Adaptive Object Segmentation from Surveillance Video Sequences

Murali S
PET Research Center
P.E.S. College of Engineering
Mandya, Karnataka, India

Girisha R
PET Research Center
P.E.S. College of Engineering
Mandya, Karnataka, India

ABSTRACT

Identifying moving objects from a video sequence is a fundamental and critical task in many computer vision applications. We develop an efficient adaptive object segmentation algorithm for color video surveillance sequences; background is modeled using Multiple Correlation Coefficient ($R_{a.bc}$) using pixel-level based approach. Segmented foreground generally includes self shadows as foreground objects since the shadow intensity differs and gradually changes from the background in a video sequence. Moreover, self shadows are vague in nature and have no clear boundaries. To eliminate such shadows from motion segmented video sequences, we propose an algorithm based on inferential statistical Difference in Mean (Z) method. Self shadow eliminated foreground contains cast shadows. Where, cast shadows produce troublesome effects for video surveillance systems, typically for object tracking from a fixed viewpoint. It yields appearance variations of objects depending on whether they are inside or outside the shadows. To eliminate cast shadows from video sequences, we propose an algorithm based on the fact that, cast shadow points are usually adjacent to object points and are merged in a single blob on the edge of the moving objects. Also cast shadow occurs only at run time (as objects move in the scene). The approach uses the Standard Scores (S) to build statistical model. This statistical modeling can deal with scenes with complex and time varying illumination. S models are constructed and updated for every inputted frame. Results obtained with different indoor and outdoor sequences show the robustness of the approach.

Keywords

Video surveillance, Object segmentation, Motion segmentation, Self shadows, Cast shadows.

1. INTRODUCTION

Analysis of human movement is currently one of the most active research topics in computer vision. Human Motion Analysis (HMA) includes detection, tracking, and recognition of people. HMA can be classified into 3 categories [1, 2], namely low level vision (Detection), intermediate level vision (Tracking) and high level vision (Behavioral Analysis). The application domains where HMA can be applied are video surveillance, content based image retrieval, gait recognition etc.

The automated video surveillance system is expected to detect people and monitor their actions and subsequently need to analyze their behavior in order to prevent any untoward incidents. To analyze the behavior of a person in a given setup, the first step is human detection and tracking. Tracking involves detection of regions of interest in a frame and then finding frame-to-frame correspondence of each region's location and shape.

Nearly, every system in the HMA starts with segmentation [1]; current motion segmentation methods mainly based on background subtraction or temporal differencing or optical flow or statistical methods [2]. Development of a reliable background models adaptive to dynamic changes in complex environments is still a challenge [1]. In this paper, we propose a novel approach to segment the motion objects using statistical $R_{a.bc}$ method. The algorithm obtains stable segmentation results even under varying environmental conditions.

Segmented foreground objects generally includes their self shadows as a foreground object since the shadow intensity differs from the background. To obtain a better segmentation quality, object tracking algorithms must correctly separate foreground objects from the shadows. Since, self shadows produce troublesome effects for video surveillance systems, typically for object tracking from a fixed viewpoint because sometimes they may appear as segmented object. Despite many attempts, the problem remains largely unsolved, due to several inherent challenges: Dark regions are not necessarily self shadow regions since foreground objects can be dark too; a commonly used assumption is that these shadows fall only on the ground plane is not valid to general scenes

One of the main challenges after self shadow elimination is identifying shadows which objects cast [2]. Segmented foreground objects generally includes their cast shadow as a foreground object since cast shadow moves with the foreground object. The inclusion of shadows as foreground points can cause serious problems while extracting moving objects such as object shape distortion, object merging, and even object losses (due to the shadow cast over another object) which also affects surveillance capability while target identification and tracking. To obtain a better segmentation quality, object detection algorithms must correctly separate foreground objects from the shadows [4].

In this paper, we propose an object segmentation method based on Multiple Correlation Coefficient for motion segmentation, Z and S methods for elimination of self and cast shadows'. The proposed object segmentation does not put any restrictions on the scene in terms of illumination conditions, geometry of the objects and size and position of object. The rest of this paper is organized as follows: Section 2 presents a review of the recent and ongoing activity in the domain of object segmentation; sections 3, 4 and 5 discuss object segmentation. Section 6 discusses experimental results; finally, section 7 concludes the proposed methodology.

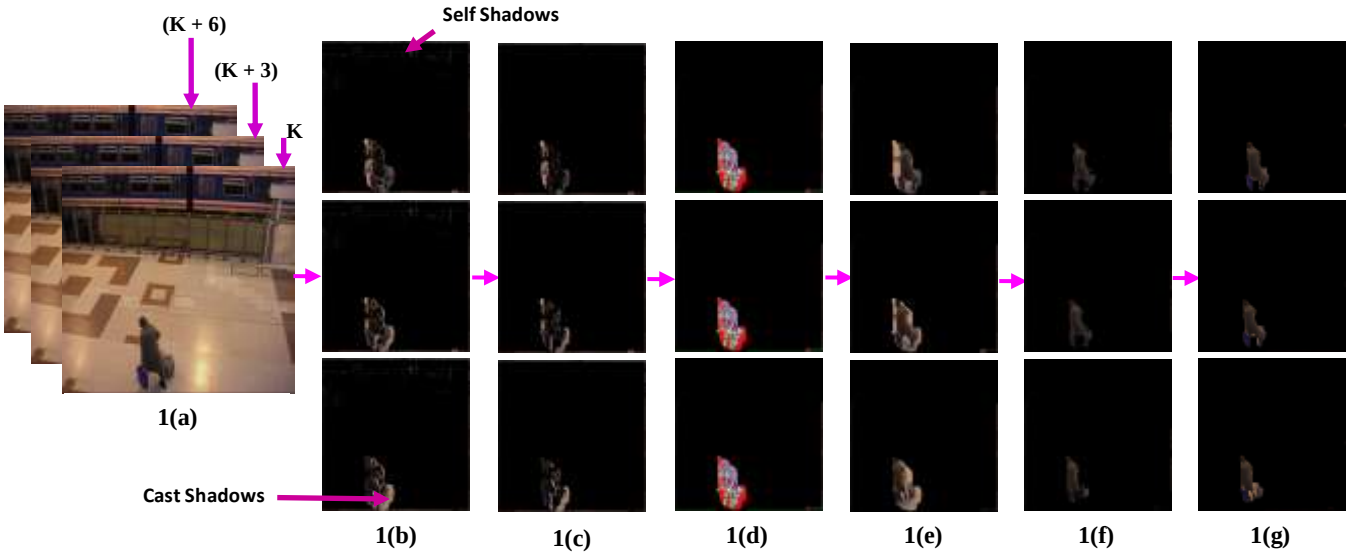


Figure 1: Overview of proposed system to segment objects from surveillance video frames. **(a)** Input frames 109, 112 and 115 of the PETS 2006 data set 7, camera 3. **(b)** Motion segmentation. **(c)** Self shadow elimination. **(d)** Spatial clustering. **(e)** Segmented object. **(f)** Cast shadow elimination. **(g)** Segmented foreground object.

2. RELATED WORK

The first step in HMA is the extraction of motion information through motion segmentation. Motion segmentation in video sequences aims at detecting regions corresponding to moving objects such as humans. Detecting moving regions provides a focus of attention for later processes such as tracking and behavior analysis. At present, all segmentation methods can be classified into four major groups such as background subtraction, temporal differencing, optical flow, and statistical methods.

Background subtraction [3, 6, 7] is a commonly used class of technique to detect moving regions in an image. It is highly dependent on a good background model to reduce the influence of dynamic scenes derived from lighting and extraneous events such as clutter, shadow, occlusion etc. Temporal differencing [8, 9] makes use of pixel-by-pixel difference between two or three consecutive frames in an image sequence to extract moving regions. Temporal differencing is very adaptive to dynamic environment, but generally does a poor job of extracting the entire relevant feature pixels, e.g., generate holes inside the moving entities. Optical flow [10, 11] based motion segmentation uses characteristics of flow to detect independently moving objects even in the presence of camera motion. However most flow methods are computationally complex and very sensitive to noise.

Recently, some statistical methods [12, 13, 14] are proposed to extract change regions from the background and these methods are inspired by the basic background subtraction methods. The statistical approaches use the characteristics of individual pixels or groups of pixels to construct more advanced background models [14]. And the statistics of the background can be updated dynamically during processing. Each pixel in the current image can be classified into foreground or background by comparing the statistics of the current background model. The majority of the statistical methods proposed so far in the literature for background subtraction use either Gaussian or Kernel distribution to model the background [1, 2].

It is very common in real world that the shadow will appear as long as an object is in front of the light source. Shadows occur when objects totally or partially occlude direct light from a light source. According to the classification reported [4] shadows are composed of two parts: self shadows and cast shadows. The self shadow is the part of the object which is not illuminated by the light source. The cast shadow is the area projected on the scene by the object and further classified into umbra and penumbra. The umbra corresponds to the area where the direct light totally blocked by the object, where as in the penumbra area it is partially blocked.

Self shadow detection and elimination algorithms can be classified into model or property based techniques. Model based techniques are usually used for specific situations such as in [19, 20, 21, 22, 23], where a priori knowledge of scene geometry and foreground objects is incorporated into a model. Property based approaches [24], uses features like geometry, brightness or color to identify shadow regions, are more robust to different scene and illumination conditions.

A very few methods for identifying self shadows have been developed in recent years. A comparative study of many self shadow segmentation algorithm can be found in [18]. The proposed method in [19] is based on extraction of dark regions from the image. The algorithm is divided into three stages. In the first stage, dark regions extracted using intensity values and assumes single light source in an indoor environments. In the second stage penumbra regions are identified based on intensity scale from the extracted dark regions and subsequently classified as self or cast shadows in final stage.

The method proposed in [20] uses photometric color invariants to extract shadow regions and subsequently classified as self (if on the object) or cast (if on the ground plane) shadows. Where, in [21] shadowing factor is derived as a function of surface roughness and in color variation; and assumes surface is homogeneous, isotropic and smooth microscopically with a Gaussian height field. Shadow light environment is estimated in

an image [23], using cast and self shadows in a real image. Both self and cast shadows are eliminated from static images in [24]. First self shadows are eliminated using gradient space and then cast shadow edges are extracted using color invariants. Finally, using Poisson equation shadow free reflectance image is obtained.

Figure 2: Frames 487, 488 & 489 of a PETS 2006 data set 5. **First row**: Input frames. **Second row**: Output of $(100\%)R^2_{a.bc}$

Figure 3: Frames 487, 490 & 493 of a PETS 2006 data set 5. **First row**: Input frames. **Second row**: Output of $(100\%)R^2_{a.bc}$

Figure 4: Frames 487, 492 & 497 of a PETS 2006 data set 5. **First row**: Input frames. **Second row**: Output of $(100\%)R^2_{a.bc}$

Presences of shadows are determined first using illumination direction in [22]. Object shapes are recovered using object edges if shadows are present. [22] eliminates the cast shadows from the outdoor images if it is on the ground plane and it keeps the shadows on the object as self shadows based on HSI color space. However, this technique cannot be applied to dynamic environments because method is based on background subtraction; assumes self shadows occur only on the object and nowhere else.

Several methods for identifying cast shadows have been developed in recent years. A few detection algorithms used monocular images as inputs. Studer [4] used features like brightness, edge and shading information to detect moving cast shadows in textured background. The proposed algorithm [4] uses the previous frame (instead of the background) as reference frame. This choice exhibits some limitations in moving region detection since it is influenced by object speed and it is too noise sensitive.

A comparative study of many cast shadow segmentation algorithm can be found in [27]. In [28] pixels are represented in HSV colour space, those pixels are classified as shadows having the approximately the same hue and saturation values compared to the background, but lower luminosity.

Several cast shadow detection algorithms have been proposed for traffic surveillance, which are based on model based shadow detection. Chen et al. [29] combines illumination properties of shadow with lane line geometry for shadow elimination. For shadow elimination, all lane-dividing lines should be first detected, after lane detection shadows are eliminated horizontally first then vertically. The system depends on dynamic environmental condition and camera viewpoint. Lo [30] proposed shadow detection and removal method, considers colour, shading, texture, neighbourhoods and temporal consistency in the scene. Experiments are conducted in known environmental conditions.

Fig. 1, depicts an overall overview of proposed system to segment motion objects from background and to eliminate shadows (Self and Cast) from the segmented motion objects using PETS video of 2006, data set 7 of camera 3 for video frames 109, 112 and 115. The proposed system uses $R_{a.bc}$ to segment motion objects from temporal differencing frames. After motion segmentation, we apply Z method to eliminate self shadows as shown in Fig. 1(c) and then spatial clustering is applied as shown in Fig. 1(d) because temporal differencing generates holes in segmented objects. Cast shadows are eliminated using S values as shown in Fig. 1(f). Once again spatial clustering applied to group motion objects to get final segmented object as shown in Fig. 1(g).

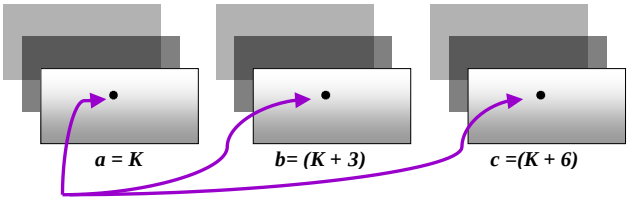
3. MOTION SEGMENATION

Segmentation is an important step in many image processing applications. The idea is to partition an image into a set of regions corresponding to objects in the image based on some feature such as motion or texture. The features used for segmentation may vary continuously between video frames at two different regions. This makes it difficult to draw the line between two regions. It may even be possible that they are in fact so similar that they should be only one region. We have proposed segmentation algorithms based solely on estimations of the motion in image sequences.

A static camera observing a scene is a common case of a surveillance system. Detecting intruding objects is an essential step in analysing the scene. Even though there exist a myriad of segmentation algorithms in the literature [1, 2]. Most of them follow a simple one or two frame differencing except [5, 17] and nearly everyone assume that the background does not vary and hence can be captured a priori. This limits their usefulness in most practical applications.

Motion segmentation is done in this paper, by checking pixel by pixel disparity (using equation (1)) in RGB color space between three (by using equation (5)) video frames simultaneously as shown in Fig. 1(a). Image subtraction is based on temporal differencing (frame gap is three) between K, (K+3) and (K+6) as shown in Fig. 3. Extensive experiments conducted by us on PETS data set revealed that, if we do temporal difference with successive frames as shown in Fig. 2 (i.e., K, (K+1) and (K+2)) motion of the objects is almost negligible and its waste of

processing time. On the other hand, if we increase frame gap beyond three frames than the objects moved very fast in the scene and generated unnecessary cast shadows in the corresponding difference images as shown in Fig. 4. The proposed motion segmentation algorithm in this paper robust to illuminations, complex backgrounds, adapts to dynamic environments and reflections can vary without significantly affecting the result.



p(x,y) and its eight neighbors (N8(p)) referred as pixel P(x,y) in each frame and 3 frames P(x,y) RGB values (in total 27) are used in each $R_{a.bc}$ calculation.

Figure 5: Pixels selection for $R_{a.bc}$ calculation.

Let, the pixel RGB value on any coordinate (x, y) is denoted by $p(x, y)$ with x in the range from 0 to w_x and y in the range 0 to h_y. Where w_x and h_y are the size of the image in the X and Y directions, respectively. Let, a pixel $p(x, y)$ along with its eight neighbors $(N8(p))$ from now on referred to as pixel $P(x, y)$ as shown in Fig. 5. Background is modeled using statistical co-efficient of multiple correlation $(R_{a.bc})$. The distance measuring function (4) is used to find current pixel $P(x, y)$ in all three frames, belongs either to background or foreground. For this, three frames pixel $P(x, y)$ RGB values are represented as $R_{a.bc}$. Threshold (T_s) is applied to co-efficient of determination $(R^2_{a.bc})$ and if $R^2_{a.bc}$ is greater than the T_{3M}, then that pixel $P(x, y)$ is classified as background in all three frames as shown in Fig.1(b).

Multiple correlation, measures the degree of linear relationship between three pixels T_{3M} RGB values from temporal differencing frames K, $(K + 3)$ and $(K + 6)$ as shown in Fig. 6. We assume linear relationship between three pixels RGB values at every position (x, y) in all input video frames.

$$R_{a.bc} = \sqrt{\frac{{r_{ab}}^2 + {r_{ac}}^2 - 2 r_{ab} r_{ac} r_{bc}}{1 - {r_{bc}}^2}} \quad (1)$$

Where, $R_{a.bc}$(where, $a = K, b = (K + 3), c = (K + 6)$) is the co-efficient of multiple correlation between the dependent frame a pixel $P(x, y)$ RGB value by keeping b and c frames pixel $P(x, y)$ RGB values constant,, like this $R_{a.bc}$ is calculated for each $(w_x \times h_y)$ pixel of the frame. Where, r_{ab}, r_{ac} and r_{bc} are correlation coefficient r_{XY} [15] computed using equation (2).

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} \quad (2)$$

$$\text{Where, } S_{XY} = \frac{\sum XY}{N}, S_X = \sqrt{\frac{\sum X^2}{N}} \text{ and } S_Y = \sqrt{\frac{\sum Y^2}{N}} \quad (3)$$

Where, $X = \{a, b\}$, $Y = \{b, c\}$ and $X < Y$. We assume $a < b < c$ based on frame numbers. In equation (3), S_{XY} is a covariance [16], S_X and S_Y are standard deviation of the pixel $P(x, y)$ depending on r_{ab}, r_{ac} and r_{bc}. $N = 27$ total number of RGB values in a pixel.

Let, I^{K_3}, $I^{(K_3+3)}$ and $I^{(K_3+6)}$ are the ${K_3}^{\text{th}}$, $(K_3 + 3)^{\text{th}}$ and $(K_3 + 6)^{\text{th}}$ corresponding frames respectively. Then the difference images $D^{K_3}_{3M}$, $D^{(K_3+3)}_{3M}$ and $D^{(K_3+6)}_{3M}$ are generated using equation (4) which contains motion objects of frames ${K_3}^{\text{th}}$, $(K_3 + 3)^{\text{th}}$ and $(K_3 + 6)^{\text{th}}$ respectively.

$$D^i_{3M_{(x,y)}} = \begin{cases} 0, & if\ R^2_{a.bc} > T_{3M} \\ RGB\ of\ I^i_{(x,y),} & if\ R^2_{a.bc} \le T_{3M} \end{cases} \quad (4)$$

$$\text{Where, } i = \{K, (K + 3), (K + 6)\}$$

$$K = (9n + 1)^{\text{th}} \text{ frame, where } n \ge 0 \text{ and } n \in N \quad (5)$$

Where, T_{3M} is a predefined threshold value empirically chosen. A coefficient of $R_{a.bc}$, lies between 0 and 1 as shown in Fig. 6(b) depending on a linear relationship between three pixels $P(x, y)$ RGB values as shown in Fig. 6(a). If the $R_{a.bc}$ is 1, the correlation is called perfect. Although a correlation coefficient of 0 indicates no linear relationship between the variables, it is possible that a nonlinear relationship may exist.

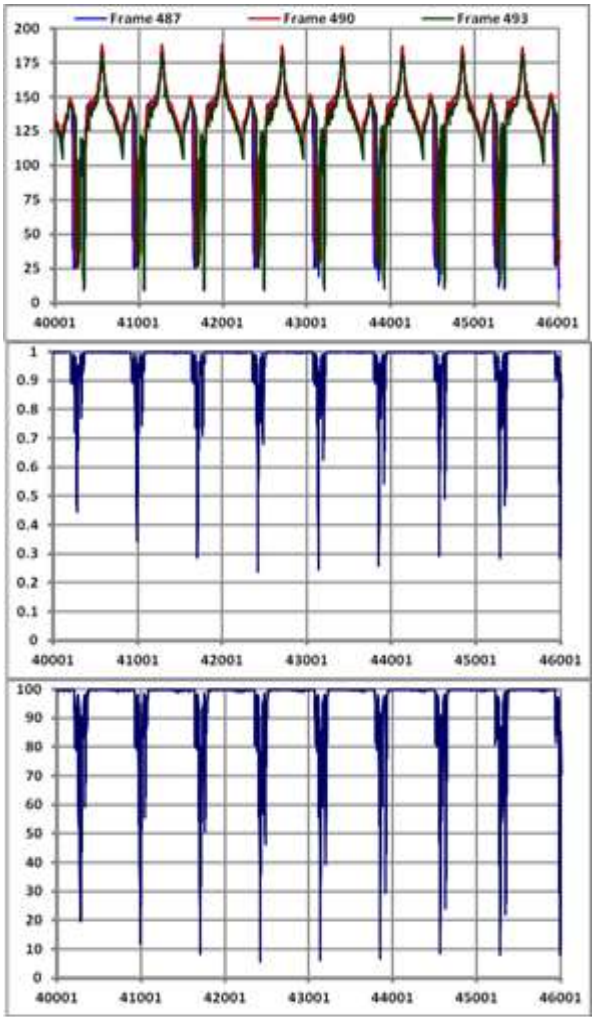


Figure 6: $R_{a.bc}$ applied on frames 487, 490 & 493 of the PETS 2006 Data Set 3, Camera 3. The X-axis indicates pixel positions in a frame. The Y-axis indicates **(a)** μ value, **(b)** $R_{a.bc}$value, **(c)** $(100\%)R^2_{a.bc}$ value.

The $R_{a.bc}$, if interpreted in terms of its squared value (that is, $R^2_{a.bc}$) is an estimate of the proportion of the total variation in pixel $P(x, y)$ RGB values which are explained by the linear relationship between the three values. This proportion is usually

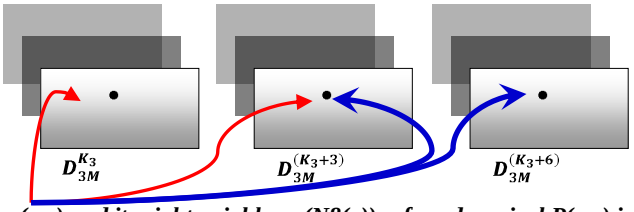
converted to a percentage, $(100\%)R^2_{a.bc}$, which is known as the coefficient of multiple determination as shown in Figure 6(c). For example, if $R_{a.bc} = 0.922$, then $R^2_{a.bc} = 0.850$, which means that 85% of the total variation can be explained by the linear relationship between pixel $P(x,y)$ RGB values and remaining 15% unexplained.

A false positive pixel either belongs to self or cast shadows as shown in Figs. 16 and 23 (row two). Shadows occur when objects totally or partially occlude direct light from a light source. The self shadow pixels are those which are not illuminated by direct light source. If we decrease the T_{3M} value false negative increases and by reducing false positive pixels. The cast shadow will increase, if the objects move fast in the scene because cast shadow points are usually adjacent to object points and are merged in a single blob on the edge of the moving objects [4]. In addition, cast shadow occurs only at run time (as objects move in the scene). However, self shadow remains almost constant because shadow intensity differs from the foreground as shown in Figs. 16 and 23 (row two).

Shadows are omnipresent in real-life setups. Detecting and removing them automatically is crucial for the quality of the segmentation. A point of the scene is shadowed if part of the light it receives in normal circumstances is occluded. The corresponding pixels on the camera images are therefore still representing the same object, but under different lighting conditions. Depending on the type of lighting and the physical properties of the object, the color of the pixels can be modified in a number of ways. Sections 4 and 5, discusses the way to eliminate self and cast shadows from surveillance video sequences using motion segmented frames.

4. SELF SHADOW ELIMINATION

Shadows occur when objects occlude light from a light source. On one hand, shadows provide rich information about object shapes and light orientations. They provide strong clues about the shapes, relative positions, and surface characteristics of the objects. They can indicate the approximate location, intensity, shape, and size of the light source(s). In fact, in some circumstances the shadows constitute the only components of the scene, as in shadow-puppet theater (is an ancient form of storytelling and entertainment using opaque, often articulated figures in front of an illuminated backdrop to create the illusion of moving images) and in pin screen animation (Pin screen animation makes use of a screen filled with movable pins, which can be moved in or out by pressing an object onto the screen. The screen is lit from the side so that the pins cast shadows. The technique has been used to create animated films with a range of textural effects difficult to achieve with traditional animation). On the other hand, shadows may cause embarrassments for visual applications. For example, objects together with their shadows form distorted figures and adjacent objects may be connected through shadows. Both can confuse object recognition systems. Segmenting objects from shadows can be a nontrivial task. Referring to Fig. 1, shadows can be broadly divided as cast and self shadows. As revealed in that figure, the self shadow is a part of the object, which is not illuminated by the light source. The cast shadow lying beside the object belongs to the background. For object recognition and many other applications, cast shadows are undesired and need to be eliminated, if self shadows are not part of objects they should be eliminated. If objects have intensities similar to those of shadows, shadow removal could become extremely difficult. Even though objects and shadows can be separated, object shapes are often incomplete.



p(x,y) and its eight neighbors (N8(p)) referred as pixel P(x,y) in each frame and at a given time 2 frames P(x,y) RGB (in total 18 from 9 pixels) values are used in each Z calculation.

Figure 7: Pixels selection for Z calculation.

Self shadows are modeled based on Z_{ML} after foreground pixel extraction. Z_{ML} value is computed between frames (($D_{3M}^{K_3}$ and $D_{3M}^{(K_3+3)}$) and ($D_{3M}^{(K_3+3)}$ and $D_{3M}^{(K_3+6)}$)), among corresponding RGB values of the pixels $P(x,y)$ using equation (6). Finally, the average of the two computed Z value is taken from equation (9) to decide if the current sample pixel $P(x,y)$ belongs to self shadow or to motion object as represented in equation (7). Those parts of the segmented motion objects, which are not illuminated by light source, become self shadows such parts are also eliminated by self shadow removal algorithm as shown in Fig. 9.

The two main theoretical branches of statistical science are descriptive and inferential statistics. The former is useful to characterize the overall set of data, called population, by assigning a proper descriptive model or distribution family to it. The latter one, adapted when the entire set of data is unknown and we want to infer the behavior of the entire population from a sub-set of sample data [12, 13, 14].

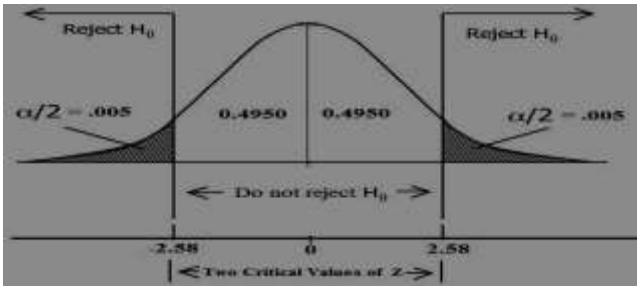


Figure 8: The rejection and nonrejection regions for hypothesis testing about (μ_1 - μ_2) for PETS 2006, Data set 7, Camera 3 video sequence.

We use inferential statistics difference in mean (Z) test in this paper, for comparing means of two independent populations. The Z uses two components null hypothesis (H_0) and alternative hypothesis (H_1) to test a claim (i.e., two sample means are equal or not). A hypothesis is a claim or statement about a property of population [25]. Where, the H_0 is a claim about a population parameter that is assumed true until it is declared false. Where, H_1 is claim about a population parameter that will be true if the H_0 is false. The Z test in this paper uses a predetermined significance level, denoted by α to test a claim (i.e., probability of rejecting H_0). A confidence interval is the range of values that we believe to be part of the H_0 population (i.e., that would lead us to retain the H_0) is constructed using critical values in such a way that the

probability of rejecting the H_0, if it is true, is equal to α [26]. Critical values (are selected from standard normal distribution table) separates the critical region (where we reject the H_0) from the values of the Z test statistics that do not lead to a rejection of the H_0 as shown in Fig. 8.

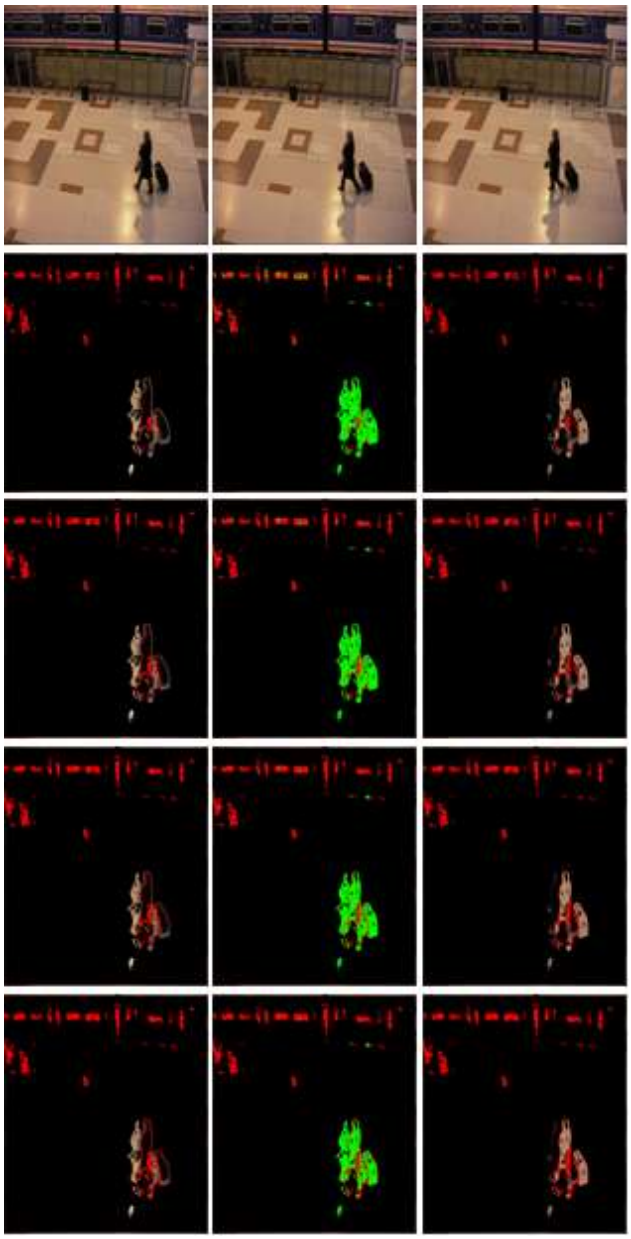


Figure 9: Confidence level analysis for Z test using frames of the PETS 2006 Data Set 7, Camera 3. **Red** pixels will be classified as SELF SHADOWS using Z method. **First row:** Input frames (2107, 2110 & 2113), **Second row** = 80%, **Third row**= 95%, **Fourth row**= 99%, **Fifth row** = 99.9%.

In this section, we consider statistical inferences of Z test, to eliminate self shadows. Let, μ_M be the mean of the first population and μ_L be the mean of the second population. To test a hypothesis about the difference between these two population means i.e. $(\mu_M - \mu_L)$ we calculate $(\bar{X}_M - \bar{X}_L)$ to make an interval estimate and to test a hypothesis. Where, $\bar{X}_M$ be the mean of a sample taken from the first population and $\bar{X}_L$ be the mean of a sample taken from the second population [25]. Considering following two possibilities H_0 and H_1, based on independent random samples of size $n_M = 27$ and $n_L = 27$ of the two temporal frames as shown in Fig. 7. Therefore, the sampling distribution of $(\bar{X}_M - \bar{X}_L)$ is large and approximately normal, and we use normal distribution to perform the hypothesis test [26].

Let, $H_0 : \mu_M - \mu_L = 0$ (Belongs to Self Shadow)

Let, $H_1 : \mu_M - \mu_L \neq 0$ (Belongs to Motion Object)

$$Z_{ML} = \frac{(\bar{X}_M - \bar{X}_L) - (\mu_M - \mu_L)}{\sqrt{\frac{S_M^2}{n_M} + \frac{S_L^2}{n_L}}} \quad (6)$$

Where, Z_{ML} is the difference in means test of the pixel $P(x,y)$ between two temporal differencing frames RGB values. The Z_{ML} is calculated for each $(w_x \times h_y)$ remaining foreground pixels of the motion segmented frames. Where, $M = \left\{D_{3M}^{K_3}, D_{3M}^{(K_3+3)}\right\}$, $L = \{(M+3)\}$ that is if $M = D_{3M}^{K_3}$ then $L = D_{3M}^{(K_3+3)}$. $n_M = n_L = 27$ are number of RGB values of the pixel P(x,y). Where, the value of $(\mu_M - \mu_L) = 0$ substituted from H_0. Where, S_M and S_N are the standard deviations of the two samples selected from the $(\bar{X}_M - \bar{X}_L)$

Let, $D_{3M}^{K_3}$, $D_{3M}^{(K_3+3)}$ and $D_{3M}^{(K_3+6)}$are motion segmented frames of K_3^{th} , $(K_3+3)^{\text{th}}$ and $(K_3+6)^{\text{th}}$ respectively (after motion segmentation using $R_{a.bc}$ method) using equation (4). Then images $D_{SD}^{K_3}$, $D_{SD}^{(K_3+3)}$ and $D_{SD}^{(K_3+6)}$ are generated using equation (7) which contains self shadow eliminated motion objects of frames $D_{3M}^{K_3}$, $D_{3M}^{(K_3+3)}$ and $D_{3M}^{(K_3+6)}$ respectively.

$$D_{SA_{(x,y)}}^{i} = \begin{cases} 0, & if\ (|Z| \leq T_{SD}) \\ RGB\ of\ D_{3M_{(x,y)}}^{i}, & if\ (|Z| > T_{SD}) \end{cases} \quad (7)$$

$$\text{Where, } i = \left\{D_{3M}^{K_3}, D_{3M}^{(K_3+3)}, D_{3M}^{(K_3+6)}\right\} \quad (8)$$

$$Z = \frac{Z_{D_{3M}^{K_3}, D_{3M}^{(K_3+3)}} + Z_{D_{3M}^{(K_3+3)}, D_{3M}^{(K_3+6)}}}{2} \quad (9)$$

Where, Z is an average value taken from $Z_{D_{3M}^{K_3}, D_{3M}^{(K_3+3)}}$ and $Z_{D_{3M}^{(K_3+3)}, D_{3M}^{(K_3+6)}}$ (both values are computed using Z_{ML}) and T_{SD} is an critical value empirically chosen from standard normal curve table [14]. The significance level (α) is 0.01. The $\neq$ sign in the H_1 indicates that the test is two-tailed. A two-tailed test has rejection regions in both tails. The area in each tail of normal distribution curve will be $\alpha/2 = 0.01/2 = 0.005$. The critical values of the Z for 0.005 area in each tail of the normal distribution curve are $\cong \pm 2.58$ from standard normal distribution table.

Fig. 8, shows Z confidence interval for PETS 2006, data set 7, camera 3 video sequence and Fig. 9 gives confidence level analysis. Critical value range ($T_{SD} = \pm 2.58$) is compared with calculated Z value at 99 confidence level and if Z lay between T_{SD} value range, then H_0 is accepted and pixel $P(x,y)$is classified as self shadow in all three frames as shown in Fig. 10.

5. CAST SHADOW ELIMINATION

Visual information, in the form of images and video, comes from the interaction of light with objects. Illumination is a fundamental element of visual information. Detecting and interpreting illumination effects are part of our everyday life visual experience. Shading for instance allows us to perceive the 3D nature of objects. Shadows are particularly salient cues for inferring depth information. However, we do not make any efforts to avoid them. Moreover, when humans are asked to describe a picture, they generally omit the presence of illumination effects, such as shadows, shading and highlights. The human visual system is able to analyze illumination in a scene and to discard it to reach a description of the scene's content that is, more useful for action. It is also able to analyze illumination effects to get information about the scene. Millions of years of biological evolution and environmental adaptation have indeed made human vision a highly developed and complex process.

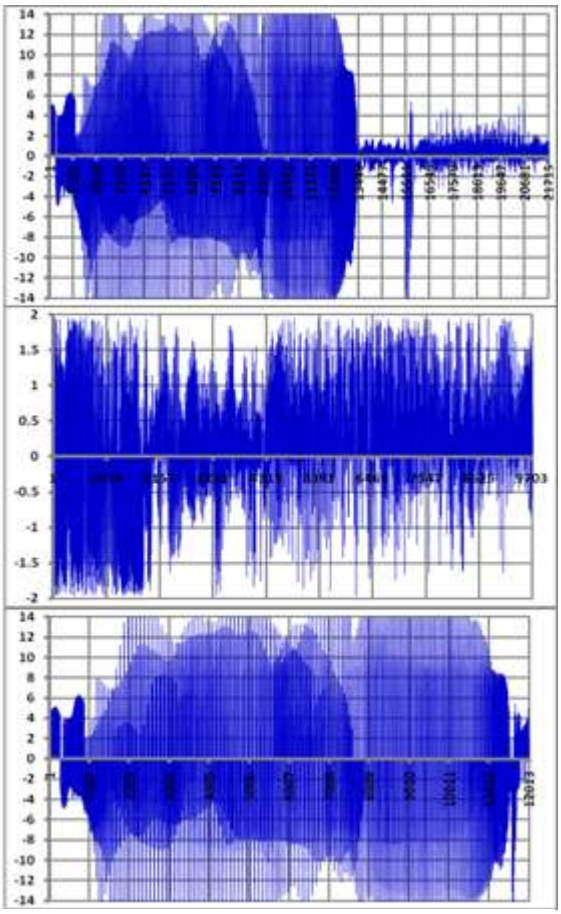


Figure 10: Z calculation for the frames 2107, 2110 and 2113. The X-axis shows number of remaining pixels after motion segmentation using $R_{a.bc}$ method. The Y-axis indicates **(a)** Average Z value, **(b)** Recognized self shadow pixels, **(c)** Recognized possible motion object pixels.

For many algorithms in computer vision, dealing with illumination effects is a challenging task. Illumination phenomena can in fact mislead fundamental tasks such as object extraction and description. For this reason, lighting conditions require careful consideration in many applications and need often to be controlled. Among illumination effects, shadows are often an integral part of natural scenes. A growing interest has emerged over the last few years within the computer vision community in the investigation of the nature of shadows in digital images.

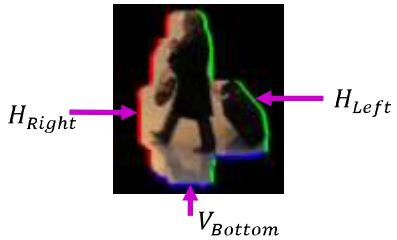


Figure 11. Possible cast shadow pixels extraction from three directions, Red pixels from $\boldsymbol{H_{Right}}$, Blue pixels from $\boldsymbol{T_{Bottom}}$ and Green pixel from $\boldsymbol{T_{Left}}$

Object moving between a light source and the background as shown in Fig. 1 generates a cast shadow on the background. In addition, the features connected with cast shadow, such as illumination, its geometry, color and the position changes with time, while the background is stable as shown in Fig. 1. For such situations, a temporal feature is a fundamental element to handle the evolution of the cast shadows. Moreover, a detected foreground object, will probably continue being in the foreground for some time. Spatial feature is another essential element to understand the structure of the cast shadows. Spatial feature, such as objects edge color helps in detecting any changes in cast shadow.

The first step toward identifying cast shadows is to extract regions as possible cast shadows. The algorithm identifies possible cast shadow candidates by scanning video frames in three directions concurrently, where cast shadows are likely to occur in normal surveillance videos. That is, in horizontal direction from right to left (H_{Right}), then from left to right (H_{Left}) and finally in vertical direction from bottom to top (V_{Bottom}) as shown in Figs. 11 and 12. For each direction, the algorithm calculates thresholds *(T_{Right}, T_{Left} and T_{Bottom})* based on S using RGB values of the possible cast shadow pixels, which will be used to identify cast shadows from the segmented objects.

Figure 12: Frames (103 & 106) of a video Sequence of PETS 2006 data set 7, camera 3, which shows possible, cast shadow points identified by the algorithm.

Let, $D_{\mathrm{Z}}^{K_3}$, $D_{\mathrm{Z}}^{(K_3+3)}$ and $D_{\mathrm{Z}}^{(K+6)}$ are $D_{SA}^{K_3\,\mathrm{th}}$, $D_{SA}^{(K_3+3)\,\mathrm{th}}$ and $D_{SA}^{(K_3+6)\,\mathrm{th}}$ corresponding frames respectively. Let the RGB value on any

pixel (x,y) be denoted by $I(x,y)$ with x in the range from 0 to n_x and y in the range 0 to n_y. Where n_x and n_y are the size of the image in the X and Y directions, respectively. A frame is scanned in the X direction at y=Y_0, the pixel with the RGB value $I(x,y)>0$, for $0 \leq x < n_x$, is stored in the vector R. Subsequently, x and y values are varied between n_x and n_y respectively to get the remaining edge values of the segmented objects in H_{Right} direction and vector R will contain first encountered pixel RGB values (the red pixels). The vector R values then represented has S, which becomes a threshold value T_{Right} for the objects H_{Right} direction.

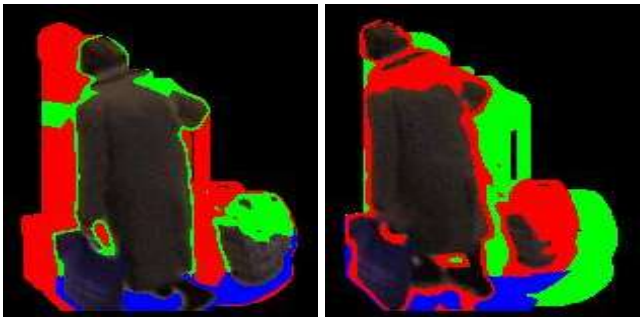


Figure 13: Frames (103 & 106) of a video Sequence of PETS 2006 data set 7 camera 3 after cast shadow points recognized by the algorithm. Red pixels from $\boldsymbol{H_{Right}}$; Blue pixels from $\boldsymbol{V_{Bottom}}$ and Green pixel from $\boldsymbol{H_{Left}}$

Figure 14: Frames (103 & 106) of a video Sequence of PETS 2006 data set 7, camera 3. **First Row**: Frames after cast shadow points are removed by the algorithm using equation (10) and (11). **Second Row**: Frames after spatial clustering.

After, threshold calculation, using equations (10) and (11), we classify foreground pixel either to a motion object or cast shadow. For that, a frame again scanned in all three directions concurrently. Wherever, two successive pixels RGB value greater than 0, S is calculated using equation (11). And then current pixel $I(x,y)$ classified either into cast shadow or motion object using equation (10). In summary, we will treat pixel $I(x,y)$ as cast shadow if it satisfies equations (10) and (11) as shown in Figs. 13 and 14.The Z score [13] is the value of standard deviations the data value falls above (positive Z score) or below (negative Z score) the mean for the data set. Z score specifies an exact location within a normal distribution and is a dimensionless value.

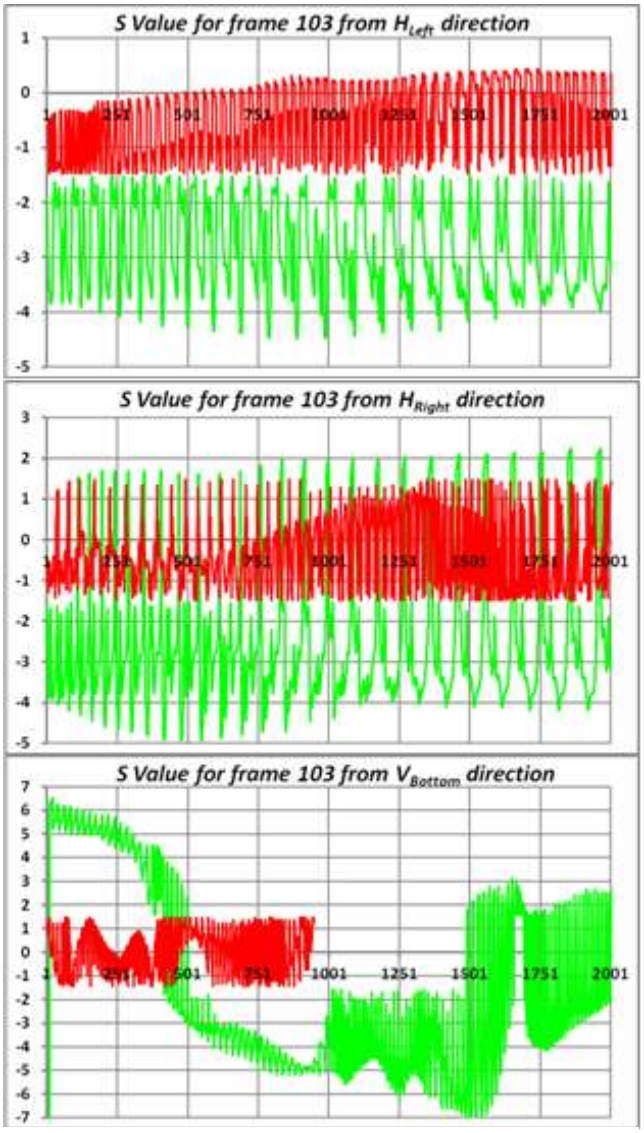


Figure 15: For the frame 103 of the PETS 2006 data set 7. The X-axis shows foreground pixels position and Y-axis shows S value. The **RED** color indicates pixels that are recognized as CAST SHADOWS and **Green** color indicates pixels that are recognized as FOREGROUND.

$$D^i_{Z(x,y)} = \begin{cases} 0, & if\ (|S| \leq T_{SS}) \\ RGB\ of\ D^i_{Z(x,y)}, & if\ (|S| > T_{SS}) \end{cases} \quad (10)$$

Where, $i = \left\{D^{K_3}_{SA}, D^{(K_3+3)}_{SA}, D^{(K_3+6)}_{SA}\right\}$

$$S = \left(\frac{\left(\Sigma^{(x+1,y+1)}_{(x,y)} \mu^i{}_S\right) - \mu^i{}_P}{\sigma^i_P}\right) \quad (11)$$

Let, $D_{\mathrm{Z}}^{K_3}$, $D_{\mathrm{Z}}^{(K_3+3)}$ and $D_{\mathrm{Z}}^{(K+6)}$ are $D_{SA}^{K_3\,\mathrm{th}}$, $D_{SA}^{(K_3+3)\,\mathrm{th}}$ and $D_{SA}^{(K_3+6)\,\mathrm{th}}$ corresponding frames respectively from equation (7), which contains cast shadow eliminated objects after applying equation (10). Where, μ_S^i is a sample mean calculated based on two successive pixel RGB values belongs to H_{Right} or H_{Left} or V_{Bottom} direction. μ_P^i (μ_{Right}^i , μ_{Left}^i or μ_{Bottom}^i) and σ_P^i (σ_{Right}^i , σ_{Left}^i or σ_{Bottom}^i) are population mean and standard deviation calculated values with respect to H_{Right} or H_{Left} or V_{Bottom} directions. T_{SS} (T_{Right}, T_{Left} or T_{Bottom}) is a predefined threshold value empirically chosen for each direction. While calculating Z in the equation (11) at any given point of time μ_S^i, μ_P^i, σ_P^i and T_{SS} must be from same direction. The threshold value T_{SS} should be small to retain as many pixels of true foreground objects but must discriminate cast shadow pixels from foreground pixels as shown in Fig. 15. Fig. 14 shows frames after cast shadow elimination. As a post-processing stage, we apply spatial clustering for remaining foreground pixels as shown in Fig. 14.

6. EXPERIMENTAL RESULTS

we analyzed and evaluated the performance of object segmentation algorithm for surveillance video sequence frames of IEEE PETS[1] (Performance Evaluation of Tracking and Surveillance) 2001, 2004, 2006 and 2009 data sets. System has been tested using several sequences of PETS data set among which there are different tracking scenario including indoor and outdoor environments, varied number of people. Results shown here are raw results, without any post treatment. For each environment, parameters were set once. Results we have selected represent a snapshot of the algorithm results and are typical of the performance throughout the sequences. Each of the sequence contains 500 to 4000 frames and resolution varied from one sequence to another sequence.

In outdoor environments, illumination changes rapidly due to fast changing weather conditions. Figs. 16, 17, 22 and 23 show frames of an outdoor video sequence in which whole image illuminated by direct sun light.

There are always variations in the illumination parameters between two frames of the same scene taken even at different times of the same day. Figs. 18 to 21 show images in indoor environment, corresponding to color video sequences acquired in varying range of fluorescent lighting systems with complex illumination. Because of the multiple light sources on the ceiling and the high reflectivity of the floor, shadows cast on the background by objects have a large variation in intensity. At a given pixel, the shadows go from being fairly light to being fairly deep as a function of the position of the object.

In Figs. 16 to 23, first row shows input frames; second row shows motion segmented output frames, which contains self shadows; third row shows frames completely free from self shadows; fourth row contains cast shadow eliminated frames.

7. CONCLUSION

In this paper, we have proposed a system capable of segmenting moving objects from surveillance videos. We employed a novel three stage method which uses Multiple correlation coefficient to segment three video frames simultaneously with each other in temporal differencing method in the first stage. In the second stage self shadow elimination method is proposed which considers sample pixel from segmented motion objects while calculating Z, to eliminate self shadows. Finally, cast shadows are eliminated using S. The proposed shadow removal techniques are applied to foreground rather than the entire image so as to save significant processing time. This is important for real time applications such as surveillance systems.

Extensive experiment conducted on different data sets of PETS (to name a few: Outdoor- 10694 frames, Indoor – 65,000 frames) reveals that results are stable and satisfactory. The object segmentation algorithm is robust to large variation in intensity (such as those caused by fluorescent lighting and as well as direct sun light), due to temporal differencing method. Therefore, we conclude that the proposed algorithm works well under various conditions.

8. ACKNOWLEDGMENT

Authors would like to thank DRDO (*Aeronautical Research & Development Board).* This publication is based on the project sponsored by DRDO file number DARO/08/2021515/M/I

9. REFRENCES

1. W.Hu et al, "A Survey on Visual Surveillance of Object Motion and Behaviors", IEEE Transactions on SMC - part C: Applications and reviews, Vol. 34, No.3.
2. Thomas B. Moeslund et al,"A Survey of Advances in Vision Based Human Motion Capture and Analysis", Computer Vision and Image Understanding, October 2006.
3. Joshua Migdal and W.E.L. Grimson,"Background Subtraction Using Markov Thresholds", Proceedings of the IEEE Workshop on Motion and Video Computing, 2005.
4. J. Stauder, R. Mech and J. Ostermann. "Detection of Moving Cast Shadows for Object Segmentation". IEEE Transactions on Multimedia, 1(1):65-76,1999.
5. Kameda, Yoshinari et al, "A human motion estimation method using three successive video frames", proceedings of the IC on virtual system and multimedia. 1996.
6. S.Wachter and H.H. Nagel, "Tracking Persons in Monocular Image Sequences", Computer Vision and Image Understanding, Vol. 74, No.3, pp. 174-192, June, 1999.
7. Chia-Jung Pai et al,"Pedestrian Detection and Tracking at Crossroads", Pattern Recognition, 2004.
8. N.Thome et al, "A Robust Appearance Model for Tracking Human Motions", IEEE DICTA-2005.
9. L.Havasi et al, "Higher Order Symmetry for Non-linear Classification of Human Walk Detection", Pattern Recognition Letters, Vol. 27, pp. 822-829, 2006.
10. S.Denman et al, (2005), "Adaptive Optical Flow for Person Tracking", IEEE DICTA-2005.
11. Daniel Freedman and M.W.Turek,"Illumination-Invariant Tracking via Graph Cuts", IEEE CVPR, June 2005.
12. Steven Cheng, et al,"A MultiScale Parametric Background Model for Stationary Foreground Object Detection", IEEE Workshop on Motion and Video Computing, 2007.

[1]**Performance data can be found at "http://homepages.inf.ed.ac.uk/rbf/CAVIAR/"**

13. Mohand Said Allili, et al,"A Robust Video Foreground Segmentation by Using Generalized Gaussian Mixture Modeling", IEEE Fourth Candian Conference on Computer and Robot Vision, 2007.
14. Jaime Gallego, et al,"Segmentation and tracking of static and moving objects in video surveillance scenarious", IEEE ICIP, 2008.
15. Stephen Bernstein and Ruth Bernstein, "Elements of Statistics II: Inferential Statistics". Schaum's Outlines, First edition, New Delhi, 2005.
16. Murray R. Spiegel and Larry J. Stephens, "Statistics". Schaum's Outlines, Third edition, New Delhi, 2000.
17. Collins et al, "A system for video surveillance and monitoring", tech. report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May, 2000.
18. Andrew Woo, Pierre Poulin and Alain Fournier, "A survey of shadow algorithms", IEEE CG&A, Volume 10, Issue 6, Nov. 1990 Page(s):13 - 32 November 1990.
19. Caixia Jiang and M.O. Ward, "Shadow identification", Proceedings of the IEEE Computer Society Conference on CVPR, 15-18 June 1992 Page(s):606 – 612.
20. Elena Salvador, Andrea Cavallaro and Touradj Ebrahimi, "Shadow identification and classification using color models", Proceedings of the IEEE IC on Acoustics, Speech, and Signal Processing, Volume 3, 7-11 May 2001 Page(s):1545 - 1548
21. Yinlong Sun, "Self shadow and local illumination of randomly rough surfaces", Proceedings of the 2004 IEEE Computer Society conference on Computer Vision and Pattern Recognition, 2004.
22. Wang. J.M et.al., "Shadow detection and removal for traffic images", Proceedings of the IEEE international Conference on Networking, Sensing & Control, 2004.
23. Takeshi Takai, A Maki and T Matsuyama, "Self shadows and cast shadows in estimating illumination distribution", , 4th European Conference on Visual Media Production 27-28 Nov. 2007 Page(s):1 - 10
24. Li Xu, Feihu Qi and Renjie Jiang, "Shadow removal from a single image", Proceedings of the IEEE international Conference on Intelligent Systems Design and applications, 2006.
25. Douglas C. Montgomery and Geroge C. Runger, "Applied Statistics and probability for engineers", Third edition, John wiley & Sons, 2003.Stephen B and Ruth B, "Elements of Statistics II: Inferential Statistics", Schaum's Outlines, Tata McGraw-Hill Edition, 2005.
26. Prem S.Mann, "Introductory Statistics", Fifth edition, Wiley India Edition, 2007.
27. Prati, I. Mikic, M.M. Trivedi and R. Cucchira. "Detecting Moving Shadows: Algorithms and evaluation". IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.25. No.7, July 2003.
28. Prati, I. Mikic, M.M. Trivedi and R. Cucchira. "Detecting Moving Objects, Ghots and Shadows in video streams". IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.25. No.10, October 2003.
29. Jun-Wei Hsieh, Shih-Hao Yu, Yung-Sheng Chen, and Wen-Fong Hu. "A Shadow Elimination Method for Vehicle Analysis". Proceedings of the 17th IEEE International Conference on Pattern Recognition, 2004.
30. Kuo-Hua Lo and Mau-Tsuen Yang, "Shadow Detection by Integrating Multiple Features", Proceedings of the IEEE 18th International Conference on Pattern Recognition, 2006.

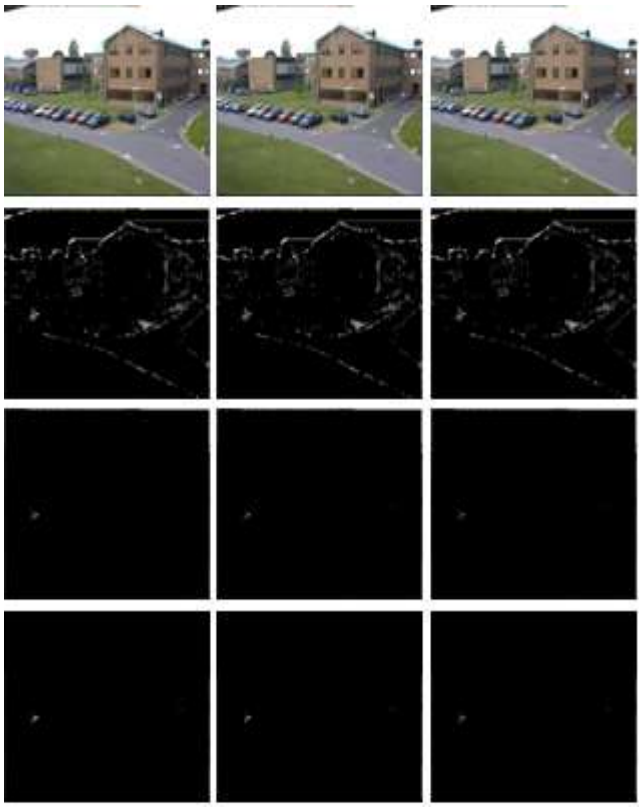


Figure 16: Frames of PETS data set 2001.

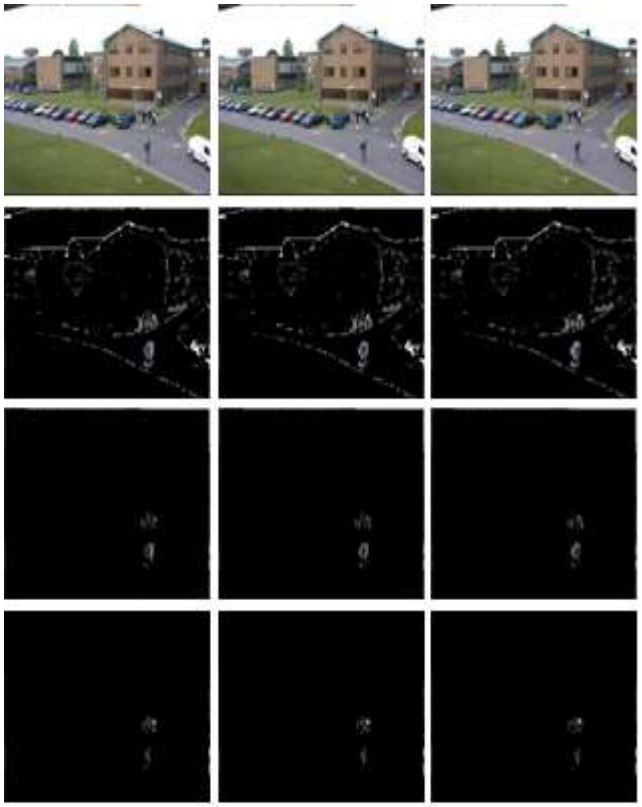


Figure 17: Frames of PETS data set 2001.

Figure 18: Frames of PETS data set 2004.

Figure 19: Frames of PETS data set 2004.

Figure 20: Frames of PETS data set 2006.

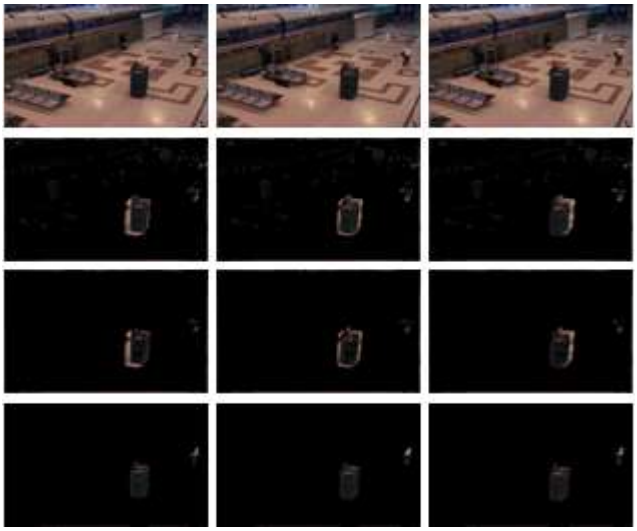


Figure 21: Frames of PETS data set 2006.

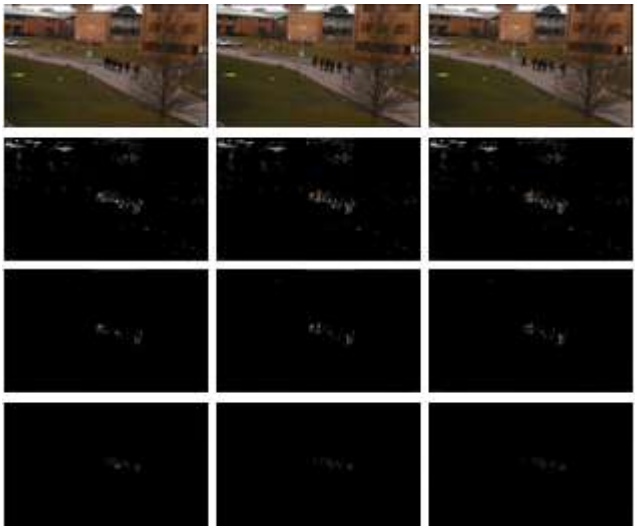


Figure 22: Frames of PETS data set 2009.

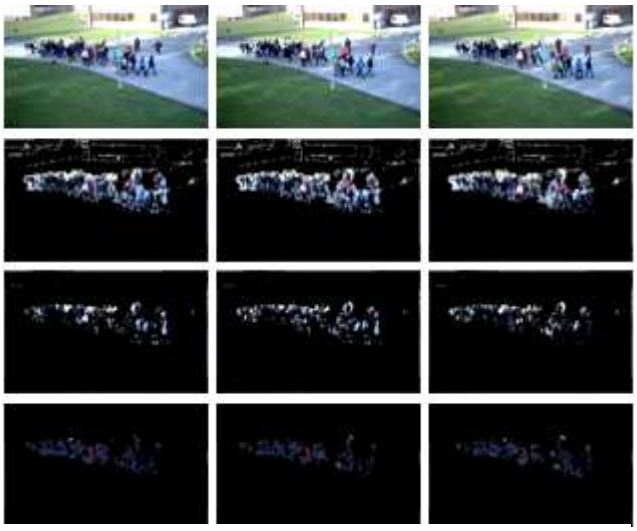


Figure 23: Frames of PETS data set 2009.