Abstract

In this research, data acquired by Unmanned Aerial Vehicles (UAVs) are primarily used to detect and track moving objects which pose a major security threat along the United States southern border. Factors such as camera motion, poor illumination and noise make the detection and tracking of moving objects in surveillance videos a formidable task. The main objective of this research is to provide a less ambiguous image data for object detection and tracking by means of noise reduction, image enhancement, video stabilization, and illumination restoration. The improved data is later utilized to detect and track moving objects in surveillance videos. An optimization based image enhancement scheme was successfully implemented to increase edge information to facilitate object detection. Noise present in the raw video captured by the UAV was efficiently removed using search and match methodology. Undesired motion induced in the video frames was eliminated using block matching technique. Moving objects were detected and tracked by using contour information resulting from the implementation of adaptive background subtraction based detection process. Our simulation results shows the efficiency of these algorithms in processing noisy, un-stabilized raw video sequences which were utilized to detect and track moving objects in the video sequences.

© 2014 The Authors. Published by Elsevier B.V.

Keywords: Enhancement; Optimization; Stabilization; Block Matching; Background Subtraction

1. Introduction

Illegal trespassing and border intrusions by immigrants are a great problem posted against the U.S. border security force and the Department of Homeland Security. Various technical and scientific approaches have been proposed over the past decade to ensure protection to the citizens of U.S. The main stay of these protection processes is video surveillance across the southern border to track movement of illegal immigrants. This video surveillance process is carried out using various methods namely, by mounting cameras on watch towers or using UAVs to constantly patrol the border. The images are captured using various types of image sensors mounted on UAVs. The main area of research is in finding out a way to extract more information from the captured data...
without complicating the hardware involved in the process. In this research, we used a UAV called “Phantom” for capturing test videos in a south Texas university area in order to test our algorithms. Phantom is an aerial filming multi-rotor system with a remote control unit range of 300 meter and a flying speed of 6 mph. Phantom is mounted with a camera on its base. The camera used is GoPro Hero 3 Black edition, which is capable of producing videos with 4K resolution and up to 30 frames per second. In the following sections we shall discuss various video enhancement algorithms and object detection algorithms which have been proposed by various researchers that could be incorporated for the processing of videos captured by the UAV and use it for the purpose of tracking illegal movement across southern border. Image enhancement is the process of converting the obtained raw image data into a more advantageous and convenient form. Image enhancement is also used to emphasize some particular regions of the image which could be used for special purposes like object tracking and detection. Various manipulations on the pixels of the captured image help in suppressing noise and also help in extricating more data from the captured image data. Further these processed and enhanced images are aimed to extract moving object that are of interest. The raw video sequences captured by the UAV are a primary source of information perused by Homeland Security. In order to make the videos more desirable and useful for the purpose of object tracking, the video needs to be made noise-free and clearer. These requisites are accomplished by the following four stages by pixel manipulation of the images in the video sequences.  

2.1. Noise Removal:
Noise present in the video is an inadmissible parameter while processing video data. In order to de-noise the video sequence we use VBM3D algorithm which is based on Box Matching and 3D Filtering (BM3D). This is a de-noising algorithm which has been used to remove specific noises. In BM3D noise removal is achieved by converting similar block in the image into similar image fragments called groups. These blocks are examined by sliding window method which helps in stacking blocks with similar features in a 3D stack called ‘groups’ and the procedure is called ‘grouping’. The similarity between the blocks in the 3D stacks result in high level correlation between the blocks. These blocks are treated with 3D de-correlating transforms which attenuates the noise present in the image. The noise attenuation is attributed to the shrinkage of transform coefficients by wiener filtering and this procedure is termed as ‘collaborative filtering’. We obtain the estimates of these blocks after an inverse of the 3D transform is applied. The estimates of the blocks are combined together to obtained the noise free reconstructed image. BM3D algorithm has been extended to process noisy image sequences in videos and is called Video Block matching 3D filtering (VBM3D). In this technique the standard deviation of the noise has not been determined by calculation before processing, instead it is assumed. In VBM3D, the noisy video captured is represented as

\[ R(x,y;t) = f(x,y;t) + n(x,y;t) \]  

Where \( R(x,y;t) \) is the noisy video signal, \( f(x,y;t) \) is the true signal and \( n(x,y;t) \) is the noise present in the video with standard deviation \( \sigma_n \).

![V-BM3D algorithm](image)

**V-BM3D algorithm**

**Step 1.** Grouping blocks with similar features into stacks and obtaining an estimate using grouping and collaborative filtering

**Step 2.** Final estimates of the 3D blocks and collaborative wiener filtering that uses frequency domain mapping of the corresponding groups from the basic estimate.

2.2. Noise Estimation:
The noise removal process carried out by the VBM3D algorithm does not take into account the exact noise standard deviation of the videos into account. Therefore, in order to overcome this issue we perform Noise Estimation using Immerkaer Method. Here the basic Noise Model is made use of on each image in the video.

\[ I(x,y) = f(x,y) + n(x,y) \]  

where \( I \) is the Noisy image, \( f \) is the Noise-free image and \( n \) is additive noise. The noise is assumed to be additive white noise with zero mean. Variance of the noise is the parameter that is required. It is used along with the VM3D to remove noise from the sequences. According to the Immerkaer Method, a Laplacian operator of the form shown below is used to calculate the standard deviation \( \sigma_n \).

\[ N= \begin{pmatrix} 1 & -2 & 1 \\ 2 & -4 & 2 \\ 1 & -2 & 1 \end{pmatrix} \]
The Standard deviation $\sigma_n$ is computed as

$$\sigma_n = \sqrt{\frac{1}{6(W-2)(H-2)} \sum_{image} |I(x,y)|}$$  \hspace{1cm} (3)

where W is the width and H is the height of the image in pixels.

2.3. Enhancement Using PSO Based Algorithm:

Image enhancement involves mathematical manipulations on pixels, filtering in spatial and frequency domain and also pseudo color transformation to provide more details about the desired region, such as targets. Image enhancement is majorly a process which needs to be optimized for improved efficiency. In the Particle Swarm Optimization (PSO) based algorithm, the global and local characteristics are used to form a parameterized transformation of the captured image and are optimized using the PSO algorithm, which was proposed by Kennedy and Eberhart, as an objective evaluation function $^5$. The transformation function is nothing but a median filter with center as the local information and global information as the correction. This transforming function produces an enhanced image based on the local characteristics such as standard deviation and global characteristics such as mean $^6$.

$$Z(i,j) = K(i,j)[f(i,j) - c \times m(i,j)] + m(i,j)^a$$  \hspace{1cm} (4)

In equation (4), $Z$ is the transformed image, $K$ is the transformation function, $a$ and $c$ are two parameters and $m(i,j)$ is the local mean obtained using a window of $n \times n$

$$K(i,j) = \frac{k \times D}{\sigma(i,j) + b}$$  \hspace{1cm} (5)

In eqn (5), $k$ and $b$ are parameters; $D$ is the global mean and $\sigma(i,j)$ is the local mean of the $(i,j)^{th}$ pixel. $a,b,c,k$ are the parameters which determine the degree of enhancement. These for parameters are set as particle and are iterated to find the best fit using PSO algorithm. Best combination of the $a$, $b,c,k$ which gives maximum enhancement is given by the evaluation function. The range of $a \in [0,1.5], b \in [0,0.5], c \in [0,1], k \in [0.5,1.5]$.

The evaluation function is derived from the mean of edge intensities, entropy of the image, i.e. the contrast range of the image and the number of edges. The edge information is obtained using Sobel operator and numbers of edge pixels are deduced using a histogram of the edge image $^6$.

$$F(I_e) = \log(\log(E(I_e))) \times \frac{n_{edges(i)}}{M \times N} \times H(I_e)$$  \hspace{1cm} (6)

In eqn (6), $I_e$ is the enhanced image, $E(I_e)$ is the sum of the pixel intensities of the sobel edge image edge detector $^6$, $n_{edges(i)}$ is the number of edge pixels in the image and $H(I_e)$ is the entropy i.e. degree of randomness of pixel M and N are the size of the image.

The classical PSO algorithm is used to enhance the images. PSO algorithm helps in optimization by considering each pixel as a random particle with some personality. The PSO algorithm helps in producing an enhanced image which is obtained by assessing the image acquired through transformation functions with the evaluation function. The performance parameters are obtained from the algorithm and the enhancement is achieved by adjusting few parameter settings in the algorithm.

2.4. Illuminance Reflectance Algorithm for Non-Uniform Lighting Conditions: This algorithm $^7$ is used to enhance images captured under non-uniform illumination. A typical example of this kind of situation is an unmanned under water vehicle carrying an imaging device and an illumination source. Due to the light absorption phenomenon deep inside the ocean, even an artificial illumination source cannot produce a uniform lighting condition.

In this method of enhancement, the captured image is considered to be a combination of luminance and reflectance components $^8$.

$$I(x,y) = L(x,y).R(x,y)$$  \hspace{1cm} (7)

In eqn(8), $I(x,y)$ is the true image, $L(x,y)$ is the luminance component and $R(x,y)$ is the reflectance component. The camera characteristics contribute two more parameters which are gained and offsets of the imaging device. As these are typical constant values, the captured image is actually a function of these characteristics and the reflectance component. The non-uniform illumination condition is modelled as an image degraded by a Gaussian function. When the captured image is smoothed using a Gaussian function, the resulting image is a rough estimate of the luminance component $^9$. 


Video stabilization is achieved by adopting the Improved Fourier Mellin Variant method or Circular Block matching. In this
A video stabilization algorithm should nullify the following parameters prevailing in the unstable video:
The process is that the motion of the UAVs capturing the video needs to be estimated to compensate for the motion in the video stream.

The Dynamic range compression of the image is achieved using windowed inverse sigmoid function defined as

$$f(v) = \frac{1}{1 + e^{-av}}$$  \hspace{1cm} (9)

The Luminance component tends to contain low frequency and mid-tone components of reflectance. They are degraded during dynamic range compression. To restore the contrast of original image a center surround type contrast enhancement technique is used. The method involves the below definitions

$$L'_{n,enh}(x, y) = L_{n,enh}(x, y)^{E(x, y)}$$  \hspace{1cm} (10)

In eqn (10), $L'_{n,enh}(x, y)$ is the mid tone frequency enhanced luminance component and $E(x, y)$ is defined as

$$E(x, y) = R(x, y)^p$$  \hspace{1cm} (11)

In eqn (11), $R$ is the reflectance component and $P$ is a parameter determined by global standard deviation of the image. Finally color is restored by dividing the intensity components to original color components.

2.5. Video Stabilization:
Videos captured by UAVs are unstable due to their motion and other external environmental conditions. Tracking of objects in an unstable video is a difficult task which leads to the important process of video stabilization. The challenge here in video stabilization process is that the motion of the UAVs capturing the video needs to be estimated to compensate for the motion in the video stream. The video stabilization requires two components:

- Motion Estimation.
- Image compensation.

A video stabilization algorithm should nullify the following parameters prevailing in the unstable video:

- Horizontal tilt.
- Vertical Tilt.
- Rotation, and
- Scaling.

Video stabilization is achieved by adopting the Improved Fourier Mellin Variant method or Circular Block matching. In this method, unlike other methods such as intensity patterns and the Kanade-Lucas-Tomasi Feature Tracker (KLT) which uses structural features of the image for motion estimation, here spectral information is used. Mosaicking is made use of here, which is based on the variant of the Fourier Mellin Invariant transformation of the image frames. Here logarithmic representation of the spectral magnitude of the iFMI descriptor is used and a filter is applied to accommodate the shift. This algorithm is based on the principal that when two signals are periodically shifted, the inverse Fourier transformation of the phase difference results in a Dirac pulse which is defined by the equation,

$$\delta(t - a) \rightarrow 1. e^{j\omega a}$$  \hspace{1cm} (12)

Algorithm
1. The spectra of two corresponding frames are calculated.
2. Magnitude of the spectral coefficient is calculated.
3. The spectra are resampled to Polar co-ordinates.
4. Logarithmic resampling of the radial axes of polar spectra.
5. The POMF is calculated for the resampled spectra.
6. Rotational/scaling parameters are calculated from Dirac pulses.
2.6. Stabilization Method:

A Sequence of image is obtained. Image frame \( n \) and frame \( n+1 \) are processed with iFMI algorithm. The result is the transformation between frames. From the transformation the camera motion is estimated. The inverse of the transformation between the frame ‘\( n \)’ and ‘\( n+1 \)’ is applied to frame ‘\( n \)’. A PID controller is used to keep the image frame of the output to the center. This part is called compensation.

2.7. Circular Block Matching Algorithm:

A circular block matching algorithm consists of three steps:
- Suitable blocks are chosen from a current frame to match with corresponding blocks of the reference frame.
- A polynomial estimation and prediction model is proposed to determine global motion parameters.
- With the help of the parameters image compensation is done.

2.8. Flow Chart:

![Flow Chart](image)

3. Object Detection and Tracking

Moving object detection is aimed at extracting moving objects that are of interest in video sequences. Moving object detection can be a challenging task, since the computer cannot differentiate certain aspects. Before the video sequence is run in real-time, there are
several obstacles to be considered such as shadows, illumination changes etc. Therefore, a technique called robust video enhancement and stabilization of video sequences is performed in MATLAB©. The robust video stabilization algorithm works in the background of the device which in this case are the UAVs.

The detection of moving objects can also be done using Histogram of Gradient (HOG)\(^{13}\) descriptors. However this method proves to be inefficient and inapplicable for detecting moving objects from UAV videos since it cannot detect objects from a top angle. Hence background subtraction\(^{14}\) algorithm is implemented. Although background subtraction model suffers from various trade-offs such as detection of shadows of moving objects it is most efficient detection technique. Given an image we want to identify the foreground objects. This is done to obtain objects of interest and not the scene. Also this method has less room for error and has very minimalistic cost.

The moving objects are detected using background subtraction algorithm which displays foreground image with contours, foreground mask and mean background image. The design aims to minimize resource usage and estimate background without user intervention. Detecting objects in real-time is implemented and the same is being extended to detect moving objects like cars and human-beings. Once the mentioned operations are performed, several edge detection techniques such as Canny-edge detection\(^{15}\) are applied to find the contours in the videos. Detecting objects in real-time is implemented and the same is being extended to detect moving objects. The background subtraction model is used to differentiate the foreground objects from the background in the sequence of video frames. The foreground and background are identified and hence separated at pixel level using a piece wise median filter\(^{16}\). If we know that many pixels inside a region belong to the foreground, it is more likely to have an object of interest located in that region. It could serve as a strong indicator that the pixels inside the box are foreground pixels. Finally, the object is tracked using the Kalman filter.

The Pseudo code for the detection and tracking of moving objects is given below:

```
• Estimate the background image.
• Implement background subtraction algorithm.
• Filter the object in the foreground using adaptive threshold.
• Perform morphological operations on the group pixels.
• Track the object using Kalman filter.
```

Figure 4. Object detection pseudo code.

The first step is to estimate the background frame and subtract it from the foreground frame. The initial frames of the video sequence are used to estimate the background image. Using a median filtering technique, the object which is in the foreground is differentiated from the background. The consecutive frames are subtracted from the previous frames in the video sequence and a global threshold \(Th\) is applied to the absolute difference to get the foreground mask. Once the objects of interest are obtained, it can be extended to find the contours in the videos. Detecting objects in real-time is implemented and the same is being extended to detect moving objects like cars and human-beings. Morphological operations such as dilate and erode are performed on the group pixels to augment the desired pixels and further calculate a bounding box for the object. Finally, track the vehicle using the Kalman filtering\(^{17}\). The background original image obtained is denoted as \(I_b^{(0)}(x,y)\) and the foreground image subtracted from the background image is denoted as \(I_f^{(0)}(x,y)\). The object detection description is shown below:

\[
FI(x,y) \begin{cases} 
1, & \text{if } I_f^{(0)}(x,y) \cup I_b^{(0)}(x,y) = 1 \\
0, & \text{else}
\end{cases}
\]  

(13)

3.2. Blob Extraction:

We use blob extraction for object pixels that are close to each other. Morphological operations are used to group edge pixels into objects. For example, pixels that represent an object such as a human being or a vehicle are grouped together. The bounding boxes of these objects are calculated. Finally, the individual bounding boxes are merged so that each object is enclosed by a single bounding
box and the corresponding contours are drawn.

### 3.3. Tracking:

Object tracking is the process of detecting the position of each moving object and tracking them in image sequences. The most general object tracking used is the region based object tracking \(^{18}\). In motion estimation the object is tracked between two successive frames by dividing image into small fixed non-overlapping blocks. Each block from first frame will be matched against second frame within a pre-defined search window. The best matching position is detected in order to estimate block displacement. However accuracy of this method is limited by the block size. Therefore when it comes to tracking, the Kalman filter can be chosen for tracking vehicles in a noisy environment. The Kalman filter utilizes the locations of the bounding boxes detected in the previous frames to predict the locations of these bounding boxes in the current frame. It is possible eliminate the shadows of the moving objects by estimating the Gaussian of the shadows of the moving objects and thresholding the shadows of the objects. Tracking of colored objects is also achieved by converting the video sequence into Hue, Saturation and Intensity and then setting an adaptive threshold by adjusting the minimum and maximum values of Hue, Saturation and Intensity. The filtered objects are binarized and the contour of the tracked object is drawn.

### 4. Results

The demonstrations were carried on with the videos captured by Phantom UAV. The algorithm was implemented on a desktop in Windows 7 environment with a 3.2 GHz Intel Pentium 7 processor. The video is captured using Go Pro Black Edition, which is capable of generating video sequences in 4K (4×1920×1280) resolution. Video frames used in this experiment are of size 320×240 and are captured at 30 frames per sec. The noise in the image was estimated by using a Laplacian operator and was found to be 20. The standard deviation value obtained is fed in to the VM3D algorithm to generate a noiseless video sequence.

Figure 5. (a) Noisy image frame. (b) De-noised frame using VBM3D. Figure 6. (c) Poor illumination image. (d) Illumination corrected frame using Illuminance reflectance model.

The de-noised videos are found to have poor illumination and are enhanced by using Illuminance Reflectance Model. The frames after de-noising are subjected to image enhancement using PSO algorithm. Depending on the best fit value of \(a, b, c, k\) the best enhancement is obtained. The quality of enhancement is examined with the fitness function.

<table>
<thead>
<tr>
<th>Image</th>
<th>nedge</th>
<th>edgesum</th>
<th>fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>frame_2</td>
<td>2899</td>
<td>74732</td>
<td>34</td>
</tr>
<tr>
<td>Frame_enh1</td>
<td>5024</td>
<td>77267</td>
<td>66</td>
</tr>
</tbody>
</table>

Table 1: Values of nedge in pixels, edgesum in intensities (0,255), Fitness(0,100) for Enhanced images in Figure 3 (g) for \(a=0.9, b=0.38, c=0.98, k=1.2\) as given by PSO.

The Image quality is assessed using eqn(6), and the fitness value for original and enhanced image is tabulated as shown in the above table. It is found that the enhancement has good effect as the fitness value which is function of entropy; number of edge pixels and sum of intensities has increased from 34 to 66.
5. Conclusion

Raw and noisy video footage of a south Texas university from the UAV was stabilized using iFMI method, de-noised using VBM3D and Enhanced using PSO technique. The processed video was used to track human being and object using back ground subtraction using Gaussian Mixture based background/foreground segmentation algorithm, image thresholding technique HOG. During these HOG descriptors were found to be inefficient in tracking human beings when images are captured exactly above a person i.e. at a top angle. Background subtraction is fast and easy to implement and can be maneuvered even with long term scene changes and repetitive motions in the clutter. An algorithm based on Kalman filtering is used to track multiple objects.

6. References