International Journal of Automation and Computing 11(1), February 2014, 59-71
DOI: 10.1007/s11633-014-0766-9

Smart Objects Identification System for Robotic Surveillance

Amir Akramin Shafie Azhar Bin Mohd Ibrahim Muhammad Mahbubur Rashid
Department of Mechatronics Engineering, Faculty of Engineering, International Islamic University Malaysia, 50728 Kuala Lumpur, Malaysia

Abstract: Video surveillance is an active research topic in computer vision. In this paper, humans and cars identification technique suitable for real time video surveillance systems is presented. The technique we proposed includes background subtraction, foreground segmentation, shadow removal, feature extraction and classification. The feature extraction of the extracted foreground objects is done via a new set of affine moment invariants based on statistics method and these were used to identify human or car. When the partial occlusion occurs, although features of full body cannot be extracted, our proposed technique extracts the features of head shoulder. Our proposed technique can identify human by extracting the human head-shoulder up to 60%–70% occlusion. Thus, it has a better classification to solve the issue of the loss of property arising from human occluded easily in practical applications. The whole system works at approximately 16−29 fps and thus it is suitable for real-time applications. The accuracy for our proposed technique in identifying human is very good, which is 98.33%, while for cars′ identification, the accuracy is also good, which is 94.41%. The overall accuracy for our proposed technique in identifying human and car is at 98.04%. The experiment results show that this method is effective and has strong robustness.

Keywords: Humans and cars identification, partially occluded human, affine moment invariants, video surveillance systems, machine vision.

1 Introduction

Video surveillance is conventionally displayed on one or several video monitors and recorded. The person who observes the continuous video is tasked to find out if there is activity that demands a response. Though this conventional surveillance performed manually is proficient for crime prevention, it requires many human resources and is costly[1]. On the other hand, intelligent video surveillance uses software to automatically identify specific objects and behaviors in real time systems. The analysis of behaviors of people and vehicles also needs identification system. Hence, humans and cars identification in videos has become an increasingly important research area in both computer vision and pattern recognition communities because of its potential applications in video surveillance system. In recent years, the development of human detection and tracking system has been going forward for several years; many real time systems have been developed[2]. However, there are still challenging technologies that need more researches: foreground segmentation, human shape description and human tracking in occluded scenes.

Reference [3] suggested that detecting humans has been proven to be a challenging task because of the wide variability in appearance due to clothing, partial occlusion and illumination conditions. To realize a more robust and secure video surveillance system, an automated humans and cars identification system is needed which can identify human and car even when the partial occlusion occurs and can analyze video streams in real-time by utilizing fast-computation techniques without compromising the accuracy and performance of that particular surveillance system. Hence, in this paper, a real-time humans and cars identification system is presented, which utilizes combinations of several not computationally expensive techniques that are proven to be robust enough against noise, scaling and partially occluded human.

We proposed the use of foreground segmentation based on adaptive background subtraction to extract foreground objects from the image. Then, to get accurate detection, we apply shadow detection and removal in our technique based on simple contrast adjustments in the HSV (hue, saturation, value) color space. Next, the extracted foreground objects will be identified as human full body, human head-shoulder or car. It is difficult to describe human only with one model, because human is non-rigid and human movement is complex. For example, human shape while standing is different from the human shape while moving[4]. Hence, we proposed the use of affine moment invariants[5] based on statistics method which was introduced in [6], but with minor modifications for feature extraction of the extracted foreground objects.

The main contributions of this work include: a new set of affine moment invariants based on statistics method which is suitable for classification purpose; humans and cars identification using aspect ratio and the new set of affine moment invariant features and extraction of human head-shoulder using both vertical and horizontal projective histograms whenever extracted objects is outside the human ratio limit.

The rest of this paper is organized as follows. Section 2 reviews related works in detection systems. Objects detection is presented in Section 3. Then, Section 4 presents the humans and cars identification process. Next, Section 5 elaborates the results. Finally in Section 6 we conclude the paper.

Manuscript received October 16, 2012; revised April 25, 2013

2 Related works

Detection of objects in video streams is the first relevant step of information extraction in many computer vision applications including video surveillance. Object detection consists of background modeling and object segmentation. The background modeling is the step where the image is considered as background in which it contains the non-moving objects in a video.

Approaches to recognize pixels of foreground objects can be categorized into three main techniques, namely background subtraction[7, 8], temporal differencing[9−11] and optical flow[12−14]. For example, [7] proposed a background subtraction algorithm to separate foreground objects from the background by compensating the input frames and comparing with the given reference frame. They also replaced the widely used morphology operations by adopting a spatial filter to suppress various noises due to the swaying of the tree branches. Reference [9] detected the moving vehicles using temporal differencing method, also known as frame differencing method. Reference [12] proposed the framework that detects the changes in the video scenes from optical flow and encodes using hidden Markov model to classify abnormal events.

When objects are successfully detected in the video surveillance system, the next step in video surveillance system is to successfully classify the moving objects. To further track objects and analyze their behaviors, it is essential to correctly classify moving objects. With the help of object type information, more specific and accurate methods can be developed to recognize higher level actions of video objects. Typical video scenes may contain a variety of objects such as people, vehicles, animals, natural phenomena (e.g., rain, snow), plants and clutter. However, the main target of interest in surveillance applications is generally humans and vehicles. Furthermore, it is necessary to provide details of a moving object such as object type to the user for recognizing what is happening at an outdoor spot. Object classification could differentiate region of interest from those by moving clouds, swaying trees, or other dynamic occurrences.

In general, for objects classification in video surveillance, there are shape-based, motion-based, and feature-based classification methods. Shape based classification using the geometry of the extracted regions such as boxes, silhouettes, blobs containing motion to classify objects in video surveillance. Reference [15] presented the shape-based classification using the simple shape ratio of human body model. We found that using the ratio of human will not give the desired results when partial occlusion occurs.

Meanwhile, for motion based classification, the idea here is to use motion characteristics and patterns to distinguish between objects. Reference [16] presented a method using residual flow to analyze rigidity and periodicity of moving objects. Non rigid moving object such as a human being has a higher average residual flow and this can be a useful cue to distinguish between human motion and other moving objects such as vehicles.

Skin color has an important feature that can be used for the classification of humans in video. Reference [17] proposed a method to classify human model based on skin color and foreground information which is foreground pixels and skin color probability maps. However, the skin color is difficult to be extracted when the objects are far from the camera and it will give the false results when the colors of skin and other objects are similar. Our proposed method, adopting shape based method using the moment invariants and aspect ratio as the features, is proven to be reliable for humans and cars identification even when partial occlusion occurs.

3 Objects detection

Fig. 1 shows overall process for objects detection in our proposed technique.

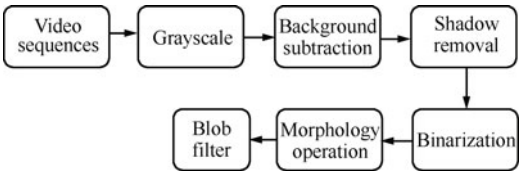


Fig. 1 Overall process for objects detection in our proposed technique

3.1 Background modeling and foreground segmentation

The background modeling is the step where the image is considered as a background which contains the non-moving objects in a video. First of all, we initialize the background of the images, which can be acquired from the background image from the video sequence, and then, we update the background image because of the changes in the scene. There are several changes that can occur in the video and can lead to the misinterpretation of the changes of scene as the background image. Hence, we adopt the simple adaptive background model as in [18] with minor modification. Here, first of all, we save the luminance value of all pixels belonging to k-th frame which does not contain any interested foreground objects (human and car). Then, after n consecutive frames, the luminance value of all pixels belonging to $(k+n)$-th frame is compared with the luminance value in k-th frame. If the value remains unchanged, the current frame will be adopted as a new background image as shown in Fig. 2. We choose 30 as the value for n.

Foreground segmentation is the next step after the background image is achieved. In order to detect the moving object, the pixels belonging to foreground objects need to be distinguished from the background. There are three main approaches in order to recognize foreground objects′ pixels, which are background subtraction, temporal differencing and optical flow. Although, temporal differencing is very suitable for dynamic environments and it does not require background initiation, it has limitation in extracting all relevant feature pixels. Meanwhile, optical flow is computationally expensive[19]. Thus we use background subtraction to recognize the foreground pixels. This can be done by simple technique of subtracting the observed frame from the background image and thresholding the result to detect the objects of interest.

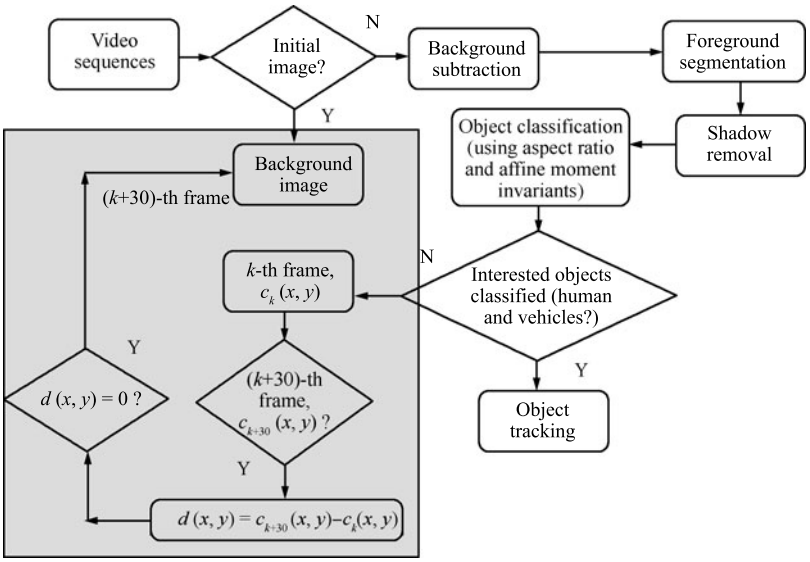


Fig. 2 Overall process in our surveillance system (background modeling showed in shaded area)

3.2 Shadow removal and morphological operations

Shadow is one of the environmental factors influencing processing of monitored images. An object is often accompanied by shadow. The shadow affects the performance of foreground detection and the regions of shadow will be detected as foreground. In order to get accurate detection of moving objects, it is necessary to separate the objects from its shadow. Reference [20] proposed a method based on shadow confidence score (SCS) to separate vehicles and cast shadows from the foreground objects in traffic monitoring system. We adopt the technique for shadow removal based on properties of color information like chromaticity and luminance as proposed in [15, 21], which are proven to be less computationally expensive.

The color distribution of shadow possesses two properties. One is that the chromaticity and luminance are lower in the shadow area. Another one is that there is a higher density in the lower chromaticity of the shadow. From the two properties of the shadow as aforementioned, we can know that luminance of the shadow area is lower. Hence, we use the global contrast adjustment method[22], in which we control the value of luminance by changing the contrast of the output image from the background subtraction to remove the lower luminance shadows while keeping the desired foreground object with higher luminance unaffected[15].

The contrast value is carefully selected to compensate other parts of extracted foreground objects which might have low luminance, because a higher contrast value can remove more shadow but can heavily deform the shapes of the foreground objects. Here, the contrast value is selected based on the experiments conducted. The results for shadow removal (Fig. 3) showed that the global contrast adjustment method for shadow removal is reliable for indoor and outdoor scenes. The indoor image in the third row in Fig. 3 was adopted from context aware vision using image-based active recognition (CAVIAR) dataset[23].

Fig. 3 Results for shadow removal demonstrated in binary images. (a) Original image; (b) Without shadow removal; (c) With shadow removal

Next, the image containing the foreground objects is transformed into a binary image which is a black and white image using global binarization method as in [24−27]. Then, the binary image is morphologically reconstructed using the closing operator. Normally the binary images have discrete noises and holes in the object region, which can be removed using morphology processing[28]. In our proposed technique, we use the closing operator which is a combination of dilation and erosion. The process of closing employs dilation, followed by erosion. Using the closing technique, noise and small holes will be eliminated while keeping the size of foreground objects.

Liu et al.[29] presented an ontology matching approach that employs mapping technique to perform similarity iter-

ative computation and discovers both linguistic and structural similarity. Liu et al.[30] proposed an image subtraction algorithm to effectively extract the target fluorescence signal. The next section describes our matching and identification technique to classify humans and cars.

4 Humans and cars identification process

Then, the blob filter is applied to extracting all the foreground objects. We use simple aspect ratio as in (1) to calculate the ratio of the extracted foreground objects. This method can be used to pre-distinguish human, vehicles and other objects. The human configuration is more than one, but the whole shape of human is in stabilization. From the CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences[31], we found out that full body human ratio is between the range of 1.8 to 3.7. Meanwhile, for vehicles (car), we use the database from [32] and we found out that side view of car ratio is between 0.26 to 0.7.

$$\text{AspectRatio} = \frac{\text{Height(pixels)}}{\text{Width(pixels)}}. \tag{1}$$

However, when the partial occlusion occurs, e.g., a human is occluded a table, then the full human body cannot be extracted. Hence, we come out with solutions by computing peak intensity value of vertical histogram for extracted blob if the ratio is not in the range of human full body as shown in Fig. 4. Based on the experiments, we found out that human in the surveillance system has the size of width less than 80 pixels and height less than 140 pixels, while a side view of the car has the size of width more than 80 pixels and height more than 25 pixels as shown in Table 1. Hence, we come out with solutions by extracting the head-shoulder of the human if the ratio is not in the range of human full body and if the peak intensity value of vertical histogram for extracted blob is less than 12 000. We choose to extract human head-shoulder because it is stable basically and it is difficult to be occluded, except that if the human is fully occluded by other humans.

Table 1 Peak intensity value of vertical histogram for human and car

Types of objects	Peak intensity value of vertical histogram	Size of pixels
Human	<12 000	Width less than 80 pixels and height less than 140 pixels
Car	>16 000	Width more than 80 pixels and height more than 25 pixels

We extract the human head-shoulder using horizontal histogram as in our previous paper[33], and the methodology for human head-shoulder extraction in our proposed technique is also shown in the shaded area in Fig. 6. We found out that human head-shoulder ratio is between 0.5 to 1.2. However, if the peak intensity value of vertical histogram for extracted blob is more than 16 000 and if the ratio is between 0.26 to 0.7, we presume that the extracted blob is a car. Since the contour of the object reflects the shape of the object itself, we use the contours of human full body, head-shoulder and car in our proposed technique as shown in Fig. 5. Then, we extract the feature of the extracted contour by using affine moment invariants using statistical method as it is invariant to translation, scaling or rotation. Here, we use moment invariants and aspect ratio to recognize the extracted objects as human full body, human head-shoulder or car. The overall process for our identification system is shown in Fig. 6.

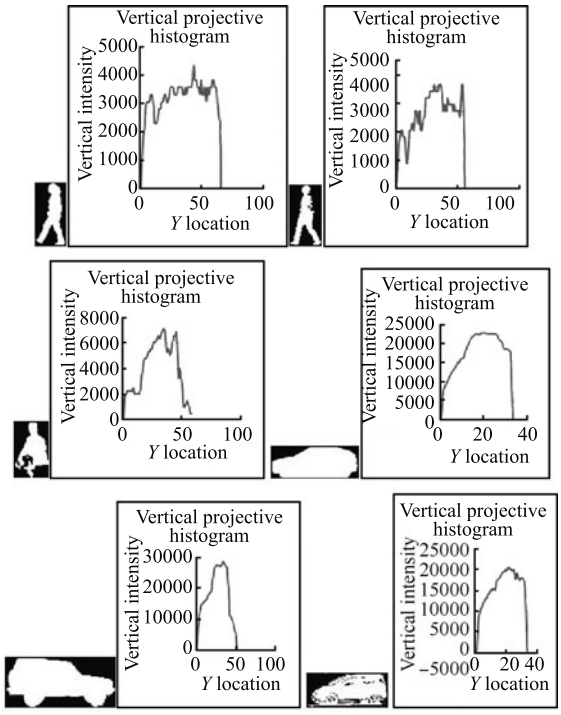


Fig. 4 Extracted target and its respective vertical projective histogram

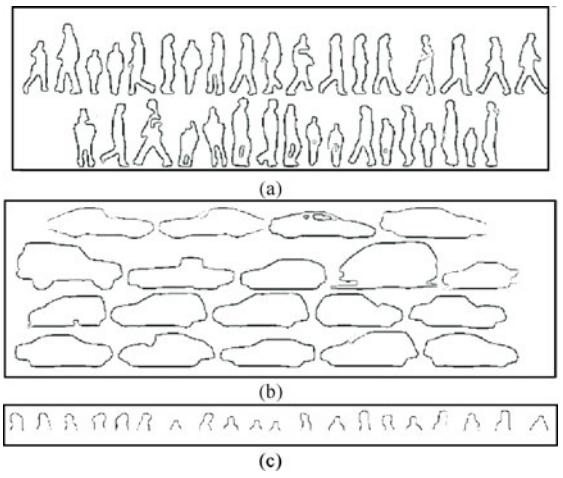


Fig. 5 Examples of training samples. (a) Human full body; (b) Vehicles (car); (c) Extracted human head shoulder

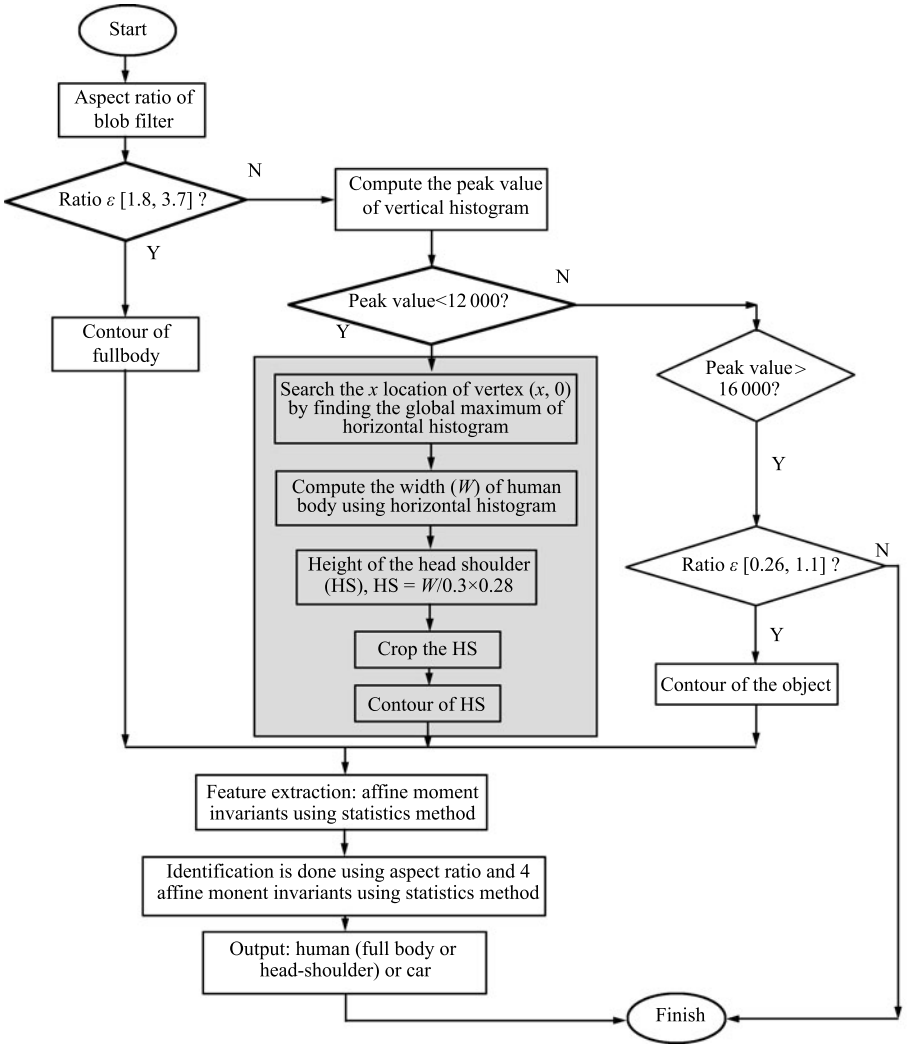


Fig. 6 Identification process in our system

4.1 Feature extraction

Moment invariants have been frequently used as features for image processing, remote sensing, shape recognition and classification. Image moments are useful to describe objects after segmentation. Moments can provide characteristics of an object that uniquely represent its shape. Several techniques have been developed that derive invariant features from moments for object recognition and representation. These techniques are distinguished by their moment definition, such as the type of data exploited and the method for deriving invariant values from the image moments.

The use of moments as invariant binary shape representations was first proposed by Hu in 1962[34]. Hu successfully used this technique to classify handwritten characters by using his seven moment invariant features. In particular, Hu[34] defined seven values as in (7)−(13), computed by normalizing central moments (6) through order three that are invariant to scaling, translation, or rotation. In 1993, Flusser and Suk[5] introduced the use of affine moment invariants derived by means of the theory of algebraic invariants. These features of moments are invariant under general affine transformations and can be used for recognition of affine-deformed objects[5].

Usually, moment invariants extraction uses the whole target region. According to [4], this extraction method requires large computation, while using the image contours requires less computation. In this paper, we extract the affine moment invariants features for every extracted blob that are pre-classified as human full body, human head-shoulder or car.

The general moment of a shape in a K by L binary image is defined as

$$n_{pq} = \sum_{x-0}^{x-X-t} \sum_{y-0}^{x-L-t} (x)^q f(x,y), \quad p,q = 0,1,2\cdots \tag{2}$$

where $f(x,y)$ is the intensity of the pixel (either 1 or 0) at the coordinates (x,y) of the contour, and $p+q$ is said to be the order of the moment. Relative moments are then calculated using the equation for central moments which are

defined as

$$\bar{x} = \frac{m_{10}}{m_{00}} \tag{3}$$

$$\bar{y} = \frac{m_{01}}{m_{00}} \tag{4}$$

$$\mu_{pq} = \sum_{x}\sum_{y}(x-\bar{x})p \cdot (y-\bar{y})^{q} f(x,y). \tag{5}$$

When a scaling normalization is applied the central moments change to

$$n_p q = \frac{\mu_{pq}}{\mu_{00}^{r}}, \quad \gamma = \left[\frac{(p+q)}{2}\right] + 1. \tag{6}$$

In particular, Hu[34] defined seven values, computed by normalizing central moments through order three, which are invariant to scaling, translation, or rotation. In terms of the central moments, the seven moments are given as

$$I_1 = (n_{20} + n_{01}) \tag{7}$$

$$I_2 = (n_{20} - n_{02})^2 + 4n_{11}^2 \tag{8}$$

$$I_3 = (n_{30} - 3n_{12})^2 + (3n_{21} - n_{03})^2 \tag{9}$$

$$I_4 = (n_{30} + n_{12})^2 + (n_{21} + n_{03})^2 \tag{10}$$

$$I_5 = (n_{30} - 3n_{12})(a_{30} + n_{12})[(n_{30} + n_{12})^2 - 3(n_{21} + n_{03})^2] + (3n_{21} - n_{03})(n_{21} + n_{03})[3(n_{30} + n_{21})^2 - (n_{21} + n_{03})^2] \tag{11}$$

$$I_6 = (n_{20} - n_{02})[(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] + 4n_{11}(n_{30} + n_{12})(n_{21} + n_{03}) \tag{12}$$

$$I_7 = (3n_{21} - n_{03})(n_{30} + n_{12})[(n_{30} + n_{12})^2 - 3(n_{21} + n_{03})^2] - (n_{30} + 3n_{12})(n_{21} + n_{03})[3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] \tag{13}$$

Invariance to translation is achieved using central moments. However, since we are using blob filter, we do not face the problems in translational variability. Hence, we adopt affine moment invariants[5] using general moments (2) instead of using central moments as follows and the reader may refer to [5] for derivations.

$$M_1 = (m_{20}m_{02} - m_{11}^2)/m_{00}^4 \tag{14}$$

$$M_2 = (m_{30}^2 m_{03}^2 - 6m_{30}m_{21}m_{12}m_{03} + 4m_{30}m_{12}^3 + 4m_{21}^3 m_{03} - 3m_{21}^2 m_{12}^2)/m_{00}^{10} \tag{15}$$

$$M_3 = (m_{20}(m_{21}m_{03} - m_{12}^2) - m_{11}(m_{30}m_{03} - m_{21}m_{12}) + m_{02}(m_{30}m_{12} - m_{21}^2)/m_{00}^7 \tag{16}$$

$$M_4 = (m_{20}^3 m_{03}^2 - 6m_{20}^2 m_{11}m_{12}m_{03} - 6m_{20}^2 m_{11}m_{12}m_{03} + 9m_{20}^2 m_{02}m_{12}^2 + 12m_{20}m_{11}^2 m_{21}m_{03} + 6m_{20}m_{11}m_{02}m_{30}m_{03} - 18m_{20}m_{11}m_{02}m_{21}m_{12} - 8m_{11}^3 m_{30}m_{03} - 6m_{20}m_{02}^2 m_{30}m_{12} + 9m_{20}m_{02}^2 m_{21}^2 + 12m_{11}^2 m_{02}m_{30}m_{12} - 6m_{11}m_{02}^2 m_{30}m_{21} + m_{02}^3 m_{30}^2)/m_{00}^{11}. \tag{17}$$

The results that we get using affine moment invariants (14−17) and Hu's 7 moment invariants (7−13) are shown in Table 2. It has clearly shown that affine moment invariants give better results for the classification purpose. However, invariants (M_1, M_2, M_3 and M_4) are not in the same magnitude. The inconsistent magnitude of the invariant features is not good for classification purpose, because it will increase the problem of large-range features domination small-range features in classification. To solve this problem, [6] normalized the values of every invariant by a statistic method as

$$M_i^{'} = M\frac{\mu_4 - \theta}{\sigma_4}, \quad i = 1, 2, 3, 4. \tag{18}$$

where M_1 is the mean of M_i, σ_i is the standard deviation of M_i, and $M = 1$.

Table 2 Feature extraction results using Hu's 7 moment invariants (7)−(13) and affine moment invariants (14)−(17)

Types	Results						
	Hu's 7 moment invariants (range)						
	$I_1(10^{-3})$	$I_2(10^{-4})$	$I_3(10^{-9})$	$I_4(10^{-9})$	$I_5(10^{-16})$	$I_6(10^{-10})$	$I_7(10^{-17})$
Human full body	[189, 310]	[80, 700]	[0, 70]	[0, 170]	[0, 270]	[0, 450]	[0, 600]
	Classical affine moment invariants (range)						
	$M_1(10^{-4})$		$M_2(10^{-6})$		$M_5(10^{-6})$		$M_4(10^{-6})$
	[450, 472]		[340, 396]		[528, 560]		[227, 257]
	Hu's 7 moment invariants (range)						
	$I_1(10^{-3})$	$I_2(10^{-4})$	$I_3(10^{-9})$	$I_4(10^{-9})$	$I_5(10^{-16})$	$I_6(10^{-10})$	$I_7(10^{-17})$
Human head-shoulder	[160, 199]	[1, 120]	[0, 1 230]	[0, 1 000]	[0, 7 500]	[0, 750]	[0, 80 000]
	Classical affine moment invariants (range)						
	$M_1(10^{-4})$		$M_2(10^{-6})$		$M_5(10^{-6})$		$M_4(10^{-6})$
	[430, 456]		[296, 355]		[500, 538]		[190, 219]
	Hu's 7 moment invariants (range)						
	$I_1(10^{-3})$	$I_2(10^{-4})$	$I_3(10^{-9})$	$I_4(10^{-9})$	$I_5(10^{-16})$	$I_6(10^{-10})$	$I_7(10^{-17})$
Car	[235, 330]	[190, 810]	[0, 40]	[0, 40]	[0, 15]	[0, 85]	[0, 30]
	Classical affine moment invariants (range)						
	$M_1(10^{-4})$		$M_2(10^{-6})$		$M_5(10^{-6})$		$M_4(10^{-6})$
	[449, 474]		[388, 404]		[554, 563]		[255, 267]

The statistics $\bar{M}_i$ and σ_i are calculated from the invariant values of the templates in Figs. 5 (a), (b) and (c). Reference [6] used this statistics method for their new affine moment invariant polynomials which were generated automatically. However, invariants in [6] include more than third-order of invariants which need more time, and this is not suitable for real-time surveillance system as in this paper. Hence, we used this statistics method (18) for classical affine moment invariants which are $M_i(M_1, M_2, M_3$ and $M_4)$ as shown in Table 3.

Table 3 Feature extraction results using statistics method of affine moment invariants as in [6]

Types	Statistics method of affine moment invariants using average value (range)			
Human full body	$M_1'(10^{-4})$ [−550, 900]	$M_2'(10^{-6})$ [−550, 900]	$M_3'(10^{-6})$ [−610, 900]	$M_4'(10^{-6})$ [−600, 900]
Human head-shoulder	$M_1'(10^{-4})$ [−2 500, −600]	$M_2'(10^{-6})$ [−2 500, −600]	$M_3'(10^{-6})$ [−2 700, 620]	$M_4'(10^{-6})$ [−2 500, −650]
Car	$M_1'(10^{-4})$ [−920, 1 300]	$M_2'(10^{-6})$ [930, 1 400]	$M_3'(10^{-6})$ [880, 1 350]	$M_4'(10^{-6})$ [900, 1 350]

However, using the statistics method in (18) gives a large range for each type of objects as shown in Table 3. So, we come out with solutions by using minimum value of M_i instead of using average (mean) value as shown in (19). The results using our method are given in Table 4. From Table 4, we observe that our method can reduce the range of values for each type of objects and is better for classification than using classical affine moment invariants. The results for statistics from the invariant values of the templates in Figs. 5 (a), (b) and (c) are given in Table 5.

Table 4 Feature extraction results using our method

Types	Statistics method of affine moment invariants using average value (range)			
Human full body	$T_1(10^{-6})$ [195, 340]	$T_2(10^{-8})$ [−190, 340]	$T_3(10^{-8})$ [195, 350]	$T_4(10^{-8})$ [180, 330]
Human head-shoulder	$T_1(10^{-6})$ [0, 190]	$T_2(10^{-5})$ [0, 185]	$T_3(10^{-8})$ [0, 206]	$T_4(10^{-8})$ [0, 170]
Car	$T_1(10^{-6})$ [340, 385]	$T_2(10^{-8})$ [335, 385]	$T_3(10^{-8})$ [340, 390]	$T_4(10^{-8})$ [325, 375]

Table 5 Results for statistics from the invariant values of the templates in Figs. 5 (a), (b) and (c)

$\bar{M}_1$	$\bar{M}_2$	$\bar{M}_3$	$\bar{M}_4$
459.96	366.13	542.49	239.62
$\min(M_1)$	$\min(M_2)$	$\min(M_3)$	$\min(M_4)$
431.25	296.7	500.05	191.17
σ_1	σ_2	σ_3	σ_4
11.56	28.25	16.18	20.34

$$T_i = \left(\frac{M_i - \min(N_t)}{\sigma_1} \right), \quad i = 1, 2, 3, 4 \tag{19}$$

where $\min(M_i)$ is the minimum value of M_i, while σ_i is the standard deviation of M_i.

4.2 Identification process

We could not identify objects by using moment invariants only due to the overlapping that occurs between the values of human full body and human head-shoulder and between the values of human full body and car. Therefore, we identify the objects using both aspect ratios as well as the moment invariants. Since the values in Table 4 are experimental based, we enlarge the values for moment invariants for testing purpose in the real scene as given in Table 6. Finally, identification in our system is done using both aspect ratios as well as the moment invariants as shown in Table 6.

5 Experimental results

We evaluate the proposed framework in two aspects: 1) human in indoor and outdoor and 2) car in outdoor. The proposed method uses single core programming with C# programming language and it was implemented on a laptop (Intel Core i5-2410M, 2.30 GHz with 2 GB of memory). The whole system works at approximately 16−29 fps.

For human, we use two public datasets and recorded videos in International Islamic University Malaysia (IIUM) to evaluate the algorithmic effectiveness. Meanwhile, for car, we use other recorded videos in IIUM. The public datasets that we used are adopted from CAVIAR Test Set[23] and SPEVI Datasets[35]. The SPEVI datasets have a frame rate of 25 fps and the frame size of 360×288 pixels, while CAVIAR datasets have a frame rate of 25 fps and the frame size of 384×288 pixels. The recorded videos in IIUM have a frame rate of 30 fps and the frame size of 320×240

Table 6 Final values for testing

Types	Aspect ratio	Statistics method of affine moment invariants using average value			
Human full body	1.8−3.7	$T_1(10^{-6})$ [170, 360]	$T_2(10^{-8})$ [180, 370]	$T_3(10^{-8})$ [180, 370]	$T_4(10^{-8})$ [160, 360]
Human head-shoulder	0.5−1.2	$T_1(10^{-6})$ [−70, 240]	$T_2(10^{-8})$ [−70, 250]	$T_3(10^{-8})$ [−70, 250]	$T_4(10^{-8})$ [−70, 250]
Car	0.26−0.7	$T_1(10^{-6})$ [310, 420]	$T_2(10^{-8})$ [310, 430]	$T_3(10^{-8})$ [310, 430]	$T_4(10^{-8})$ [300, 420]

pixels. The results of the proposed framework are categorized into two types which are true positives and false negatives. Generally, true positives happen when the objects are successfully identified when they are supposed to be identified, while false negatives happen when the objects are not identified when they are supposed to be identified. Human full body is indicated using dark blue colored bounding box, human head-shoulder is indicated using magenta colored bounding box and car is indicated using orange colored bounding box.

5.1 Human

Here, we use two videos from CAVIAR datasets, one from SPEVI datasets[35] (Courtesy of EPSRC funded MOTINAS project (EP/D033772/1)) and four from recorded videos in IIUM to test our proposed technique. The videos that were used are in different environments and lighting variations. The videos from SPEVI datasets and recorded videos in IIUM include occlusion where the human full body cannot be identified. Hence, the system extracts and identifies the human head-shoulder when the human full body cannot be identified. In this case, true positives happen when human full body or human head-shoulder is successfully identified when it is supposed to be identified, while false negatives happen when human full body or human head-shoulder is not identified when it is supposed to be identified.

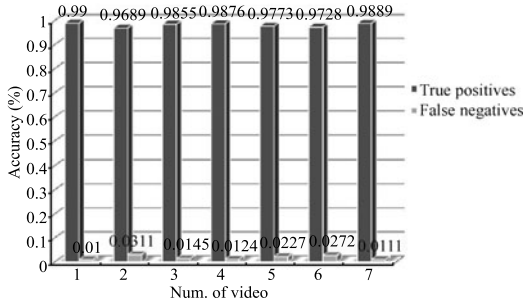


Fig. 7 Graphical representation of human identification result

During evaluation, the following criteria are taken into account: humans with less than 50% visibility at image boundaries and humans outside effective detection area region of the datasets are not used in the evaluation process. The effective detection area corresponds to area in which human can be successfully detected due to preservation of human size. The results for human identification using proposed technique is shown in Table 7. The graphical representation of human identification result for each video is shown in Fig. 7. The seven test videos used can give a clear picture on the accuracy of the system in the scenes with several numbers of human presences even in the presence of partial occlusion. The overall accuracy for our proposed technique is summarized in Table 8, where the overall accuracy is very good, which is 98.33%.

The examples for human identification using our proposed technique are shown in Figs. 8–14. Our proposed technique gives robust results by identifying human even in the presence of occlusion and under different environments and lighting variations. We used IIUMvideo 1 as our benchmarked video to identify effectiveness of our algorithm to extract and detect human head-shoulder in the presence of occlusion. Our system can identify human by extracting the human head-shoulder up to 60–70% occlusion as shown in Fig. 8.

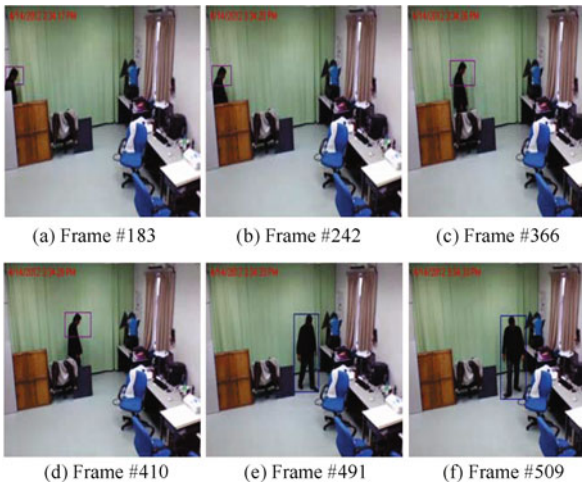


Fig. 8 Results from IIUMvideo 1

Table 7 Human identification result using proposed technique

No.	Name of video	Total Num. of frames	Total Num. of Humans		Angle of camera (degree)	Height of camera (m)	Num. of true positives	Num. of false negatives	Accuracy (%)
			Total humans	Different humans					
1	IIUMvideo 1	609	431	1	−30	3.0	427	4	99
2	IIUMvideo 2	382	257	1	−30	2.7	249	8	96.89
3	IIUMvideo 3	207	207	2	−30	3.0	204	3	98.55
4	Motinas_room 160_audiovisual (SPEVI datasets)	1072	1295	2	–	–	1279	16	98.76
5	OneStopMoveNoEnter 2front(CAVIAR datasets)	1034	925	2	–	–	904	21	97.73
6	IIUMvideo 4	1670	2020	3	−30	3.0	1965	55	97.28
7	OneStopEnter2cor (CAVIAR datasets)	2724	3864	6	–	–	3821	43	98.89

From this video sequences, we can observe that whenever occlusion is not occurring, our system identifies human full body as shown in Figs. 8 (d) and (e). Average frame rate for our proposed system is shown in Table 9. As we can see, the frame rate decreased a little bit as the number of humans in the scene increased. The whole system works at approximately 16–29 fps and it is suitable for real-time applications.

Table 8 Overall human identification result using proposed technique

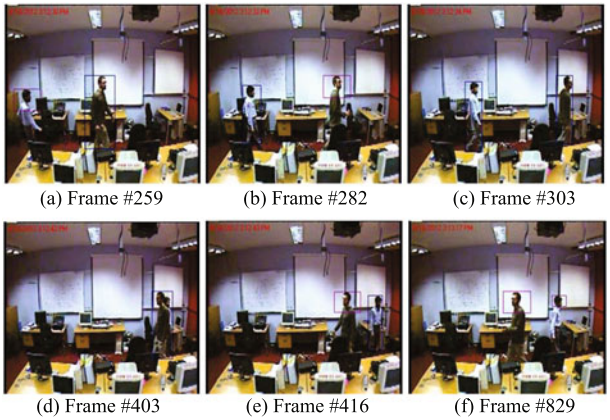
Total humans	Num. of true positives	Num. of false negatives	Accuracy (%)
8 999	8 849	150	98.33

(a) Frame #58 (b) Frame #89 (c) Frame #162

Fig. 9 Results from IIUMvideo 2

(a) Frame #159 (b) Frame #172 (c) Frame #199

Fig. 10 Results from IIUMvideo 3



(a) Frame #259 (b) Frame #282 (c) Frame #303

(d) Frame #403 (e) Frame #416 (f) Frame #829

Fig. 11 Results from motinas_room160_audiovisual (SPEVI datasets)

(a) Frame #408 (b) Frame #433 (c) Frame #447

Fig. 12 Results from OneStopMoveNoEnter2front (CAVIAR datasets)

(a) Frame #991 (b) Frame #1284 (c) Frame #1600

Fig. 13 Results from IIUMvideo 4

(a) Frame #2374 (b) Frame #2499 (c) Frame #2610

Fig. 14 Results from OneStopEnter2cor (CAVIAR datasets)

Table 9 Average frame rate for our proposed technique

Num. of human	Frame size (width × height) (pixels)	Original frame rate (fps)	Frame rate for our proposed technique (fps)
	320 × 240	30	27−29
1	360 × 288	25	19−21
	384 × 288	25	19−21
	320 × 240	30	27−29
2	360 × 288	25	16−19
	384 × 288	25	16−19
	320 × 240	30	–
3	360 × 288	25	–
	384 × 288	25	16−19

5.2 Car

Here, we use five recorded videos in IIUM to test our proposed technique. In this case, true positives happen when car is successfully identified when it is supposed to be identified, while false negatives happen when car is not identified when it is supposed to be identified. During evaluation, the following criteria are taken into account: cars with less than 65% visibility at image boundaries and cars outside effective detection area region of the datasets are not used in the evaluation process. The effective detection area corresponds to area in which car can be successfully detected due to preservation of car size.

The result for car identification using the proposed technique is shown in Table 10. The graphical representation of car identification result for each video is shown in Fig. 15. The five test videos used can give a clear picture on the accuracy of the system in the scene. The overall accuracy for our proposed technique is summarized in Table 11, where the overall accuracy is very good, which is 94.41%. The examples for car identification using our proposed technique are shown in Figs. 16−20. Our proposed technique gives robust results by identifying car under different environments as shown in Figs. 16–20. Human is also identified when human is present in the scene as shown in Fig. 20 (b). The whole system works at approximately 24–29 fps and it is suitable for real-time applications.

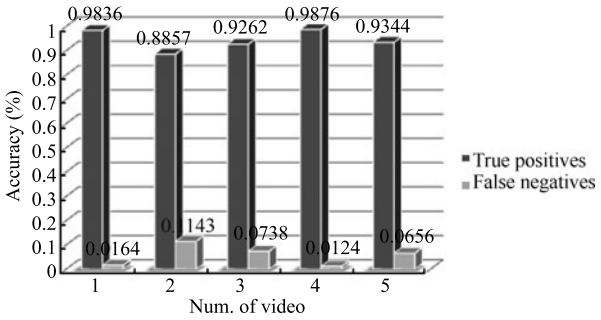


Fig. 15 Graphical representation of car identification result

Fig. 16 Results from IIUMvideo 5

Fig. 17 Results from IIUMvideo 6

5.3 Comparison of results

We compare the results of our technique for human identification with the technique in [15], the results of accuracy is given in Table 12. In order to do the comparison, we used four videos which include partially occluded humans with background objects. Using our method gives a much higher accuracy of the system as compared to the method of human identification as in [15] as shown in Table 13. We found out that using aspect ratio only as in [15], the system cannot identify the human whenever partial occlusion occurs or the extracted human is outside the human ratio limit as shown in the first row of Fig. 21 (a). In contrast, using our technique, human is still identified by extracting and detecting the human head-shoulder as indicated using magenta coloured bounding box as shown in the second row of Fig. 21 (a). In addition, we found out that using aspect ratio only, the system sometimes detects non human as human also, which is known as false positives error. Meanwhile, our proposed technique uses moment invariants and aspect ratio to identify the extracted objects as human. Hence, our proposed technique can reduce this false positives error as shown in Figs. 21 (b) and (c). Fig. 22 shows the graphical representation of the comparison results of accuracy in identifying human with the technique in [15].

Fig. 18 Results from IIUMvideo 7

Fig. 19 Results from IIUMvideo 8

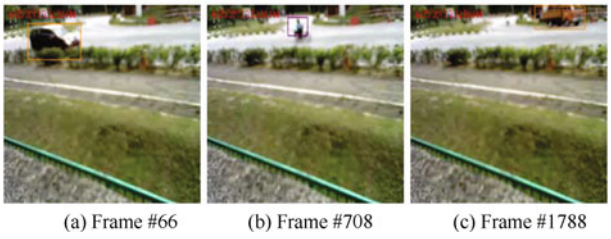


Fig. 20 Results from IIUMvideo 9

Table 10 Car identification result using proposed technique

No.	Name of video	Total Num. of frames	Total Num. of cars		Angle of camera (degree)	Height of camera (m)	Num. of true positives	Num. of false negatives	Accuracy (%)
			Total cars	Different cars					
1	IIUMvideo 5	288	61	2	0	0.5	60	1	98.36
2	IIUMvideo 6	479	70	2	0	0.5	62	8	88.57
3	IIUMvideo 7	486	122	3	0	0.5	113	9	92.62
4	IIUMvideo 8	425	161	4	0	0.5	159	2	98.76
5	IIUMvideo 9	1999	320	9	−5	1	299	21	93.44

5.4 Overall results

The overall accuracy for our proposed technique in identifying human and car is summarized in Table 14, where the overall accuracy is very good, which is 98.04%.

Fig. 21 Comparison of results: first row using technique in [15] and second row using our proposed technique

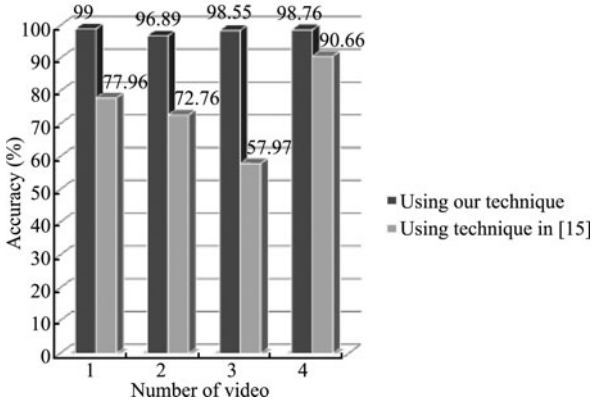


Fig. 22 Graphical representation of comparison results

Table 11 Overall car identification result using proposed technique

Total cars	Num. of true positives	Num. of false negatives	Accuracy (%)
734	693	41	94.41

Table 12 Results of human classification using method in [15]

No.	Name of video	Total num. of humans	Num. of true positives	Num. of false negatives	Num. of false positives	Accuracy (%)
1	IIUMvideo 1	431	336	95	0	77.96
2	IIUMvideo 2	257	187	70	6	72.76
3	IIUMvideo 3	207	120	87	0	57.97
4	motinas_room160_ audiovisual (SPEVI Datasets)	1 295	1 174	121	18	90.66

Table 13 Comparison results of human classification using proposed method and using method in [15]

	Total human in four videos	Num. of true positives	Num. of false negatives	Accuracy (%)
Using method in [15]	2 190	1 817	373	82.97
Using proposed method	2 190	2 159	31	98.58

Table 14 Overall identification result using proposed technique

Total humans and cars (in all videos)	Num. of true positives in identification	Num. of false negatives in identification	Accuracy (%)
9 733	9 542	191	98.04

6 Conclusion

In this paper, we presented a humans and cars identification system employing background subtraction, foreground segmentation, shadow removal, feature extraction and identification, which is suitable for real-time video surveillance system. The system has a better identification result on solving the issue of partially occluded human as it can detect the human head-shoulder whenever the human full body cannot be detected. Our system identifies human by extracting the human head-shoulder up to 60%−70% occlusion. The algorithm presented produces a good result, reduces false negatives and false positives error and also in real-time. The computation is fast and does not consume much processing power which makes this technique very suitable for analyzing fast video streams at high frame rates. This technique also could be implemented into objects tracking system. However, the technique used could be further improved to identify humans and cars in fully occluded scenes. Moreover, in the future the code for the software will also be revised to multi-thread implementation for better performance with multi-core processors.

Acknowledgments

The authors thank the IIUM for supporting this work. Portions of the research in this paper use the CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences.

References

[1] K. Goya, X. Zhang, K. Kitayama, I. Nagayama. A method for automatic detection of crimes for public security by using motion analysis. In *Proceedings of the 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, IEEE, Kyoto, Japan, pp. 736–741, 2009.

[2] J. Connell, A. W. Senior, A. Hampapur, Y. L. Tian, L. Brown, S. Pankanti. Detection and tracking in the IBM peoplevision system. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, Taipei, China, pp. 1403–1406, 2004.

[3] N. Dalal, B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, San Diego, CA, USA, pp. 886–893, 2005.

[4] Y. F. Mao, X. N. Huang. Human recognition based on head-shoulder moment feature. In *Proceedings of IEEE International Conference on Service Operations and Logistics, and Informatics*, IEEE, Beijing, China, pp. 622–625, 2008.

[5] J. Flusser, T. Suk. Pattern recognition by affine moment invariants. *Pattern Recognition*, vol. 26, no. 1, pp. 167–174, 1993.

[6] J. Liu, D. R. Li, W. B. Tao, L. Yan. An automatic method for generating affine moment invariants. *Pattern Recognition Letters*, vol. 28, no. 16, pp. 2295–2304, 2007.

[7] T. Lv, B. Ozer, W. Wolf. A real-time background subtraction method with camera motion compensation. In *Proceedings of IEEE International Conference on Multimedia and Exhibition*, IEEE, Taipei, China, vol. 1, pp. 331–334, 2004.

[8] C. R. Wren, A. Azarbayejani, T. Darrell, A. P. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, 1997.

[9] H. A. Rahim, U. U. Sheikh, R. B. Ahmad, A. S. M. Zain, W. N. F. W. Ariffin. Vehicle speed detection using frame differencing for smart surveillance system. In *Proceedings of the 10th International Conference on Information Sciences Signal Processing and Their Applications (ISSPA)*, IEEE, Kuala Lumpur, Malaysia, pp. 630–633, 2010.

[10] A. J. Lipton, H. Fujiyoshi, R. S. Patil. Moving target classification and tracking from real-time video. In *Proceedings of the 4th IEEE Workshop on Applications of Computer Vision*, IEEE, Princeton, NJ, USA, pp. 8–14, 1998.

[11] H. Fujiyoshi, T. Komura, I. Eguchi, K. Kayama. Road observation and information providing system for supporting mobility of pedestrian. In *Proceedings of the 4th IEEE International Conference on Computer Vision Systems*, IEEE, Washington, DC, USA, pp. 37, 2006.

[12] E. L. Andrade, S. Blunsden, R. B. Fisher. Characterisation of optical flow anomalies in pedestrian traffic. In *Proceedings of the IEE International Symposium on Imaging for Crime Detection and Prevention*, IEE, London, UK, pp. 73–78, 2005.

[13] J. Li, S. G. Nikolov, N. E. Scott-Samuel, C. P. Benton. Reliable real-time optical flow estimation for surveillance applications. In *Proceedings of the Institution of Engineering and Technology Conference on Crime and Security*, IEEE, London, UK, pp. 402–407, 2006.

[14] S. Denman, C. Fookes, S. Sridharan. Improved simultaneous computation of motion detection and optical flow for object tracking. In *Proceedings of 2009 Digital Image Computing: Techniques and Applications (DICTA'09)*, IEEE, Washington, DC, USA, pp. 175–182, 2009.

[15] F. Hafiz, A. A. Shafie, O. O. Khalifa, M. H. Ali. Foreground segmentation-based human detection with shadow removal. In *Proceedings of 2010 International Conference on Computer and Communication Engineering (ICCCE 2010)*, IEEE, Kuala Lumpur, Malaysia, pp. 1–6, 2010.

[16] A. J. Lipton. Local application of optic flow to analyse rigid versus non-rigid motion. In *Proceedings of International Conference on Computer Vision Workshop Frame-Rate Vision*, 1999.

[17] S. Harasse, L. Bonnaud, M. Desvignes. Human model for people detection in dynamic scenes. In *Proceedings of the 18th International Conference on Pattern Recognition*, IEEE, Hong Kong, China, vol. 1, pp. 335–354, 2006.

[18] F. H. B. K. Zaman. Automated Human Recognition and Tracking for Video Surveillance System, Master dissertation, International Islamic University Malaysia, Malaysia, pp. 47–51, 2010.

[19] W. M. Hu, T. N. Tan, L. Wang, S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 34, no. 3, pp. 334–352, 2004.

[20] G. S. K. Fung, N. H. C. Yung, G. K. H. Pang, A. H. S. Lai. Effective moving cast shadow detection for monocular color image sequences. In *Proceedings of the 11th International Conference on Image Analysis and Processing (ICIAP)*, IEEE, Palermo, Italy, pp. 404–409, 2001.

[21] C. X. Wang, W. J. Zhang. A robust algorithm for shadow removal of foreground detection in video surveillance. In *Proceedings of 2009 Asia-Pacific Conference on Information Processing (APCIP 2009)*, IEEE, Washington, DC, USA, vol. 2, pp. 422–425, 2009.

[22] I. V. Safonov. Automatic correction of amateur photos damaged by backlighting. In *Proceedings of International Conference on Computer Graphics and Vision (GraphiCon 2006)*, Novosibirsk Akademgorodok, Russia, pp. 80–89, 2006.

[23] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/.

[24] A. S. Abutaleb. Automatic thresholding of gray-level pictures using two-dimensional entropy. *Computer Vision, Graphics and Image Processing*, vol. 47, no. 1, pp. 22–32, 1989.

[25] J. N. Kapur, P. K. Sahoo, A. K. C. Wong. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics and Image Processing*, vol. 29, no. 3, pp. 273–285, 1985.

[26] J. Kittler, J. Illingworth. Minimum error thresholding. *Pattern Recognition*, vol. 19, no. 1, pp. 41–47, 1986.

[27] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[28] Q. Ye. A robust method for counting people in complex indoor spaces. In *Proceedings of the 2nd International Conference on Education Technology and Computer (ICETC)*, IEEE, Shanghai, China, vol. 2, pp. 450–454, 2010.

[29] L. Liu, F. Yang, P. Zhang, J. Y. Wu, L. Hu. SVM-based ontology matching approach. *International Journal of Automation and Computing*, vol. 9, no. 3, pp. 306–314, 2012.

[30] F. Liu, X. Liu, B. Zhang, J. Bai. Extraction of target fluorescence signal from *in vivo* background signal using image subtraction algorithm. *International Journal of Automation and Computing*, vol. 9, no. 3, pp. 232–236, 2012.

[31] CASIA Gait Database, http://www.sinobiometrics.com.

[32] B. Leibe, B. Schiele. Interleaved object categorization and segmentation. In *Proceedings of British Machine Vision Conference (BMVC' 03)*, Norwich, UK, pp. 759–768, 2003.

[33] A. M. Ibrahim, A. A. Shafie, M. M. Rashid. Human identification system based on moment invariant features. In *Proceedings of the International Conference on Computer and Communication Engineering (ICCCE 2012)*, IEEE, Kuala Lumpur, Malaysia, pp. 216–221, 2012.

[34] M. K. Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[35] http://www.eecs.qmul.ac.uk/ andrea/spevi.html.

Amir Akramin Shafie received his B. Eng. (Honours) degree in mechanical engineering from University of Dundee, UK, his M. Sc. degree in mechatronics from University of Abertay Dundee, UK, and his Ph. D. degree in engineering from University of Dundee. From 2000–2005, he was a researcher in Standards and Industrial Research Institute of Malaysia (SIRIM) Berhad. He is currently working at Department of Mechatronics Engineering, International Islamic University Malaysia, Malaysia as associate professor. He has published various articles in books, refereed journals and various international conferences, some of which have been highly cited.

His research interests include machine vision, intelligent system and autonomous system.

E-mail: aashafie@iium.edu.my (Corresponding author)

Azhar Bin Mohd Ibrahim received his B. Eng. (Honours) degree in mechatronics engineering from International Islamic University Malaysia in 2010. Currently, he is a master student in mechatronics engineering in International Islamic University Malaysia.

His research interests include machine vision, intelligent system, artificial intelligence, and surveillance system.

E-mail: mazary86@gmail.com

Muhammad Mahbubur Rashid received his B. Sc. (Eng.) degree in electrical and electronic engineering from Bangladesh University of Engineering and Technology, Bangladesh. He received his M. Sc. and Ph. D. degrees from the University of Malaya, Malaysia in 2003 and 2007, respectively. From 1994 to 2000, he was a sub divisional engineer (instrumentation and control) with the Bangladesh Power Development Board. Since 2007, he has been an assistant professor in Department of Mechatronics Engineering, International Islamic University Malaysia, Malaysia. He has published more than 65 papers in journals and conference proceedings.

His research interests include advanced control system and simulation and nonlinear modeling, process control and industrial automation, instrumentation, neural networks, artificial intelligence, power electronics, and renewable energy.

E-mail: mahbub@iium.edu.my