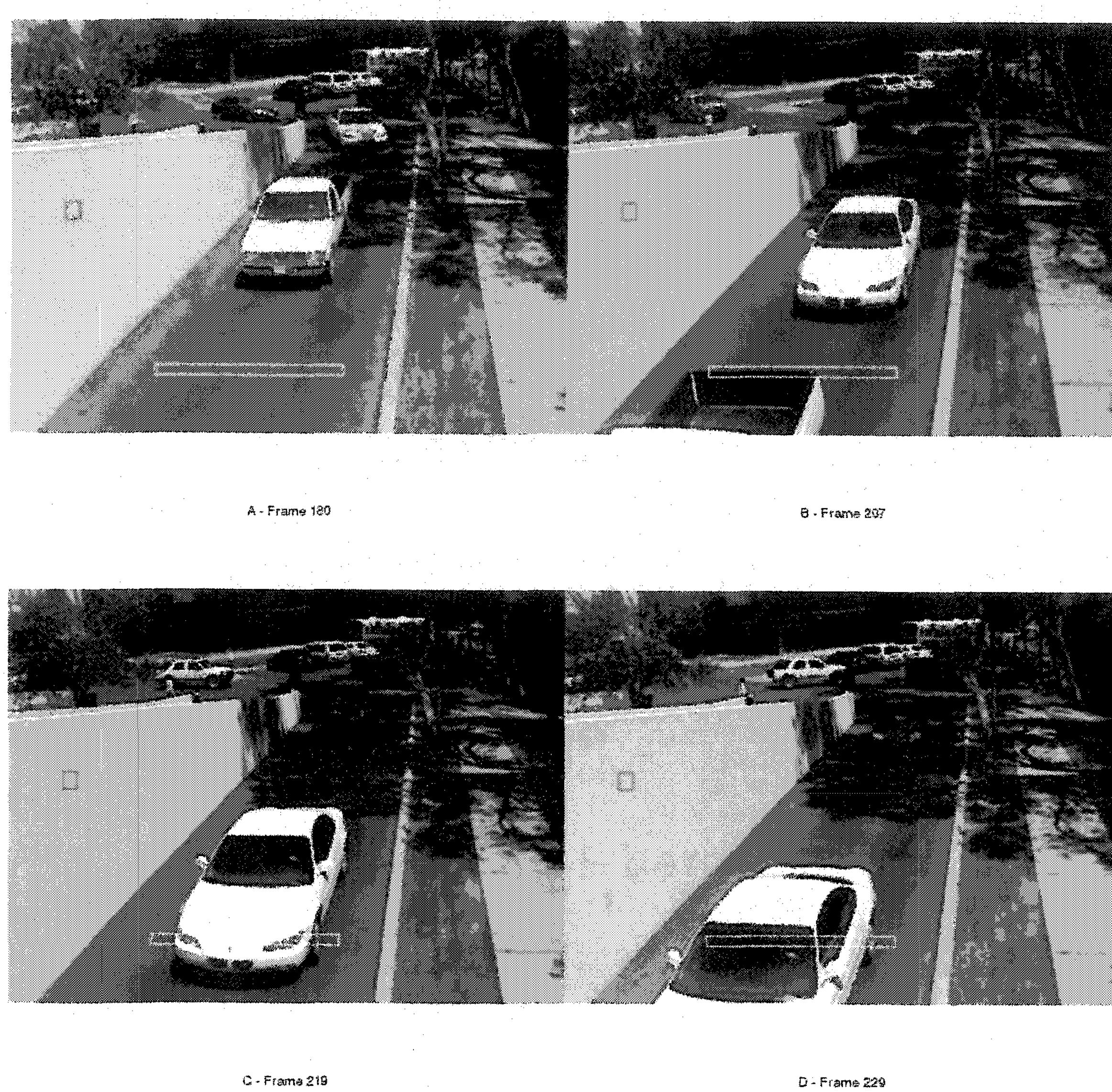


Figure 7 Video Sequence 1 Sample Images

Figure 7 shows four sample images from the video sequences which are the frames indicated in Figure 6 by the arrows. The image labeled A shows the preceding of the object's entry into the area of interest by the shift in color by the automatic white balance. The image labeled B shows the vehicle about to leave the area of interest and examination the graph in Figure 6 shows that the return to the base condition is not present because of the shift in color brought about by the automatic white balance. This video exhibited the worst example of the problem caused by the automatic white balance. The baseline detection level was between a value of one and two for the RMS of the area of detection. When the automatic white balance adjustment was made, the same area without an object in the area of detection was shifted to a value of 50. This made it extremely difficult to discriminate the end of the passing of the first car and the entry of the second car. With the hard scaling coded to the area on the wall on the left of the scene, the RMS of the area of interest between the two events was brought down to about 30. The histogram equalization method applied to this video sequence brought the RMS of the area between the two events to the RMS of the area of interest down to a value of 7. Figure 8 shows the comparison of the effects of using a scaling region (as shown in the red square box in the image) compared to the original raw video and Figure 9 shows the comparison of the histogram equalization plotted against the original raw video.

Also of note in images C and D in Figure 7 is the ability to discern structural features of the vehicles which are evident in the graphs. The different intensities of various features such as the bumpers and windshields of the vehicles can be seen as distinct features in the graph.

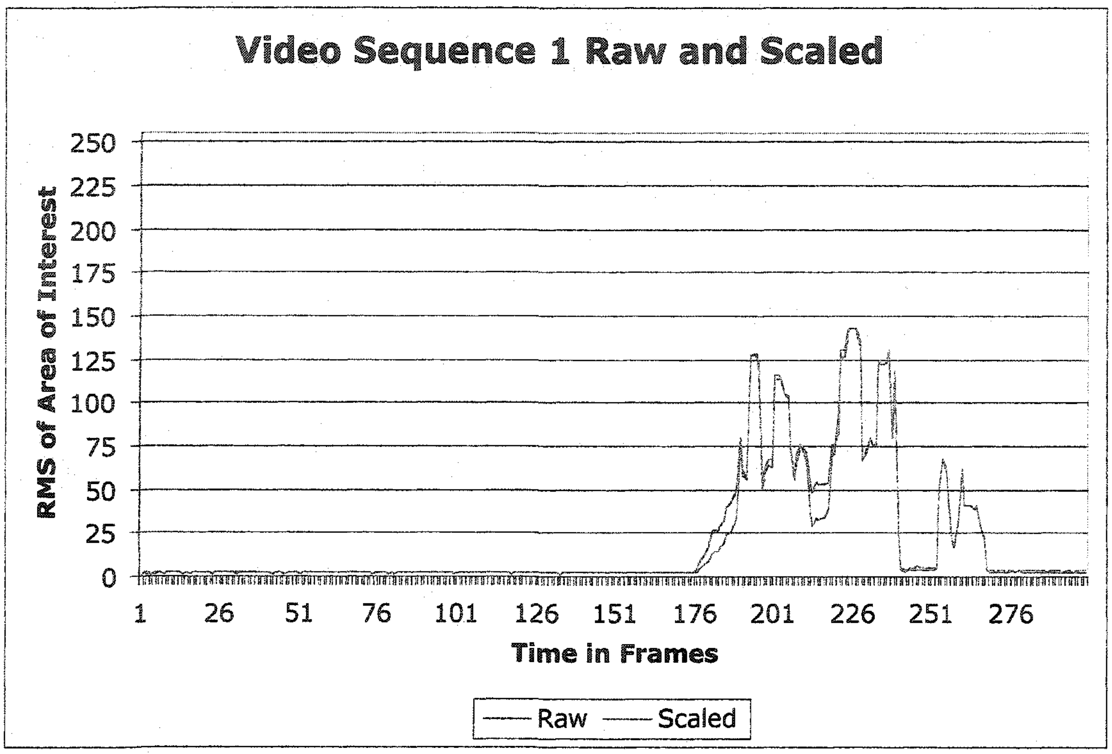


Figure 8 Comparison of raw video to scaled video for sequence 1.

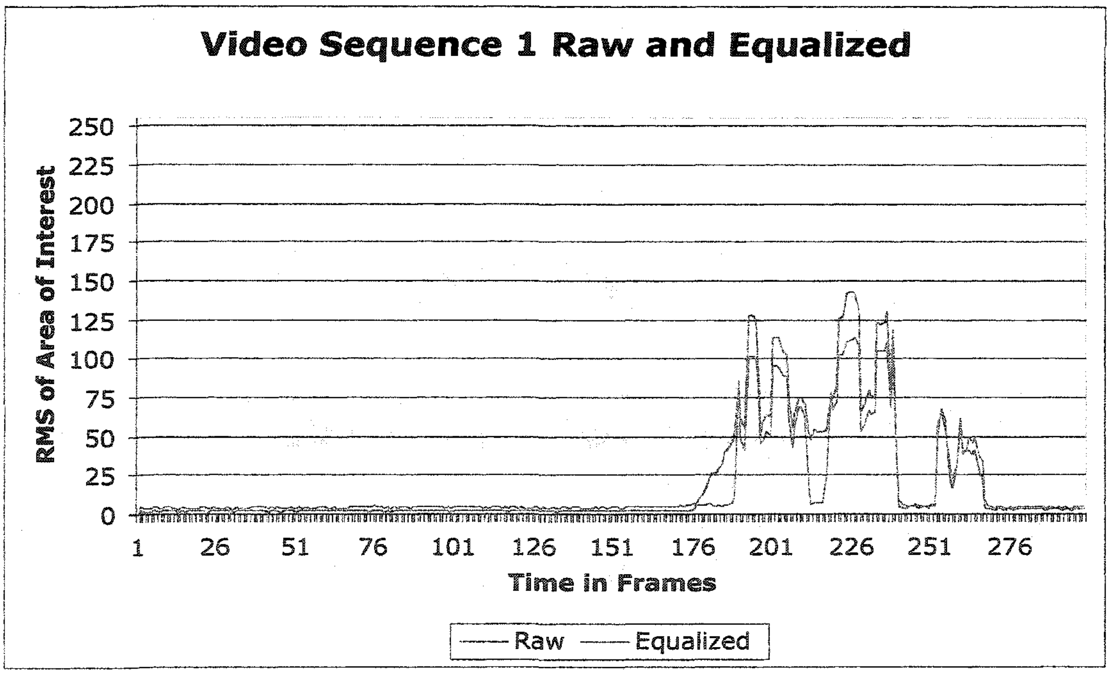


Figure 9 Comparison of raw video to histogram equalized video for sequence 1.

The results in this video were positive. It is easy to discern the gap between the two cars in the graph of the values obtained with histogram equalization. The automatic white balance problem was reduced to the point where it would be possible to set a threshold level which would discriminate each of the vehicles passing as a separate event. The histogram equalization method outperformed the scaling method by reducing the level between the two cars from a RMS of 30 found in the scaled version to an RMS of 7 for the histogram equalized version and had none of the negative issues associated with the hard scaling.

VIDEO SEQUENCE 2

The second video sequence is of a series of vehicles entering the access road for an expressway frontage road. Each vehicle pulls up to a stop sign and waits for an opportunity to enter the frontage road. Figure 10 shows the graph of the raw video.

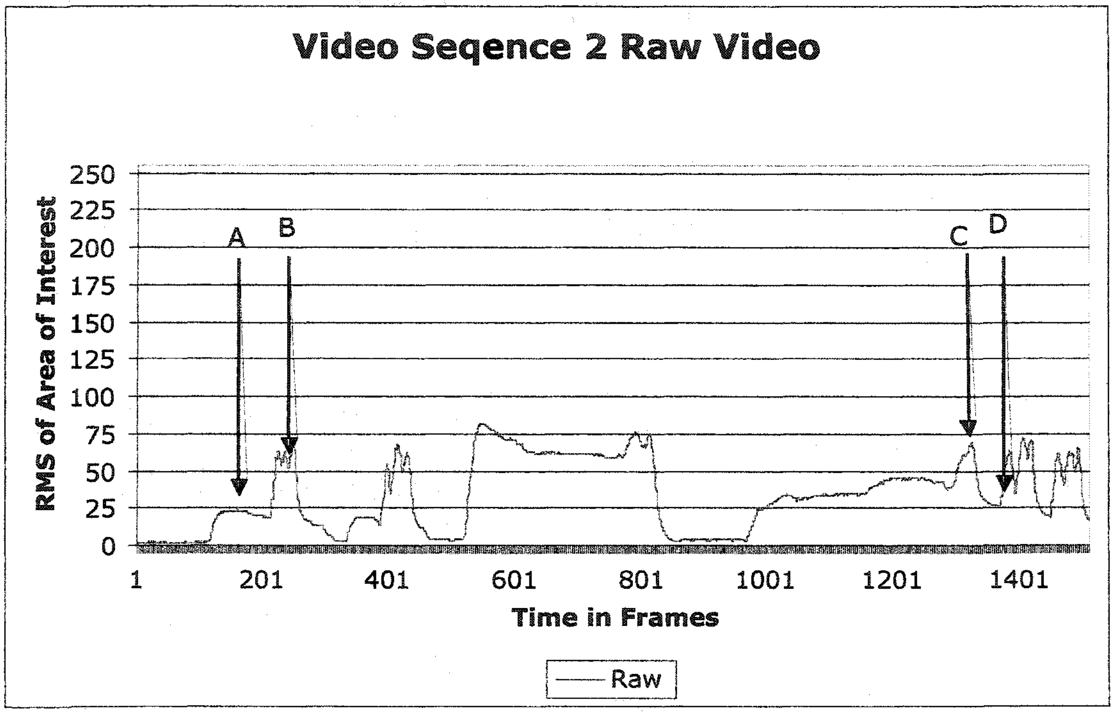
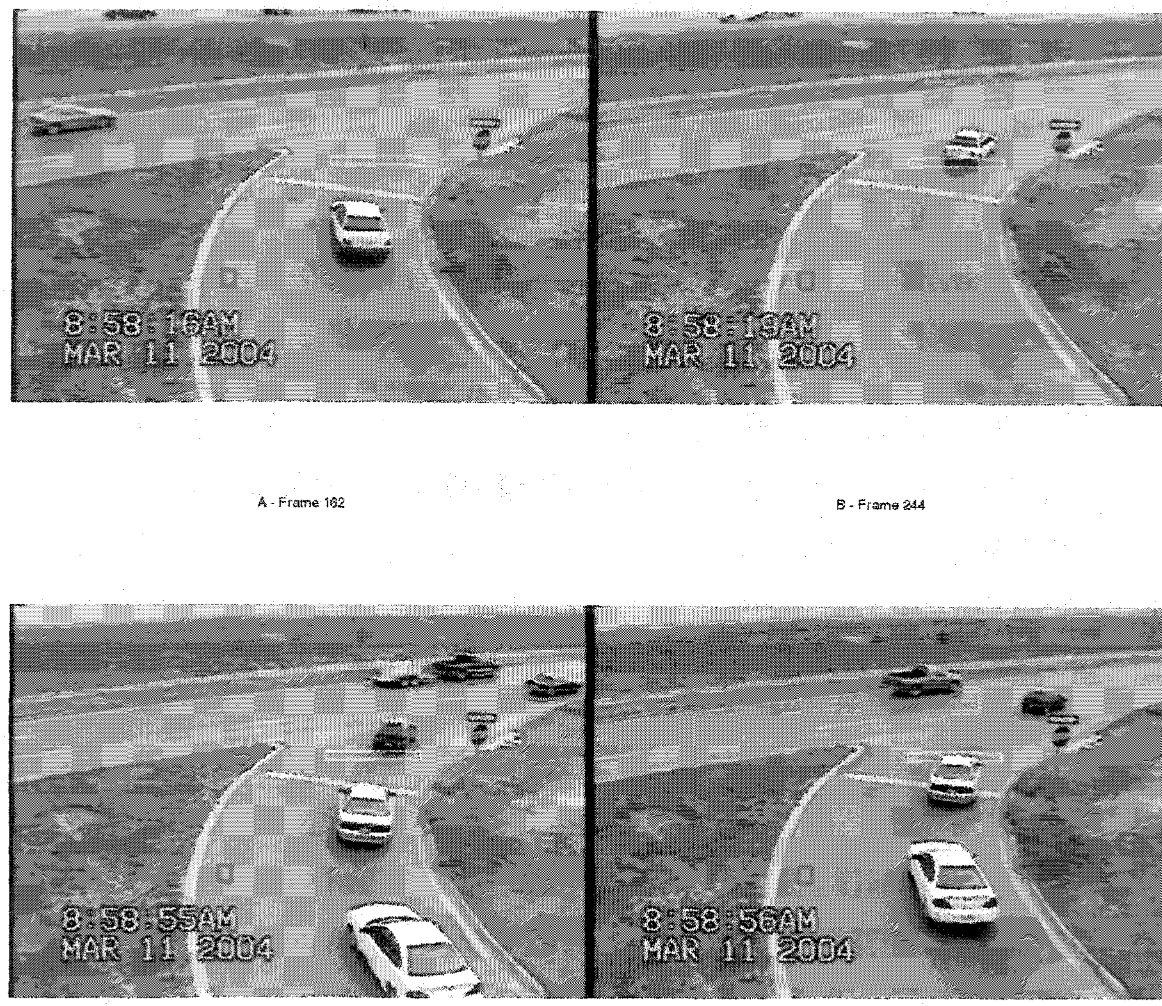


Figure 10 Video Sequence 2 Raw Video Data



C - Frame 1324

D - Frame 1371

Figure 11 Video Sequence 2 Sample Images

Figure 11 shows four frames of the video taken at the points labeled on Figure 10 with arrows. The selected area of interest is across the road near the stop sign, as shown in the white rectangle box in the image. The long areas of detection between frames ~600 and ~800 and between ~1000 and ~1300, as shown in Figures 12 and 13, are where a vehicle was waiting at the stop sign. There are a total of 6 vehicles with the third and fourth having an extended wait. Of special note are the first two white cars. Surrounding them on the graph of the raw data, the shift of the values in the AOI caused by the automatic white balance adjustment precedes the object's entry into the area of interest. As the individual events are well separated in time, the automatic white balance has the opportunity to rebalance between the vehicles. The only other difficulty in the raw video at this point is to discern the actual entry into the area of interest near frame 210 and 400 for each car respectively. The two segments between the last three cars show the most drastic example of the problem of the white balance adjustment made by the camera. The baseline of the graph between vehicles four and five is almost at the same level as when the fourth vehicle entered the area of interest. With the scaling method, it would be impossible to set a reliable threshold which would identify the passing of the last three cars as separate events. Of further concern, the automatic white balance adjustment made preceding the vehicle passing through the area of interest makes the exact frame in which the vehicle enters difficult to determine. Figure 12 compares the scaled video method to the raw video and Figure 13 compares the histogram equalized video to the original raw video.

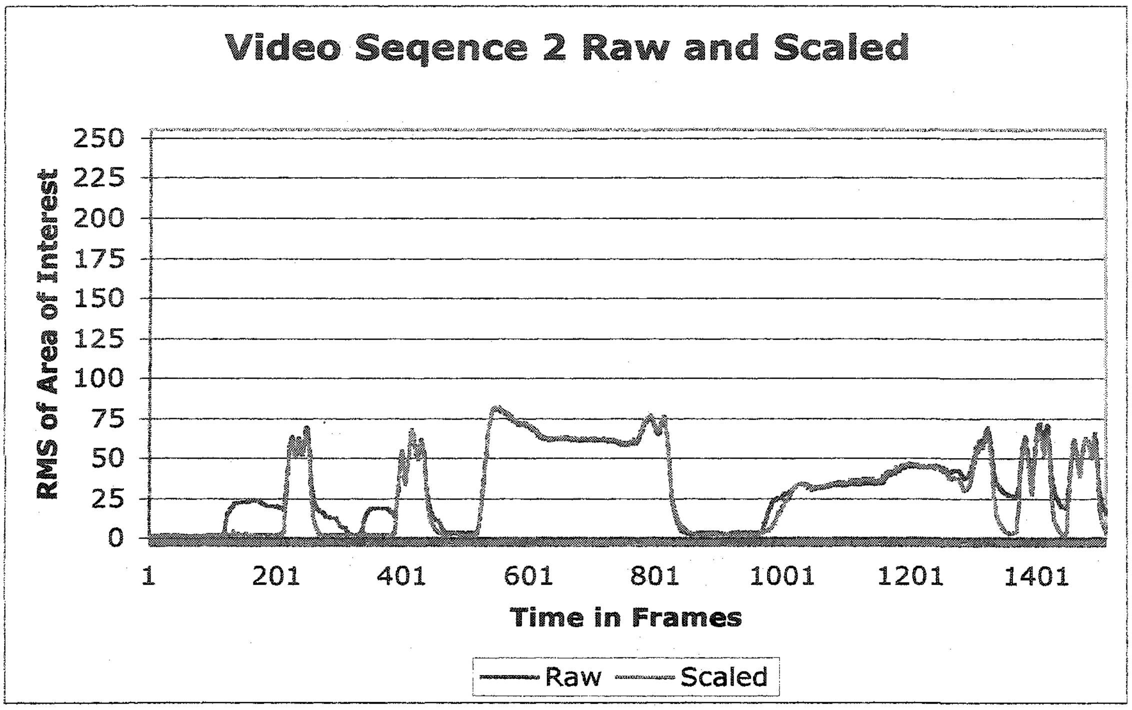


Figure 12 Comparison of raw video to scaled video for sequence 2.

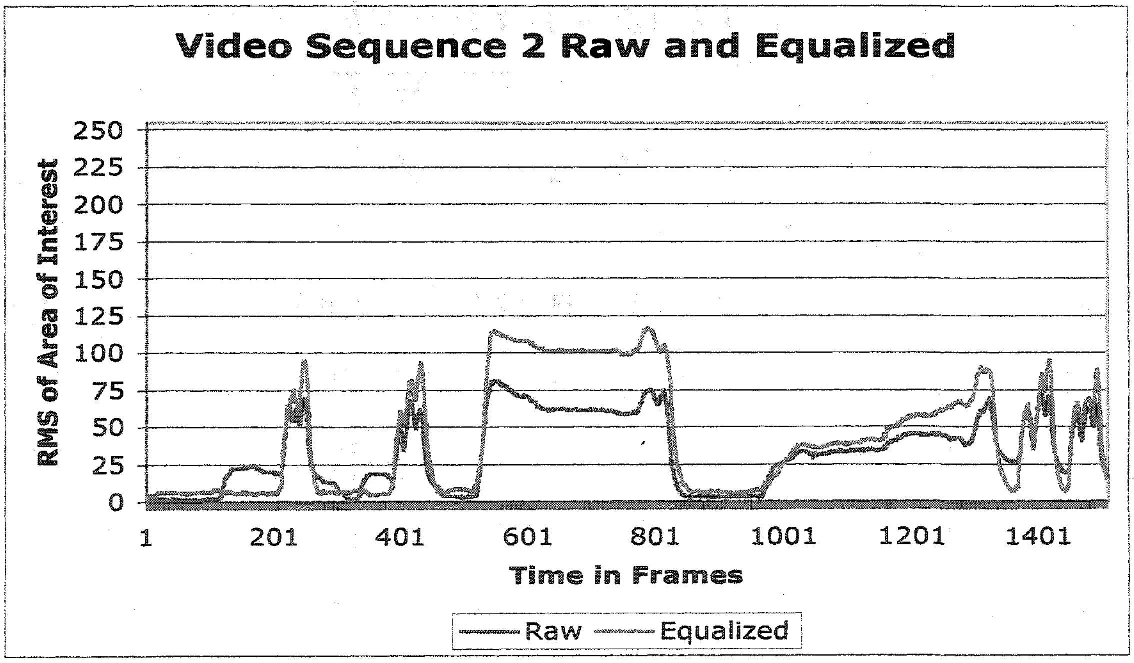


Figure 13 Comparison of raw video to histogram equalized video for sequence 2.

In this sequence, the hard scaling method actually worked well but not only did the histogram equalization method work as well, it also had the added benefit of increasing the signal to noise ratio by increasing the signal strength. Not only did it reduce the problem originally intended, it had introduced a new benefit.

VIDEO SEQUENCE 3

The third video sequence is also the entrance to a frontage road but from a different viewing angle. In this video sequence, two cars enter the frontage road from an entrance lane. Figure 14 shows the RMS plot of AOI of the raw video and Figure 15 shows the four video frames referenced by the arrows in Figure 14.

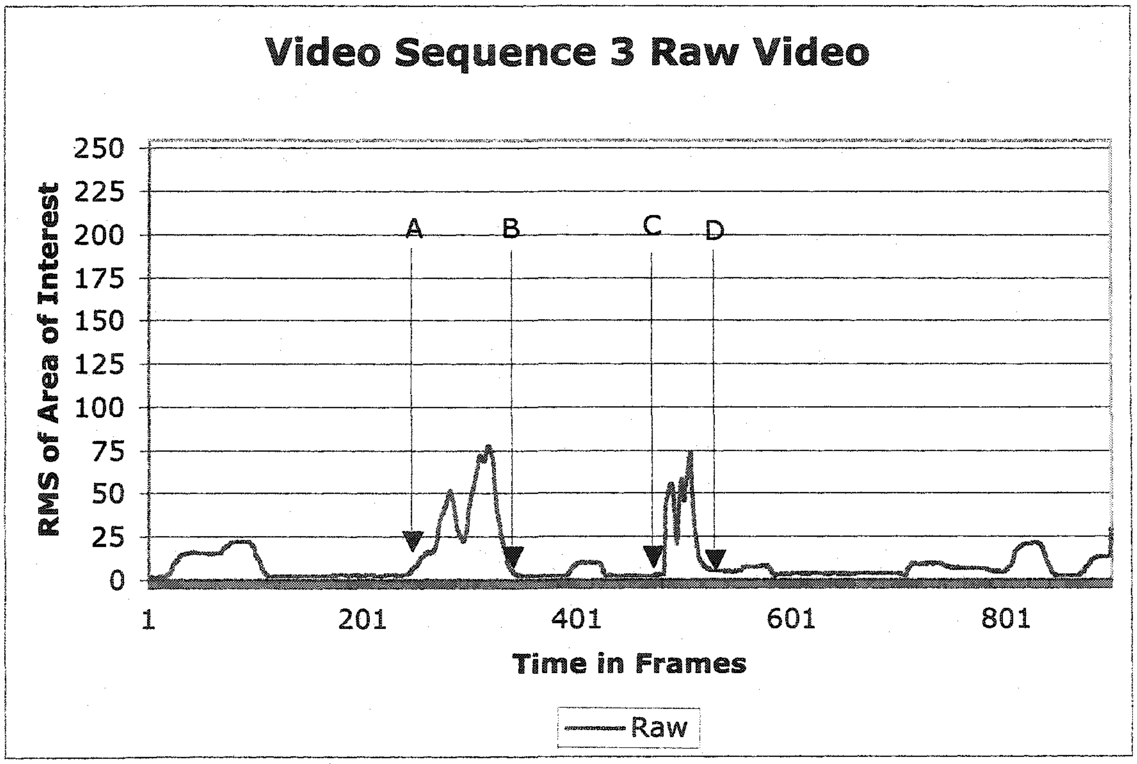


Figure 14 Video Sequence 3 Raw Data

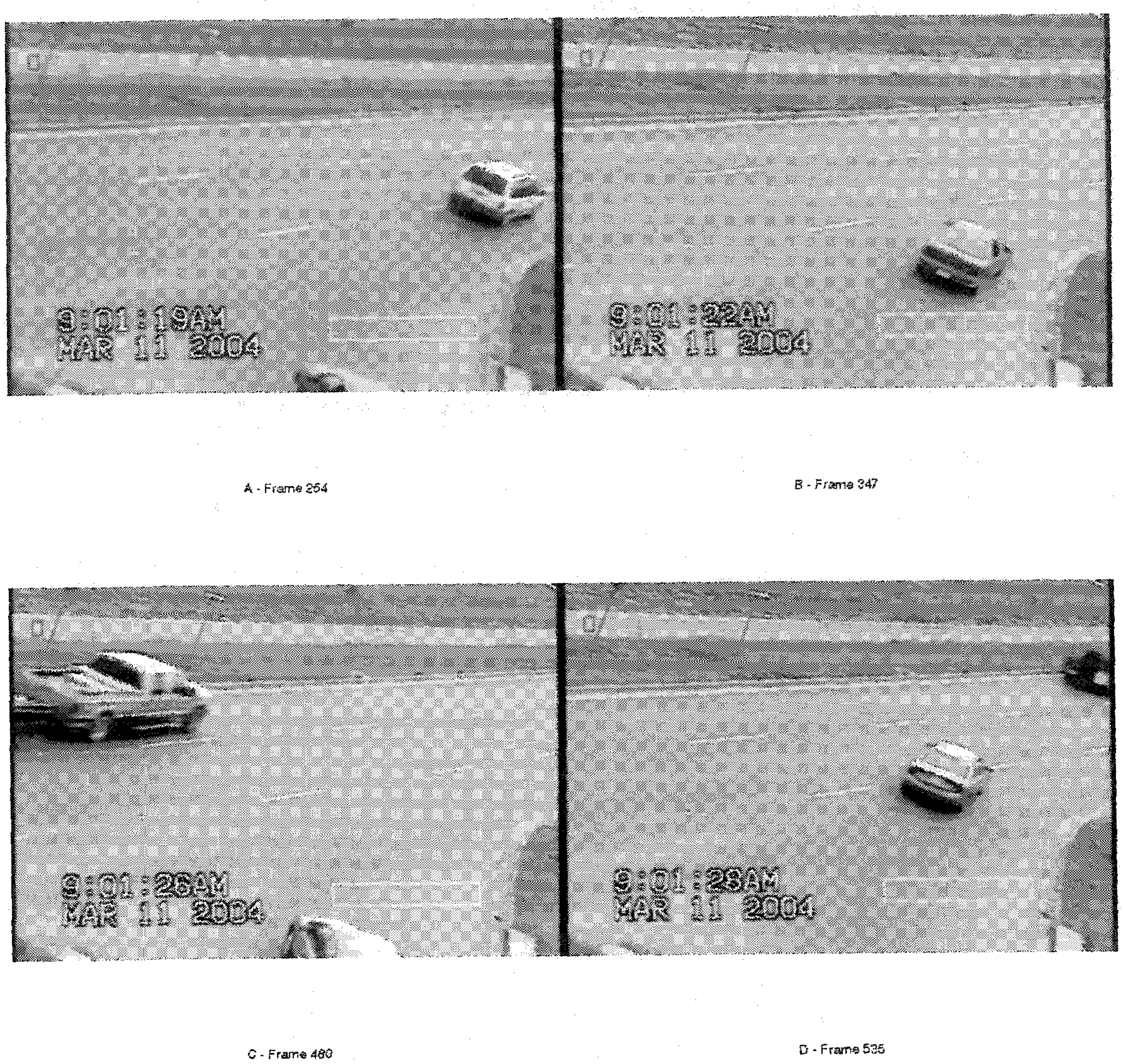


Figure 15 Video Sequence 3 Sample Images

The raw video in this sequence is affected by the passing of seven cars in the frontage road lane. Although they do not enter the area of interest, they each cause a shift in the white balance which somewhat lower in strength, can be seen as a definite pulse on the baseline of the graph which could be misinterpreted as a vehicle moving into the area of interest.

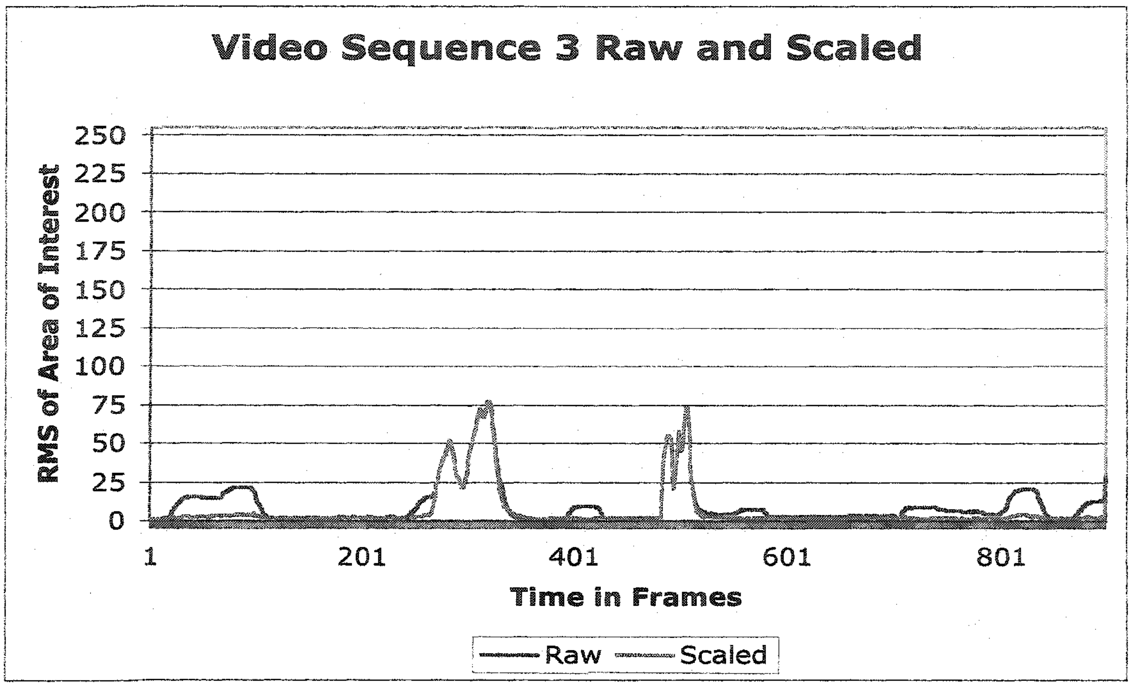


Figure 16 Comparison of raw video to scaled video for sequence 3.

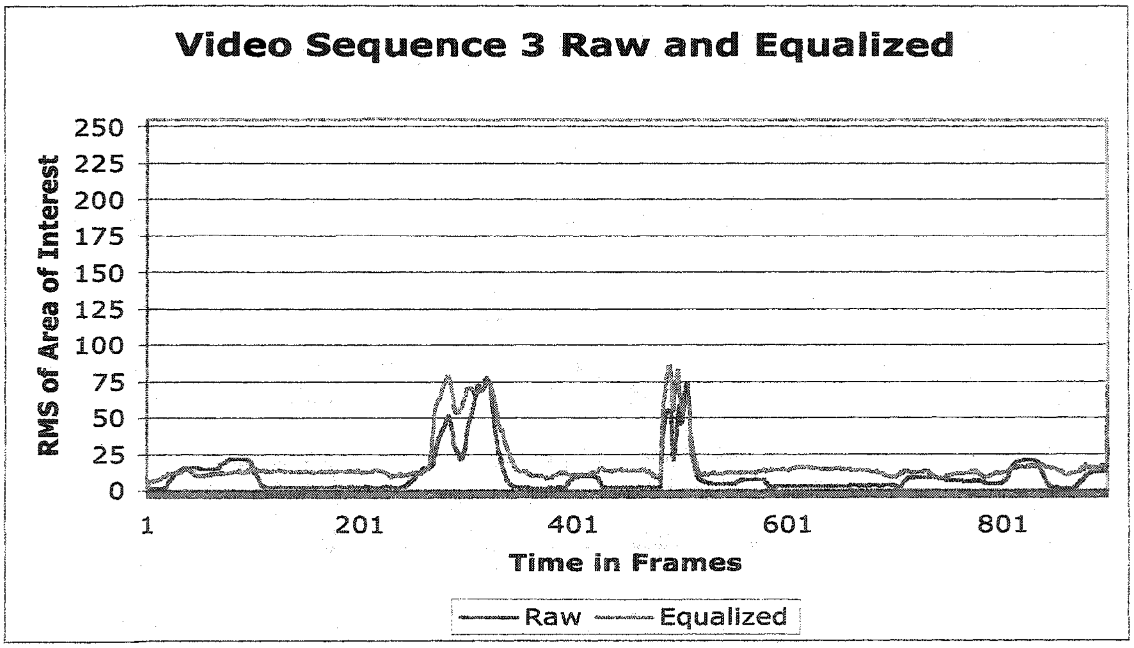


Figure 17 Comparison of raw video to histogram equalized video for sequence 3.

Again, as shown in Figures 16 and 17, the scaling performed well but the histogram equalization method performed comparably without the negative issues associated with scaling the video images. In addition, the individual variations in during the detection caused by irregularities in the vehicle passing through the area such as the darker area of the windshield were lessened by this method.

VIDEO SEQUENCE 4

This video is taken from an overpass looking down on traffic traveling away from the viewer. This scene is longer than the other scenes and has several issues that make scaling the image difficult. The graph of the original raw video is shown in Figure 18 with the points labeled to represent the frames found in Figure 19. Without enhancement, it is difficult to see the diffuse shadows caused by the vehicles but they are shown by the rise in the RMS of the graph.

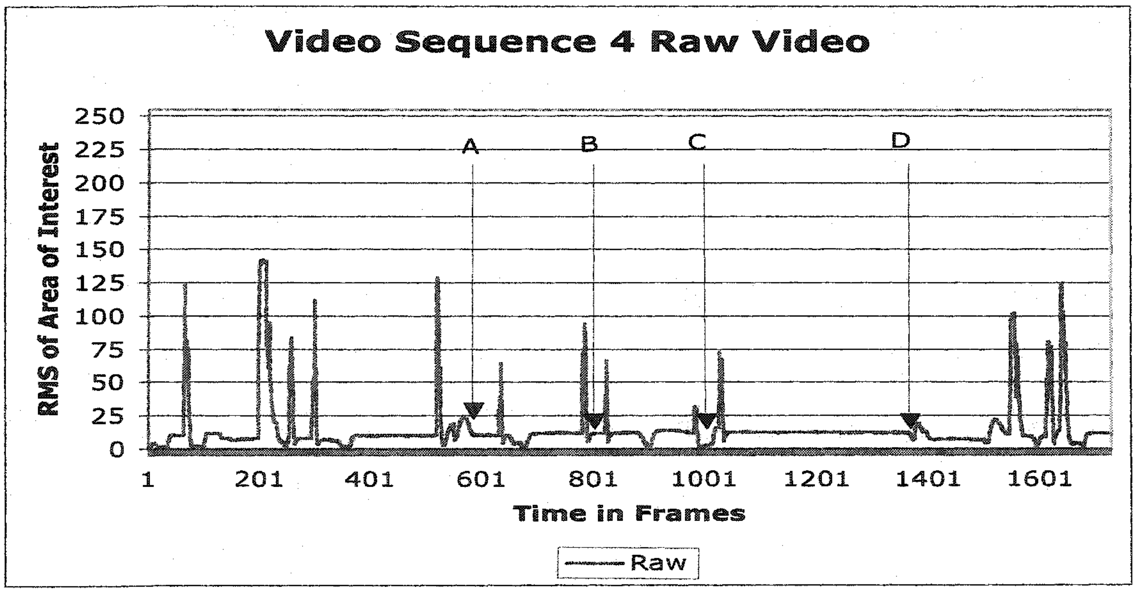
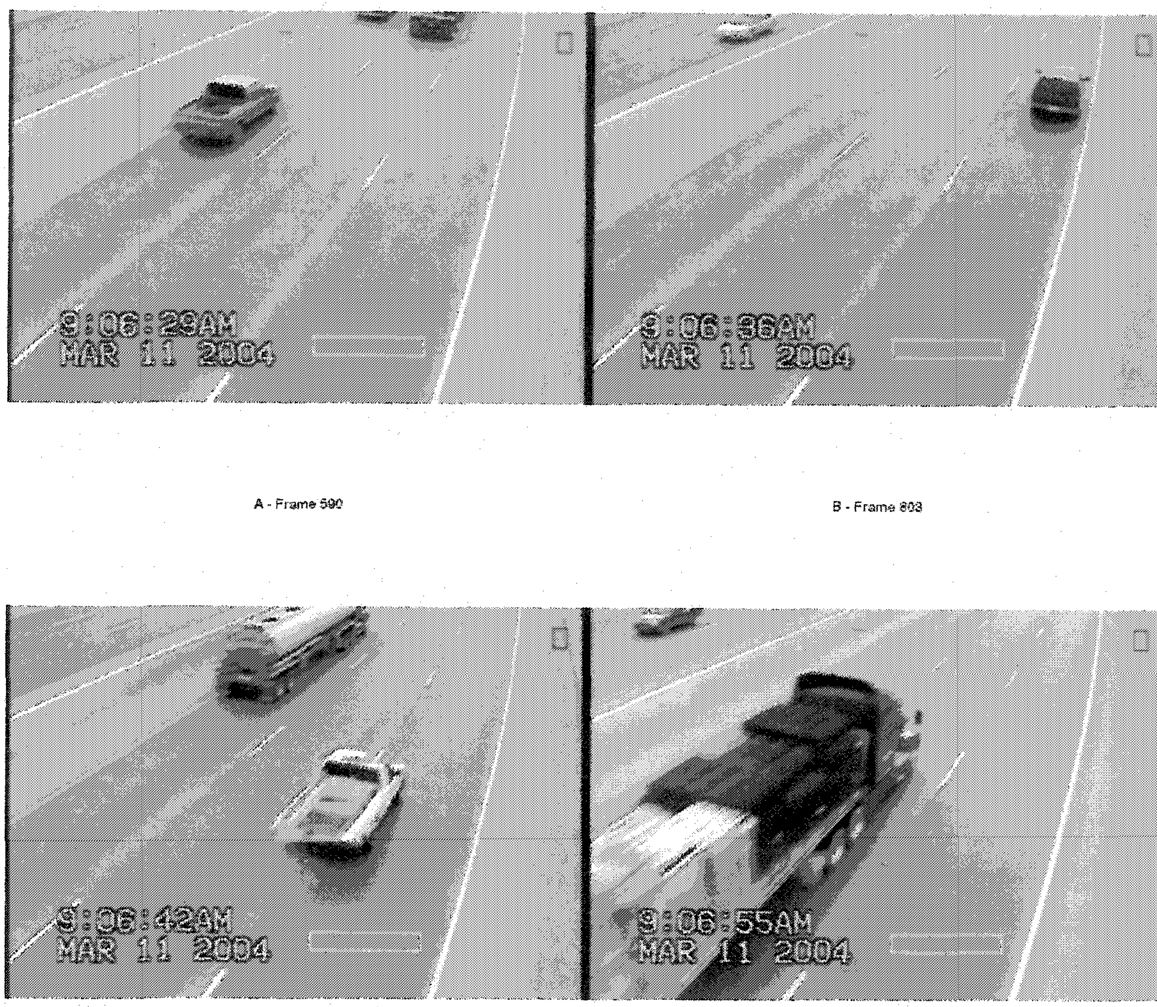


Figure 18 Video Sequence 4 Raw Data



A - Frame 590 B - Frame 803

C - Frame 969 D - Frame 1376

Figure 19 Video Sequence 4 Sample Images

There is not any area that would be acceptable to set a safe scaling area. The traffic lanes extend almost to the edge of the frame. The narrow shoulder of the road is crossed by the shadows of the vehicles which make for a shift that should be there rather than the shift caused solely by the automatic white balance.

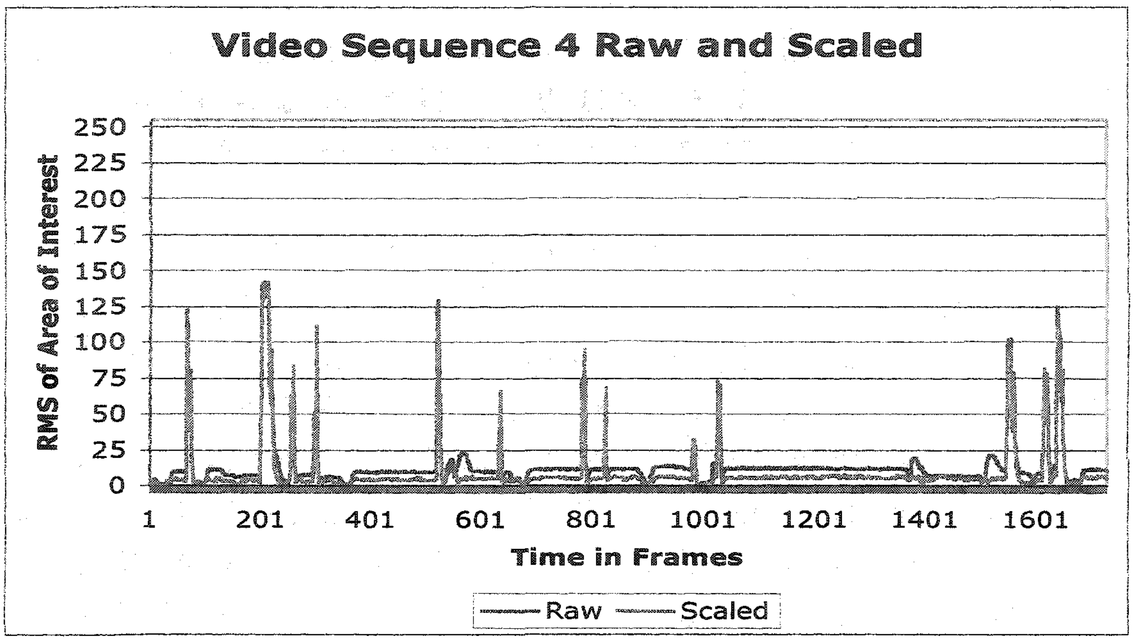


Figure 20 Comparison of raw video to scaled video for sequence 4.

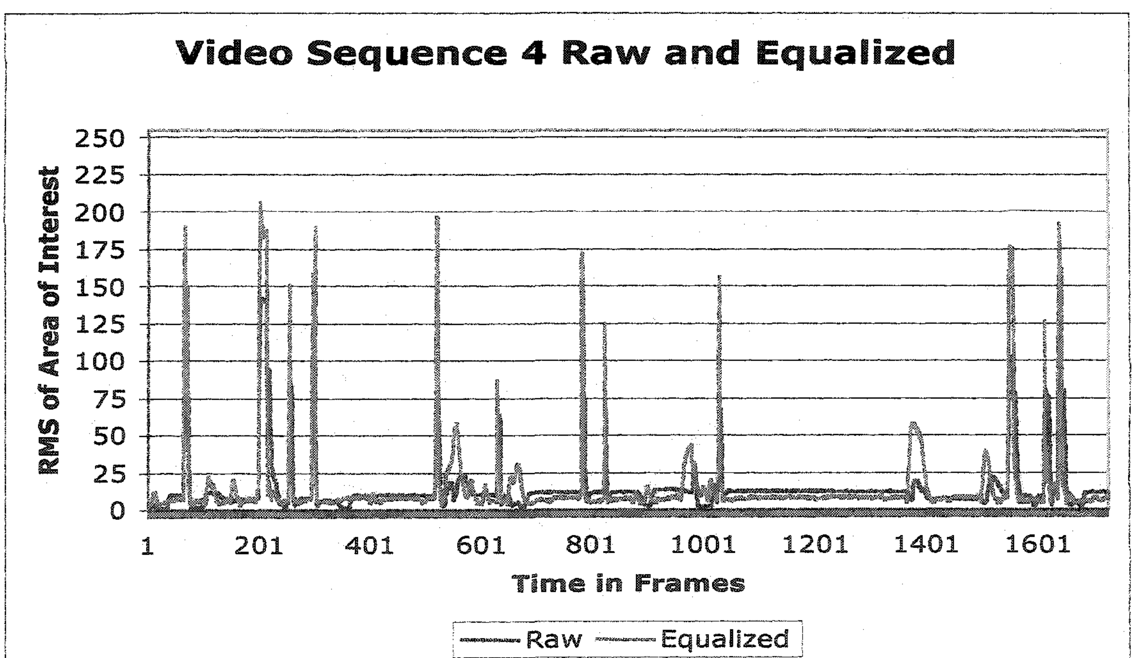


Figure 21 Comparison of raw video to histogram equalized video for sequence 4.

At first glance, the scaled version, as shown in Figure 20, appears to have a smoother response during the section between vehicles but on closer inspection, something more interesting is happening. The increase in the level caused by the shadows of the passing vehicles is preserved in the histogram equalization method as shown in Figure 21. The scaled version eliminated change caused by these shadows. As the shadow is a valid portion of the image of an object, the scaling method removed information which should be associated with the object.

VIDEO SEQUENCE 5

This video sequence is from the same overpass overlooking the freeway but in this case, the traffic is coming from the opposite direction. Figure 22 shows the graph of the raw video and Figure 23 shows the frames indicated by the arrows on Figure 22. This video sequence shows the indeterminate origin of the changes brought about by the automatic white balance. Only in the frame labeled D in Figure 23 is there something in the frame which could be attributed to causing the change. The lighting of the scene can be affected by the objects passing on the other side of the expressway which is not visible in this view.

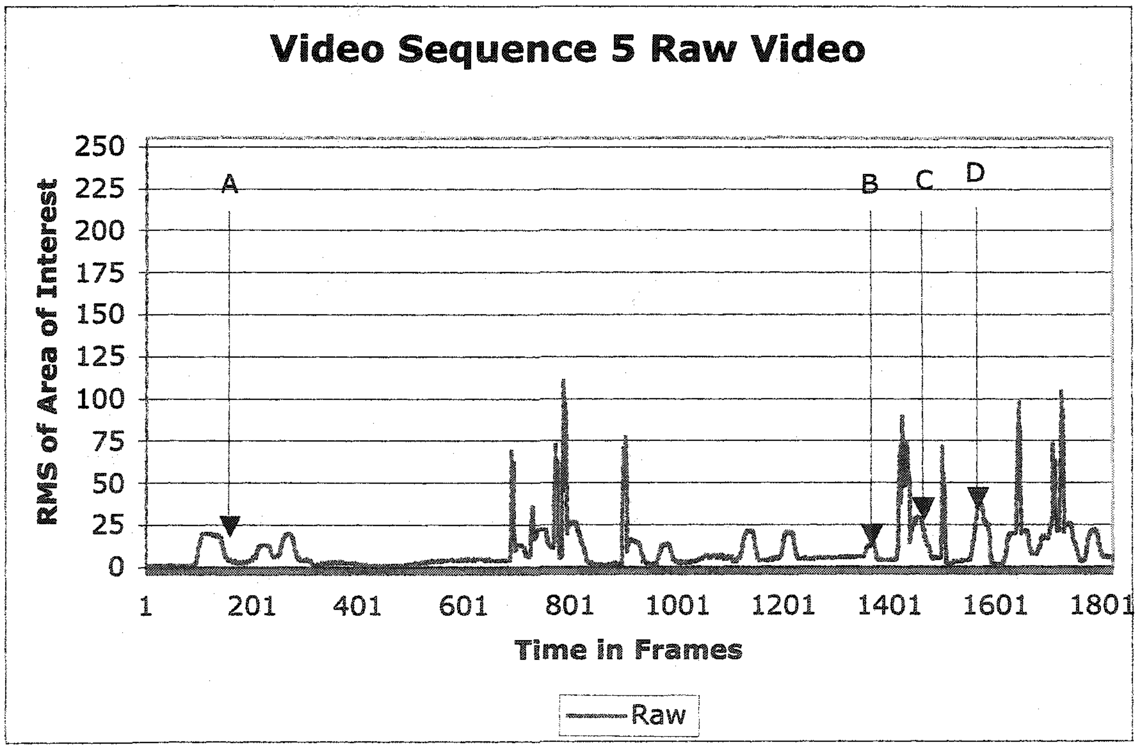


Figure 22 Video Sequence 5 Raw Data

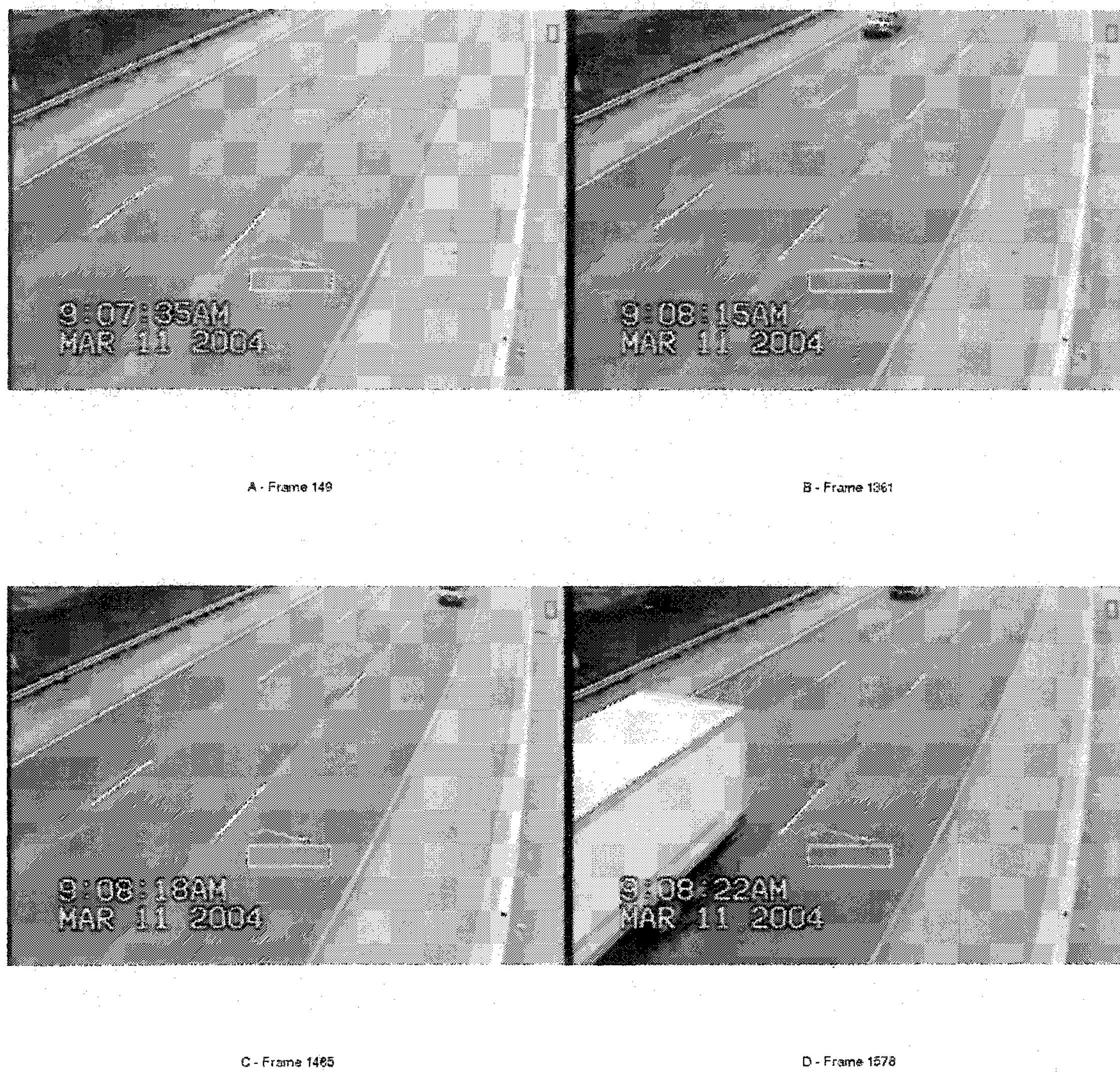


Figure 23 Video Sequence 5 Sample Images

The same problem of shadows and lack of an area free from passing traffic for selection of a scaling area is present as in video sequence 4. In this case, the results in the comparison of applying a scale factor to the frame with the results obtained by performing histogram equalization confirmed the same results as in the previous sequence as can be seen in Figures 24 and 25.

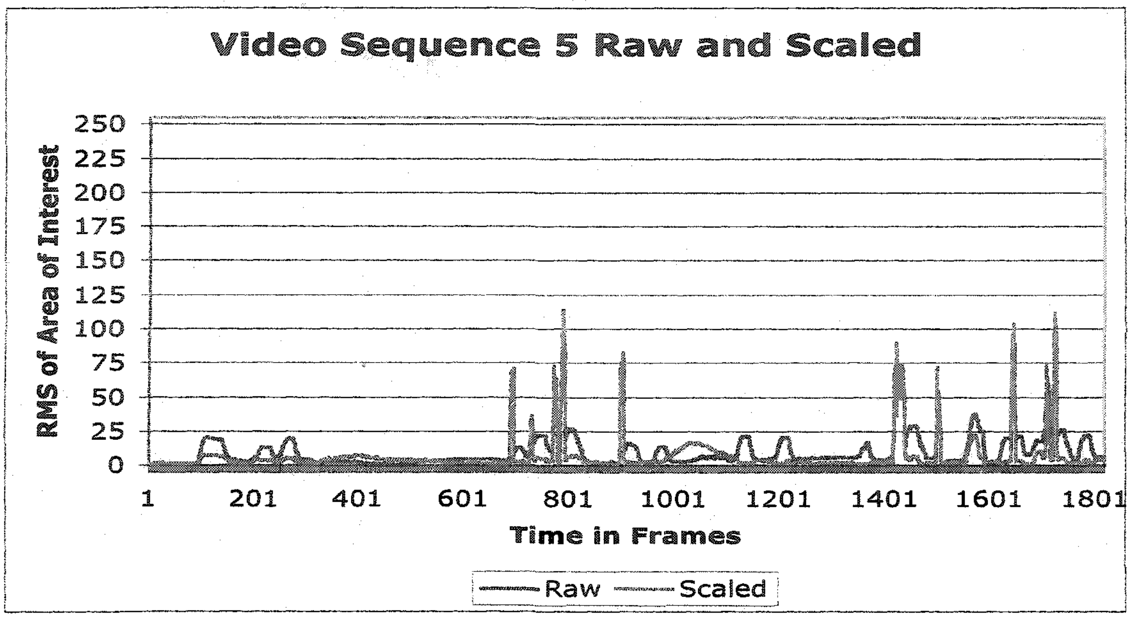


Figure 24 Comparison of raw video to scaled video for sequence 5.

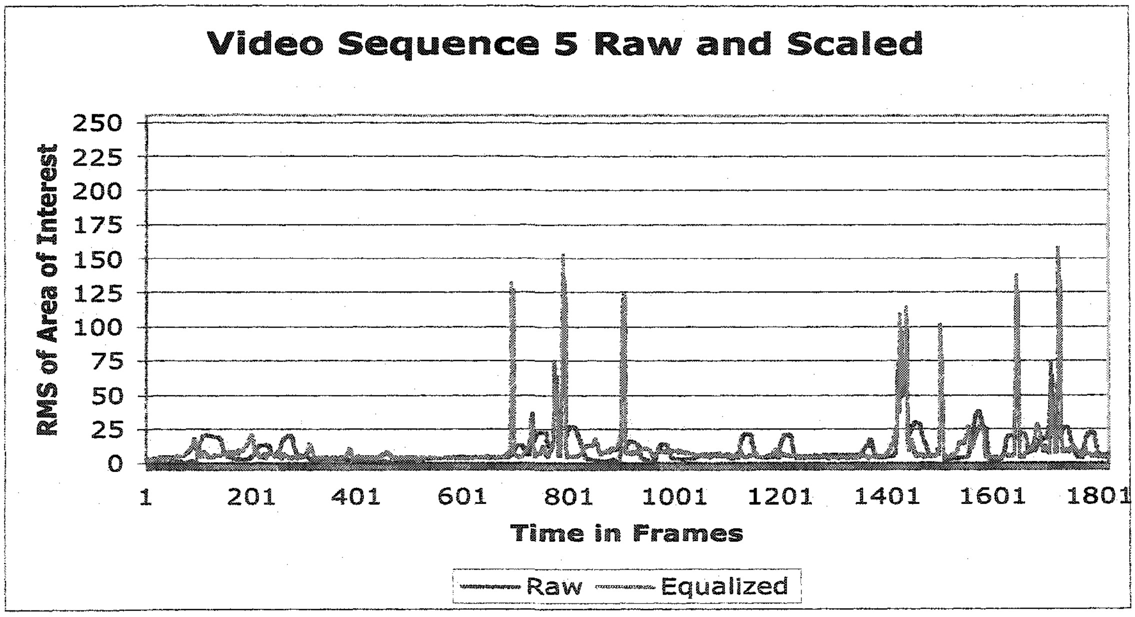


Figure 25 Comparison of raw video to histogram equalized video for sequence 5.

RESULTS

In all cases, the results of the histogram equalization performed at least as well as the scaled method but had the additional benefit of not having the negative problems associated with the scaling method. All the video sequences showed more accurate detection of objects moving through the AOI using the histogram equalization method over the hard scaling method. The actual point of entry of the objects into the area of interest was easily detectable by normalizing the effects of automatic white balance corrections made by the video capture equipment. The Histogram Equalization method avoided the problems inherent in the hard scaling method. The histogram equalization method did not require the previewing of the sequence. It allowed arbitrary selection of an AOI in the frame. Histogram equalization also preserved less distinct features of the video such as the shadows of passing objects which would allow more sophisticated techniques needed for object recognition.

CHAPTER V

CONCLUSIONS

In the samples tested in this research, the use of histogram equalization is shown to be a useful tool in compensating the effects of automatic white balance in analyzing motion video. The elimination of the need to preview the video sequence dramatically reduces the time needed to collect data from motion video. The changes caused by the automatic white balance preceding the entry of a vehicle into the area of interest allows more accurate timing of the entry and exit events. The reduction of the effects of automatic white balance correction reduces the influence of objects outside the area of interest creating changes within the area of interest. The elimination of a restricted area devoted to a scaling area allows potentially every pixel or group of pixels to become a potential area of interest. By establishing multiple areas of interest, it is possible to perform more complex analysis of the video. The preservation of subtle associated changes such as the shadows of objects opens the possibility to use more advanced image recognition techniques such as shape determination by three dimensional analysis.

Most of the uses for video data are currently under investigation but the real contribution of this research is to enable the use of video taken under less controlled circumstances such as outdoor video to be used where previously, it was restricted to using video sources which collected data under controlled conditions.

FUTURE WORK AND OPEN ISSUES

The main benefits of this technique are especially applicable to data mining of video. The arbitrary selection of AOI allows for detection of more dynamic events such as objects passing between two or more AOI. For example, with two AOIs in a bank ATM drive through location, we can use the video recoded by the security camera to answer the question like "What are the average waiting/service time per customer?" With overhead traffic video it is possible to monitor waiting times at intersections, traffic patterns collected per lane for selection of commercial properties, and monitoring parking lot utilization. Using similar techniques coupled with programming triggered by events, it would be possible to monitor things such as average wait time in single feed queues such as monitoring a bank teller line to signal the need to open or close teller booths based on average wait time in line.

Future study into the application of filters to enhance edge detection is shown in preliminary study to be potentially useful. Application of this could be used to focus in on certain portions of the image. With two image capture sources, it would be possible to determine timing for capturing the secondary video source to collect higher resolution images of selected interest such as waiting until the windshield is viewable to capture an image of the vehicle driver. Combined with image registration, it would be possible to detect motion from a video sequence where the camera is moving as well such as using aerial cameras passing overhead.

The research presented here dealt only with grayscale images. Using the color information as well would be useful in further characterizing the traffic through the area of interest by being able to answer queries such as how many red cars have passed the area of interest.

The application of this technique is not limited to vehicular traffic. Any area where motion video can be captured, there is the possibility to analyze the movement of objects through the frame. By delineating multiple areas of interest and more exact definition of the object, it is possible to create moving areas of interest to create motion tracking and characterization to signal other events and perform more useful queries on the video such as when monitoring a cash register for times when the cashier is present without customers present to aid in loss prevention in stores. By analyzing live video in combination with image capture, it is possible to programmatically decide which type of event to save thereby decreasing the amount of storage space required for the archival of critical segments of video.

Planned improvements of the application created for this analysis include extracting individual frames from the video file directly to avoid the overhead of using an external program for pre-processing the video. It is also possible to grab individual frames from a live video source with the video for Linux drivers and under DirectX under Microsoft Windows. The inclusion of different filtering methods will be incorporated to allow further enhancement of the video frames to relieve the effects of object color on the strength of the signal to allow detection of objects closer to the background color in the area of interest. Preliminary results of this have shown promise in use of edge enhancement filters.

REFERENCES

[1] D. Finlayson and M. S. Drew, "White-point Preserving Color Correction," *5th Color Imaging Conference, Nov. 1997.*

[2] R. M. Evans, "Method for Correcting Photographic Color Prints," *US Patent 2,571,697,* 1951.

[3] T. Miyano, "Auto White Adjusting Device," *US Patent 5,659,357,* 1997.

[4] M. Shan and R. Jain (Eds.) *Motion Based Recognition*, Kluwer Academic Publishing, 1997.

[5] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Inc., 1989.

[6] D. S. Taubman and M. W. Marcellin. *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, MA, 2002.

[7] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Compression Standard.* Chapman & Hall, New York, 1993.

[8] T. Acharya and P. S. Tsai. *JPEG200 Standard for Image Compression – Concepts, algorithms and VLSI Architectures*. John Wiley & Sons, Inc., NJ, 2004.

[9] MPEG-1: *"Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s,"* ISO/IEC 11172-2: Video.

[10] MPEG-2: *"Generic Coding of Moving Pictures and Associated Audio Information,"* ISO/IEC 13818-2: Video.

[11] MPEG-4: *"Coding of Audio-Visual Objects,"* ISO/IEC 14496-1: systems.

[12] ITU-T Recommendation H.261: *"Video Codec for Audiovisual Services at px64 kbit/s,"* Geneva 1990.

[13] ITU-T Recommendation H.263: *"Video Codec for Low Bite Rate Communication,"* Geneva 1996.

[14] Mark D. Fairchild, *Color Appearance Models*, Addison Wesley Longman, Inc. Reading Mass.,1998.

[15] The Image Processing Handbook, 2nd Ed. John C. Russ, CRC Press, Boca Raton FL., 1995.

[16] J. D. Murray and W. Vanryper, *Encyclopedia of Graphics File Formats*, O'Reilly and Associates, Inc., 1994.

[17] http://www.transcoding.org/cgi-bin/transcode

[18] http://www.python.org

[19] http://www.pythonware.com/products/pil/

[20] http://www.python.org/moin/TkInter

[21] http://www.gnuplot.info/

[22] http://www.imagemagick.org/

[23] http://www.cs.panam.edu/~dkirtley/video

VITA

David D. Kirtley

724 Jonquil Avenue
McAllen, Texas 78501

Education:

University of Texas – Pan American Teacher Certification, 1989

Texas A&I University, Kingsville, Texas B.S. Major in Geology, 1986

University of Texas – Pan American

Work Experience:

2001 – 2005 Software Systems Specialist, University of Texas - Pan American, Edinburg, Texas

1992 – 2000 Teacher, Rio Grande City C.I.S.D. Rio Grande City, Texas

1992 – 1993 Field Engineer, Anadrill-Schlumberger, Houston, Texas

1989 – 1992 Teacher, Rio Grande City C.I.S.D. Rio Grande City, Texas

Publications:

Compensation for Automatic White Balance Correction with Histogram Equalization, To be published in the proceedings of the *6th International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2005) in conjunction with the 8th Joint Conference on Information Sciences (JCIS 2005).*