An Algorithm for Intelligent Identification of Moving Objects in Natural Environment

Bhupendra Kumar Yadav and Jian Xiaogang

Abstract—**The intelligent moving object detection has become one of the key research areas in the computer vision. Although, there are a lot of researches and methods have been proposed related to the intelligent moving object detection, and visual surveillance and intelligent recognition system. However, still it's a great challenge of intelligent identification of moving object detection in the natural environment, due to the natural factors such as wind, sunlight, lighting and sudden illumination change which has been affecting the accuracy of moving object detection and intelligent recognition. For example, wind makes swaying trees and water rippling; sunlight makes shadows; lighting causes sudden change of light. To eliminate these problems, we have proposed a hybrid novel method based on Gaussian mixture model (GMM); Background subtraction; HSV color model; Feature extraction; and Neural networks. First, background is modeled with Gaussian Mixture Model (GMM), to eliminate the effect caused by the natural environment. Second, foreground image is extracted with background subtraction method. Third, the shadows of moving objects are detected and removed in HSV color model and morphological operation is done to get the clean foreground. That means detection is completed. Then, it is updated the background to adapt the dynamic background. After object detection, it is extracted shape features by using the Hu's seven moment invariants of the training samples of the image data, which is used to train the back propagation neural network (BPNN) as input. Finally, we have done the intelligent identification process on the trained BPNN to recognize and distinguish the detected object whether it is human or pets. The algorithm can not only eliminate the effect of natural conditions, like wind, sunlight and lightning, but also automatically update the background when the illumination changes suddenly, or moving objects stop to move, or the background objects turn to move. The advantages of the proposed algorithm are accurately moving object detection, and the detection result is not affected by the body pose. The experimental results have shown that the proposed algorithm has good robustness and real-time performance in natural environment.**

Index Terms—**Feature extraction, HSV color model, intelligent identification, neural networks.**

I. Introduction

The intelligent video surveillance system is currently of huge interest in computer vision research due to its implications in the several fields related to the security and intelligence area. It has enormous benefits over the traditional human operated surveillance system, because of this it has become part of our daily life now. It has been widely used in security control, traffic control, abnormal human behavior detection, face detection, control systems mostly everywhere such as schools, malls, markets, companies etc. However, it is very challenging tasks in many computer vision applications because it is influenced by the several things like weather, lighting, background, pose, and so on. Basically, intelligent video surveillance uses the computer to analyze image sequences and video frames automatically with the several techniques and processes. For the moving object detection the video analyze to detect the objects, then track such objects and analyze it frame by frame and recognize the specific behavior of the objects intelligently [1].

There are three main basic methods of moving object detection which are using widely all the time till now such as background subtraction, frame difference and optical flow method. Background subtraction method [2] is a commonly used for moving object detection in the static scene. It detects the moving object by subtraction the current image from the background image pixel-by-pixel. This method is very simple and easy to implement, but it's very sensitive to the dynamic changes like sudden illumination changes, parked car moved out etc. In frame differencing method [3] the moving objects are detected by taking differences of consecutive frames in a video sequences. This method is adaptable to the dynamic changes in the scenes, but mostly fails to detect the complete region of the moving objects and usually occurs the 'holes' phenomenon, when the moving objects move too fast or too slow. Optical flow method [4] uses the flow vectors of moving objects over time to detect moving objects in image sequences. This method is effective and adaptable to the dynamic scenes, but very time consuming and computationally very complex. Usually, it is very difficult to meet the requirement for real-time performance without any special hardware.

Background modeling and estimation is very essential task for the intelligent video surveillance. As it models the background and detect the foreground objects in the moving object detection. In the complex environments, like dynamic background, illumination changes, some objects being introduced and removed suddenly from the video scene, we need a robust and adaptable background modeling algorithms and methods to adjust and resolve these problems. There are many different background modeling methods have been proposed and developed by the researchers, and they are categorized in to pixel based and region based. Most popular and simple methods are as Gaussian mixture model

Manuscript received October 18, 2019; revised March 6, 2020.

The authors are with mechanical engineering department in Tongji University, Shanghai, China (e-mail: 1793129@tongji.edu.cn).

doi: 10.18178/ijmlc.2020.10.5.985

(GMM), SOBS, SACON, Vibe, kernel density estimation (KDE), local binary pattern (LBP), Codebook, PBAS, and AGGM. Gaussian mixture model (GMM) method is a most famous and easy to implement in the complex background. GMM is a pixel based method, in which every pixel is processed with the mixture of K Gaussians [5]-[8].

Generally, the intelligent video surveillance system have two steps, first moving object detection, second feature extraction and object classification. We have already discussed about moving object detection methods and their applications above. The feature extraction and object classification are very important processes in the intelligent video surveillance. Basically, it is related to the features of the object like shape, color, texture, size and so on. After feature extraction the object is need to be classified as the final result for the intelligent identification. There are many deferent methods have been use for the feature extraction such as Haar wavelet, Gabor wavelet, Hu moment invariants, local binary pattern (LBP), Zernike moments, Fourier descriptors and so on. The Hu moment is one of the popular and easy to implement for feature extraction. The object classification is the final step of the intelligent video surveillance system, because it classifies the shape and characters of the given targets. K-NN classifier, artificial neural networks (ANN), Support vector machine (SVM), Decision tree and so on. The ANN is one of the most popular classifier in the intelligence area. As it has a very intelligent property to adjust the result according to the errors and targets. ANN gives the very accurate and satisfied results especially for the image processing applications [9]-[13].

In this paper, we have proposed an algorithm for intelligent moving object detection and object classification. The background modeling and the moving object is detected by combining of Gaussian mixture model with the background subtraction and three-frame difference methods. Then, the HSV color model is used to detect and remove the shadow. Finally, the Hu moment with the artificial neural network (ANN) is used to extract the features and classify the characters of the detected moving object. This paper is organized as follows. Section II describes the difficulties in detection. Section III describes algorithm processes and methodology in detail. Section IV shows the experiments and results. Section V concludes and summarizes the paper.

II. Challenges and Difficulties

Intelligent moving object detection and video surveillance in natural environment remains a broad and an open research problem even after the several years of research in this field. A robust, accurate and high performance approach is still a great challenge today. The difficulty level of this problem highly depends on how you define the object to be detected. It may depend on variation in object pose or deformations, variation in illumination and partial/full occlusion of the object [14].

The typical challenges for the research algorithm of the intelligent moving object detection and video surveillance are summarized as below:

A. Illumination Changes

It is consider that background model adapts to gradual changes of the appearance of the environment. For example in outdoor settings, the light intensity typically varies during day. Sudden illumination changes can also occur in the scene. This type of change occurs for example with sudden switching on/off a light in indoor environment. This may also happen in outdoor scenes (fast transition from cloudy to bright sunlight). Illumination strongly affects the appearance of background, and cause false positive detections.

B. Dynamic Background

The background scenery may contain movement (a fountain, movements of clouds, swaying of tree branches, wave of water etc.). Such movement can be periodical or irregular (e.g., traffic lights, waving trees). Handling such dynamics background is a very difficult and crucial task in the process of intelligent object detection.

C. Shadows

The shadows made by moving objects that often has complication in processing steps subsequent to object detection. Overlapping shadows of foreground regions may misclassify as moving objects, which is the very difficult part to make their separation and classification.

D. Noise

In video processing, the noise is also one of the difficult tasks to deal with. The video signal is generally superimposed with noise. The video surveillance has to cope with such degraded signals affected by different types of noise, such as sensor noise or compression artifacts.

E. Occlusion

When the occlusion (partial/full) occurs, it may affect the process of computing the background frame. However, in real life situations, occlusion can occur anytime a subject passes behind an object with respect to a camera. It is hard to control.

F. Clutter

When there is presence of background clutter that makes the task of segmentation difficult. It is hard to model a background that reliably produces the clutter background and separate the moving foreground objects from that.

G. Camouflage

It may occur intentionally or not, some objects may poorly differ from the appearance of background, making the correct classification should be difficult. This is especially important in surveillance applications. Camouflage is particularly a problem for temporal differencing methods.

H. Speed of the Moving Objects

The speed of the moving object plays an important role in its detection. If the object is moving very slowly, the temporal differencing method will fail to detect the portions of the object preserving uniform region. On the other hand a very fast moving object leaves a trail of ghost region behind it in the detected foreground mask.

III. Algorithms and Methodology

This section describes the proposed algorithm and their implementation for the intelligent moving objects detection in natural environment in detail (Fig. 1).

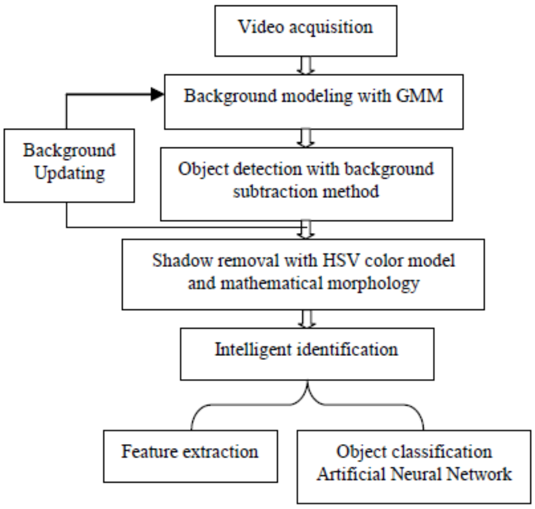

Video acquisition
Background modeling with GMM
Background Updating
Object detection with background subtraction method
Shadow removal with HSV color model and mathematical morphology
Intelligent identification
Feature extraction
Object classification Artificial Neural Network

Fig. 1. Algorithm flow chart.

A. Gaussian Mixture Model

To model a background which contains not only static objects but also dynamic objects and texture (such as water ripple, waving trees, water fountain and so on), to manage these situations, Stauffer and Grimson [15] proposed an adaptive Gaussian Mixture Model (GMM) for modeling the background. It models each pixel with the mixture of K (generally 3-5) Gaussians distributions. The Gaussian Mixture Model, also known as statistical background modeling method. The pixels of the background are distributed randomly. Zivkovic and Lee proposed some improved algorithms based on this [16], [17].

When K Gaussians distributions describe the recent history of the pixel, then, consider the pixel value at time t is the image sequences $I=\{X_1,X_2,.....,X_t\}$, and then the probability of the current pixel value X_t at time t is observed as:

$$P(X_t)=\sum_{t=1}^{k}\omega_{i,t}.\eta(X_t,\mu_{i,t},P_{i,t}) \tag{1}$$

where, $\omega_{i,t}$ is the weight of the i^{th} Gaussian at time t, and $\eta(X_t,\mu_{i,t},P_{i,t})$ is the probability density function of the i^{th} Gaussian at time t is:

$$\eta(X_t,\mu_{i,t},P_{i,t})=\frac{1}{(2\pi)^{\frac{n}{2}}|P_{i,t}|^{\frac{1}{2}}}e^{-\frac{1}{2}(X_t-\mu_{i,t})^T P_{i,t}^{-1}(X_t-\mu_{i,t})} \tag{2}$$

where, n is the number of color channels, $\mu_{i,t}$ is the mean of the i^{th} Gaussian at time t, and $P_{i,t}$ is the covariance matrix of the i^{th} Gaussian at time t. For computational reasons, GMM model assumes that the RGB color components are independent to each other and have the same variances. So, the covariance matrix is formed as:

$$P_{i,t}=\sigma_{i,t}^2.I \tag{3}$$

where, $\sigma_{i,t}^2$ is the variance of the i^{th} Gaussian at time t, and I is the image sequences acquired at time t. All the new pixel value X_t is compared with the existing K Gaussian distributions until it's matched. If a pixel value is 2.5 times of the standard deviation of distributions, then it is considered as matched with the current pixel. It can be formulated as:

$$|X_t^c-\mu_{i,t-1}^c|<2.5\sigma \tag{4}$$

(where, c is R, G, B colors channel).

When the Gaussian models matched the current pixel, then it is needed to be updated. For the matched models, the parameters updated as follows [18], [19]:

$$\omega_t=(1-\alpha)\omega_{t-1}+\alpha \tag{5}$$

$$\mu_t=(1-\rho)\mu_{t-1}+\rho X_t \tag{6}$$

$$\sigma_t^2=(1-\rho)\sigma_{t-1}^2+\rho(X_t-\mu_t)^T(X_t-\mu_t) \tag{7}$$

where α is the learning rate,

$$\rho=\alpha.\frac{1}{(2\pi)^{\frac{n}{2}}|P_i|^{\frac{1}{2}}}e^{-\frac{1}{2}(X_t-\mu_t)^T P_t^{-1}(X_t-\mu_t)} \tag{8}$$

For the unmatched models, mean μ and variance p are unchanged, only the weights are updated as follows:

$$\omega_t=(1-\alpha)\omega_{t-1} \tag{9}$$

After the weight is updated, then normalization of the weight is performed. If the current pixel does not match any of the existing K Gaussian models, then a new model is created, the least probable distribution is replaced with new one. The new value is replaced by ω/σ with the new model. In this paper α set to 0.1.

B. Background Subtraction

Background subtraction method is very simple and easy to implement for moving object detection [20]. In the background subtraction method, the moving object detects by subtracting the current image from the background image. If the difference pixel is above predefined threshold, then it is classified as foreground object. If $B(x,y)$ is a background image, and $G_i(x,y)$ is the current image, then the difference $D_i(x,y)$ is represented as,

$$D_i(x,y)=\begin{cases}G_i(x,y), & when\ D_i(x,y)\ge T_h\\ 0, & when\ D_i(x,y)<T_h\end{cases} \tag{10}$$

where, T_h is the predefined threshold. In this paper, the thresholds, $T_h = \frac{1}{2}(1-\sigma B(x,y))*(\max B(x,y)-\min B(x,y))$ where, σ is standard deviation.

C. Proposed Background Model

In this paper, we have proposed a novel approach for background modeling with combined Gaussian mixture model and the HSV color model to detect the moving objects in natural environment. In order to detect moving objects in the scene, it is first necessary to determine the area in the scene as the background. In a fixed period of time, a pixel with a small change in pixel parameters can be regarded as a background point.

Consider that the video frame size is $m \times n$, and each frame has $(m \times n)$ pixels. Convert the image to the HSV color model and create a vector G with elements of each pixel point (x, y), H value, S value, and V value, then G(x, y, 1), G(x, y, 2) and G(x, y, 3) represent the H value, the S value, and the V value of the pixel at (x, y) respectively. Select the first N frames of the video as the training frame, and Calculate the mean value of each of the three values of H, S, and V in the N frame, $\mu(x,y,c)$, and the standard deviation $\sigma(x,y,c)$, where c=1, 2 3. The decision formula is,

$$|G_i(x,y,c)-\mu(x,y,c)| < \lambda.\sigma(x,y,c) \qquad (11)$$

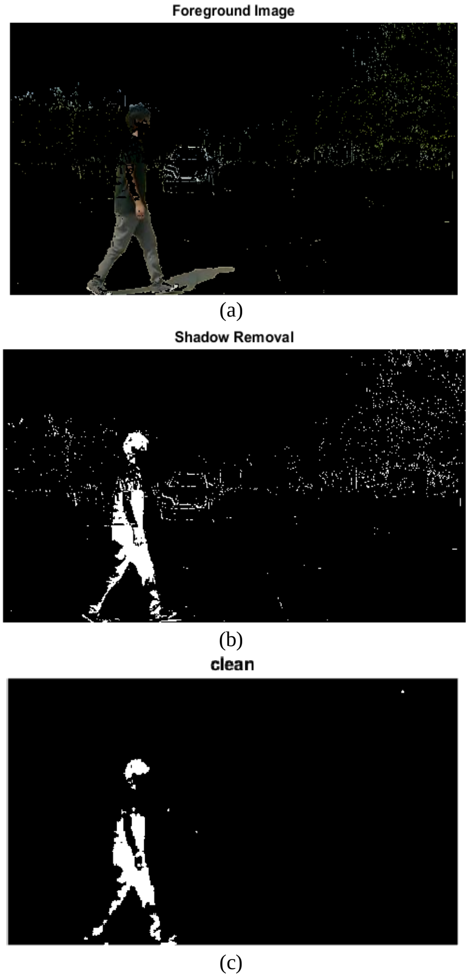
(λ=2) Obtain steady-state pixel values, and calculate the mean values $\mu_S(x,y,c)$, standard deviation $\sigma_S(x,y,c)$, and minimum of the three values of these steady-state pixels H, S, and V. The value $\min_S(x,\ y,\ c)$ the maximum value $\max_S(x,\ y,\ c)$, then the initial background reference image $B_1(x,\ y)$ can be determined by $\mu_S(x,y,c)$, the pixels in the image $B_1(x,\ y)$ are background points. Initially, we have used first 30-50 frames for background training and analysis according to the complexity of the background. The Gaussian mixture model will detect and normalize the dynamic background in natural environment like wind swaying trees or small gradual change in illumination.

D. Shadow Removal

Shadow detection and removal is one of the challenging tasks in the moving object detection. It can create serious problems during the intelligent video surveillance, as the shadow can be misclassified as moving object instead of real moving object. A shadow occurs when an object occluded by the direct light from the light source. Shadow can be divided into two categories self-projected shadow and cast shadow. Shadow could be any shapes and sizes which depend on the different objects [21]. Due to the misclassification of shadow as foreground the shadow should be deleted or suppressed. To solve this problem, we have proposed Hue-Saturation-Value (HSV) color model to analyze the shadows point. The main advantage of the HSV color model is as, it corresponds closely to the human perception of the color, and it has revealed more accurate in distinguish shadows and the object than RGB color model. In HSV color model the shadows change or reduce the background brightness (V-value) greatly while saturation (S-value) and hue (H-value) have little effect [22], [23]. The Value component of the HSV color model gives very useful information of the image, whether in color or gray scale images. So, by using these properties of the shadows in the HSV color model, the shadows point can be distinguished from the actual moving object point. According to the result, it can be formulated as follows:

$$D_i(x,y,c)=\begin{cases} 0,\ \text{if}(\alpha \le \dfrac{G_i(x,y,3)}{B_i(x,y,3)} \le \beta) \\ \qquad \cap(|G_i(x,y,2)-B_i(x,y,2)| \le Th_S) \\ \qquad \cap(|G_i(x,y,1)-B_i(x,y,1)| \le Th_H) \\ B_i(x,y,c), \qquad \text{otherwise} \end{cases} \qquad (12)$$

where, G is current image, B is the reference background frames, H, S, V represents the independent component of hue, saturation and brightness of the HSV color model. Where, α and β are threshold values which depends on light intensity. Its value is between 0 and 1 as $0<\alpha<\beta<1$. Th_S, and Th_H are the threshold values for saturation and hue, and these threshold values are set according to the experimental experience manually.



(a)

(b)

(c)

Fig. 2. (a) Foreground image with shadow (b) shadow removal binarized image (c) cleaned foreground after mathematical morphology.

After shadow removal we have used the open and closed operation of mathematical morphology to clean the noise and structuring element (SE) to define the size of the noise from the detected foreground image. The result has shown in the Fig. 2.

E. Background Update

The background parameter has to be updated in real time to adapt to the illumination gradient, to eliminate the interference caused by the illumination gradual change to the moving object, and to update the mean value, standard deviation, minimum and maximum value of the background point [24], the specific update formula can be expressed as,

$$\mu_i(x,y,c) = (1-\alpha)*\mu_{i-1}(x,y,c) + \alpha * G_i(x,y,c) \quad (13)$$

$$\sigma_i^2(x,y,c) = (1-\alpha)*\sigma_{i-1}^2(x,y,c) + \alpha*(G_i(x,y,c) \\ -\mu_{i-1}(x,y,c)) \quad (14)$$

$$\max_i(x,y,c) = \max(\mathrm{G}_i(x,y,c), \max_{i-1}(x,y,c)) \quad (15)$$

$$\min_i(x,y,c) = \min(\mathrm{G}_i(x,y,c), \min_{i-1}(x,y,c)) \quad (16)$$

The most initial values of the mean, standard deviation, and maximum and minimum values are the values of the parameters of the steady-state pixel.

F. Background Updates in Particular Cases

During the moving object detection, sometime some special circumstances occurs which affects the object detection accuracy in natural environment. To solve these problems the background reference image should be updated automatically. To deal with these cases, in this paper, we have proposed automatic updated algorithms as follows:

1) Sudden illumination change

When the lighting conditions suddenly change, then the color characteristics of the image change almost in every pixel, and it changes the area of the image pixel. So, it can be detected as moving object falsely. Therefore, it is necessary to check the situation of light mutation and adopt corresponding algorithm correction to ensure the correct and effective detection of moving objects. Since, the illumination change often causes the color characteristics of image change almost in every pixel in the scene. In this paper, we have used the area feature of the foreground image to determine whether a light mutation has occurred. If the detected foreground image area occupies most images scene, then it can be determined that the light is abrupt. To eliminate this situation, we have used 4-neibhours connected component method in the binary image. It has assumed that all 4 connected pixels are on the same object, so that each moving object in the binary image can be distinguished, and the number of moving objects is also obtained, and then the centroid and area of each object are calculated. When the total area of the detected moving object is greater than 80% of the image area, it can be determined that the illumination condition suddenly changes. If this situation does not last for a certain period of time, it can be considered as a short-lived light mutation such as lightning. It can skip the frame directly and choose to ignore the processing. If this situation continues for a certain period of time, it can be considered that the lighting condition has changed, and it is not temporary. At this time, it is necessary to re-analyze and train the background reference image.

2) When the moving objects stop to move or the background objects starts to move

When the moving object stopped to move, and if we use same background reference image to calculate, then it is detected as moving object. Also, when the background object turned in to motion, and if we use same background reference image to calculate then the color information changes, a contour appears, called "artifact". This situation will detect two moving objects, one is object and the other is artifact. These are the detection problem caused by this situation.

For these two cases, this paper decides to use the method of establishing feature vector for color information identification to determine the similarity of moving objects detected by frames. If it is determined to be similar, the moving objects detected by the current and previous frames are the same object. Euclidean Distance [25] is commonly used in signal science to calculate the similarity of two signals. Therefore, the Euclidean distance is used to characterize the similarity of the feature vectors of two foregrounds. Assuming that the vector $A = (a_1, a_2, \cdots, a_n)$, and the vector $B = (b_1, b_2, \cdots, b_n)$, then the Euclidean distance *d* between the vector A and vector B can be expressed as,

$$d = \sqrt{(a_1-b_1)^2 + (a_2-b_2)^2 + \ldots + (a_n-b_n)^2} \quad (17)$$

To establish the feature vector, this paper extracts the mean value μ_c, standard deviation σ_c and skewness S_c of the H, S and V values of each pixel in the moving object region from the corrected color foreground image, which is

$$\mu_c = \frac{1}{N}\sum_{(x,y)\in A} G(x,y,c) \quad (18)$$

$$\sigma_c = \sqrt{\frac{1}{N}\sum_{(x,y)\in A}(G(x,y,c)-\mu_c)^2} \quad (19)$$

$$S_c = \sqrt{\frac{1}{N}\sum_{(x,y)\in A}(G(x,y,c)-\mu_c)^3} \quad (20)$$

where, *A* is the area of the moving object in the colored foreground image.

If the Euclidean distance between the object feature vector of the current frame and the object feature vector of the previous frame is less than a certain threshold T_1, it is considered that the object features detected by the two frames before and after are similar, and it can be determined that not the new object enter the scene. If the linear distance between the centroid of the current frame object and the centroid of the previous frame object is less than a certain threshold *T*2, then the two object feature vectors are judged to be similar, it can be considered that the moving object is stationary, or the background object is turned into motion. When this situation continues for a certain period of time or longer, then the background reference image need to be re-established.

G. Feature Extraction

For the intelligent identification in moving object detection, the feature extraction process is very important for the image classification. It allows to extract the features of an image as ideal as possible. Feature extraction is applied to extract the most relevant features of an image, which will help to analyze the properties of an image feature, then, it will use to classify and recognize the specific features. There are several different types of feature extraction techniques such as color feature, texture feature and shape feature [26]. These all techniques have their own advantages for image classification. However, in this paper, we have proposed to use moment invariants feature extraction technique. Originally moment invariants method was introduced by Hu by using the central moments to construct the seven invariants, which can describe the shape of the region [27], [28]. Hu derived six absolute orthogonal invariants and one skew orthogonal invariant based on algebraic invariants, which are independent to the size, position and orientation. Due to the characteristics of translation, scale and rotation, invariants under the assumption of images with continuous functions and noise free. Because of, this properties moment invariants has become an important feature extraction method and widely used in image detection [29], [30].

Let's consider the two dimensional function $(p+q)^{th}$ order geometric moments are expressed as m_{pq}

$$m_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^p y^q f(x,y)dxdy \qquad (21)$$
$$p,q = 0,1,2....$$

When the image function $f(x, y)$ is a piecewise continuous bounded function, then the moments sequence m_{pq} and image function $f(x, y)$ is uniquely determined by each other. Then, the $(p+q)^{th}$ order central moments defined as follows:

$$\mu_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-x_0)^p (y-y_0)^q f(x,y)dxdy \qquad (22)$$
$$p,q = 0,1,2....$$

where, (x_0, y_0) are the pixel points of centroid of the image $f(x, y)$.

For $m \times n$ binary image $f(x, y)$, $x_0 = \frac{m_{10}}{m_{00}}$, $y_0 = \frac{m_{01}}{m_{00}}$

The centroid moments μ_{pq} computed using the centroid of the images $f(x, y)$ is equivalent to m_{pq} whose center has been shifted to centroid of the image. Therefore, normalization is performed with a zero-order central moment to obtain $(p+q)^{th}$ order normalized central moments as follows:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{r}}, \text{ where, } r = \frac{p+q}{2}+1$$

Based on normalized central moments, Hu introduced seven moment invariants as shown in equation,

$$\begin{aligned}
&\varphi_1 = \eta_{20} + \eta_{02} \\
&\varphi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \\
&\varphi_3 = (\eta_{30} + 3\eta_{12})^2 + (3\eta_{21} + \eta_{03})^2 \\
&\varphi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
&\varphi_5 = (\eta_{30} + 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\
&\quad (\eta_{21} + \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
&\varphi_6 = (\eta_{20} + \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + \\
&\quad 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
&\varphi_7 = (\eta_{21} + \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\
&\quad (\eta_{12} + \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
\end{aligned} \qquad (23)$$

The classification between people and pets are relatively simple and easy, because people mostly walking on two legs, while pets are walking on four feet, and there is very big difference in shape and pattern. Considering the simplicity of the classifier and the validity of the selection feature, this paper directly used 7 moment invariants as the input feature quantity of the classifier that distinguishes people from pets.

H. Artificial Neural Networks (ANN)

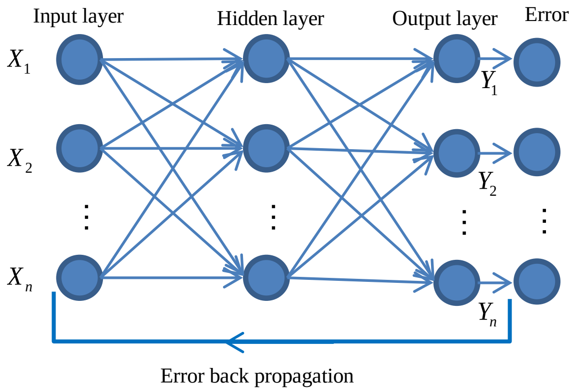


Fig. 3. BP neural network structure.

The artificial neural networks are a set of algorithms, as it modeled after the human brain, but not identical to it. Neural networks interpret through data labeling and clustering with raw input. They are used to classify the given data as input. It can identify and learn related shapes between input data sets and relevant target values. Neural networks are nonlinear system which can use for classification even complex shapes and patterns. In order to get desired results in different function and research, there are many different types of neural network models. However, in this paper, we have used back propagation neural network (BPNN) for classification as our output models. BPNN is a feed-forward neural network trained based on the neural network learning algorithm the error back propagation (Fig. 3). BPNN is very simple and it has capability in supervised shape matching [31], [32].

Basically, the BPNN is constructed of input layer, output layer, and hidden layer. The nodes are interconnected through the layers with the weights and biases. The basic principle of the BPNN has two steps as signal forward propagation and the error back propagation as shown in the Fig. 3. The input is processed from the input layer to each of the nodes through the hidden layer and then to the output layer. This process is a forward propagation process. If the desired output is not obtained at the output layer, the expected output of the output layer is the error between the actual input and the actual output. Then the error value is propagated back to the input layer in the form of iterative process layer by layer, then assigned to each connected nodes and calculated the error of the each node. On the basis of this error, the weights of each layer of nodes are adjusted until the final output reaches to the desired target value [33], [34].

Let's consider three layers BP neural network with one hidden layer. Let p is the input of the neural network, r is the nodes in the input layer, $s1$ is the nodes in the hidden layer, and $f1$ is the transfer function in the hidden layer. Then, $s2$ is the node in the output layer, $f2$ is the transfer function in the output layer, ω is the weight, b is the threshold, A is the output, and T is the target. Then, relationship of the BP neural network process can be shown as follows:

1) Forward propagation process:

The output of i^{th} nodes in hidden layer is as,

$$a1_i = f1(\sum_{j=1}^{r} \omega 1_{ij} p_j + b1_i), i = 1, 2, ..., s1 \tag{24}$$

The output of the k^{th} nodes in the output layer is as,

$$a2_k = f2(\sum_{j=1}^{s1} \omega 2_{ki} a1_i + b2_k), k = 1, 2, ..., s2 \tag{25}$$

2) Error back propagation process:

Error back propagation starts from the output layer and calculates the error through the nodes in the layers. Then, the weight and threshold of the layers adjusted based on the error gradient descent for the final output to correct the expected value. The error function E is as,

$$E = \frac{1}{2}\sum_{k=1}^{s2}(t_k - a2_k)^2 \tag{26}$$

Then, the weight changes from the i^{th} input of the hidden layer to the k^{th} output of the output layer is as,

$$\begin{aligned}\Delta\omega 2_{ki} &= -\eta\frac{\partial E}{\partial \omega 2_{ki}} = -\eta\frac{\partial E}{\partial a2_k}.\frac{\partial a2_k}{\partial \omega 2_{ki}} \\ &= \eta(t_k - a2_k).f2'.a1_i = \eta.\delta_{ki}.a1_i\end{aligned} \tag{27}$$

where,
$$\delta_{ki} = e_k.f2' \tag{28}$$

$$e_k = t_k - a2_k \tag{29}$$

Here, η is the learning rate. Similarly, the threshold changes from the i^{th} input of the hidden layer to the k^{th} output of the output layer is as,

$$\begin{aligned}\Delta b2_{ki} &= -\eta\frac{\partial E}{\partial b2_{ki}} = -\eta\frac{\partial E}{\partial a2_k}.\frac{\partial a2_k}{\partial b2_{ki}} \\ &= \eta(t_k - a2_k).f2'.a1_i = \eta.\partial_{ki}\end{aligned} \tag{30}$$

The weight changes from the j^{th} input of the input layer to the i^{th} output of the hidden layer is as,

$$\begin{aligned}\Delta\omega 1_{ij} &= -\eta\frac{\partial E}{\partial \omega 1_{ij}} = -\eta\frac{\partial E}{\partial a2_k}.\frac{\partial a2_k}{\partial a1_i}.\frac{\partial a1_i}{\partial \omega 2_{ij}} \\ &= \eta\sum_{k=1}^{s2}(t_k - a2_k).f2'.\omega 2_{ki}.f1'.p_j \\ &= \eta.\delta_{ij}.p\end{aligned} \tag{31}$$

where,
$$\delta_{ij} = e_i.f1' \tag{32}$$

$$e_i = \sum_{k=1}^{s2}\delta_{ki}.\omega 2_{ki} \tag{33}$$

Similarly, the threshold changes from the j^{th} input of the input layer to the i^{th} output of the hidden layer is as,

$$\Delta b1_{ij} = -\eta.\delta_{ij} \tag{34}$$

Repeatedly, the systematic error of the network will gradually decrease, and the learning will eventually converge to a stable set of weights. When the training meets the expected requirements, the weight of the interconnection between the nodes in each layer of the network is completely determined, then the BP neural network is trained, and the pattern can be identified for the unknown input.

IV. Experimental Results and Analysis

A. Training BP Neural Network

In MATLAB R2015a, we have designed a three-layered (input layer, hidden layer, and output layer) feed-forward BP neural network. Then, we have determined the parameters for training the network with training samples by using the function newff. In our case, we have collected 50 training samples among them 26 samples are human and 24 samples are for pets. And, also we have used 10 samples as test samples. The numbers of the nodes in the each layer of the networks have set as follows; the input layer has set to 7 nodes as we have used Hu's seven moment invariants for feature extraction which has given 7 inputs, and the output layer has set to 1 node. There are 26 nodes in the hidden layer as shown in Fig. 4. If the network does not converge during the training, then the number of nodes in the hidden layer can be appropriately increased. The transfer function of both the hidden layer and output layer has set to “logsig” because the output range of the function is 0~1, just as the output requirements. The initial value of the weights has chosen randomly between -1 and 1 which is distributed evenly. Which help to the rapid convergence of the network and avoid the situation where the network falls into local minimum point. The initial learning rate has set between 0.01-0.6. The learning function has selected the gradient descent momentum learning function “learngdm”. In addition, the training performance function is set to the mean squared error sum “sme”, the error index is 0.01. Finally, the BP network is trained using the function “traingda”. During training, when the network mean squared error is less than 0.01, or the number of the training cycles reaches 10,000 then the training of the network will stop. Fig. 4 and Fig. 5 shows the network training tool and best training performance at 96 epochs. Fig. 6 shows the trained network tested classification of human and pets.

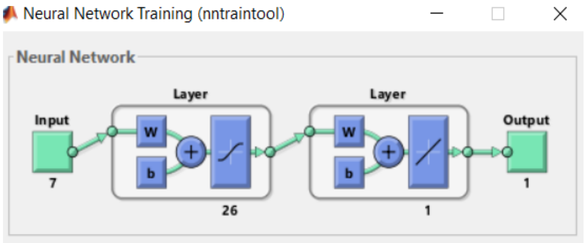


Fig. 4. Neural network training tool.

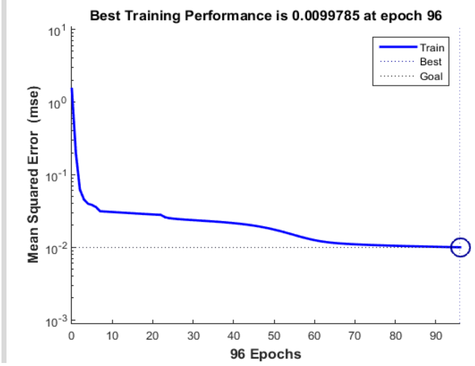


Fig. 5. Network training performance.

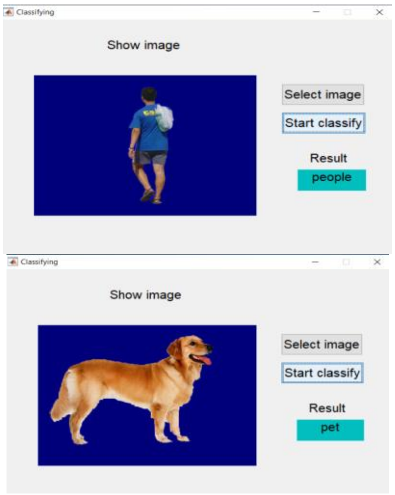


Fig. 6. Trained network classify people and pet.

B. MATLAB GUI Design for Identification

In order to implement the proposed algorithm and visualize the experiments, in this paper, we have designed GUI in MATLAB R2015a to detect the moving objects in natural environment. We have used USB webcam and phone camera to acquire the videos for the experiment, it was live stream surveillance. It has designed with three push buttons and four axes as boxes. The blue, green, and red buttons has set as video input, start detection and stop detection respectively as shown in (Fig. 7). In first box, it plays video and take snapshot, second box display foreground object, third box display trained background image, fourth box display binarized detected foreground object respectively. There has one result zone where it can be displayed the results of the detected objects and classify whether it is human or pet and save figure button use to save the required figures. There is counter zone which counts the per detection time, when it reached 100 counts then the detection system stop automatically.

We have tested our proposed algorithm through experiments in several conditions in natural environment as indoor and outdoor. To check the feasibility of the algorithm in natural environment, we have done the experiments in park, on road and in school area. We have done 50 experiments on human, and 4 times misdetected as pets. And we did 25 experiments for pets, and 2 times misdetected as human. In general, it shows that the proposed algorithm has 92% recognition rate in natural environment.

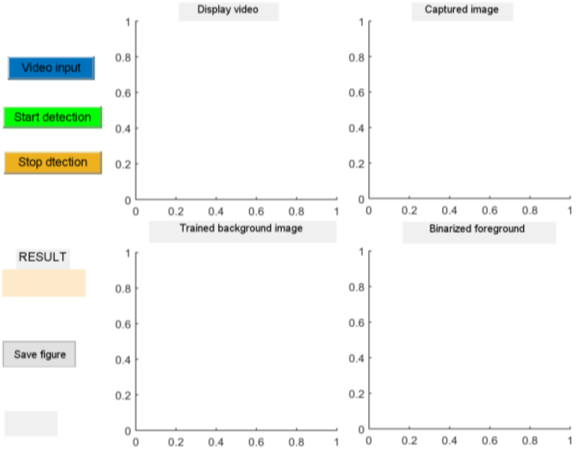
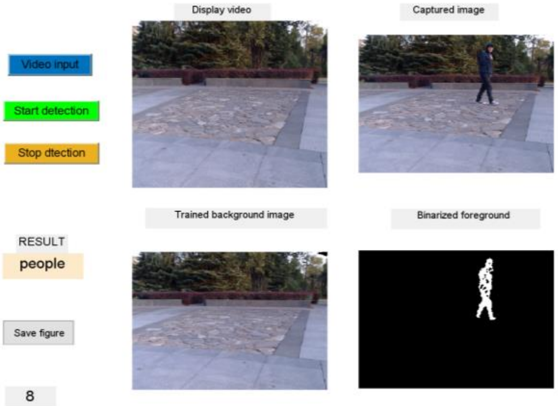
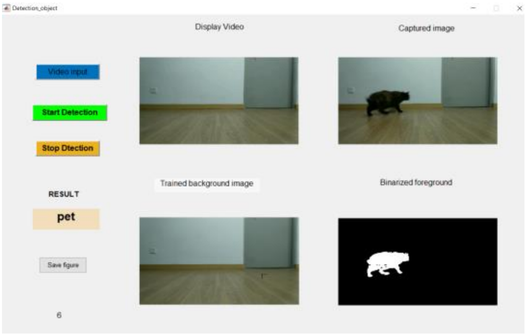


Fig. 7. GUI interface design for the experiment.

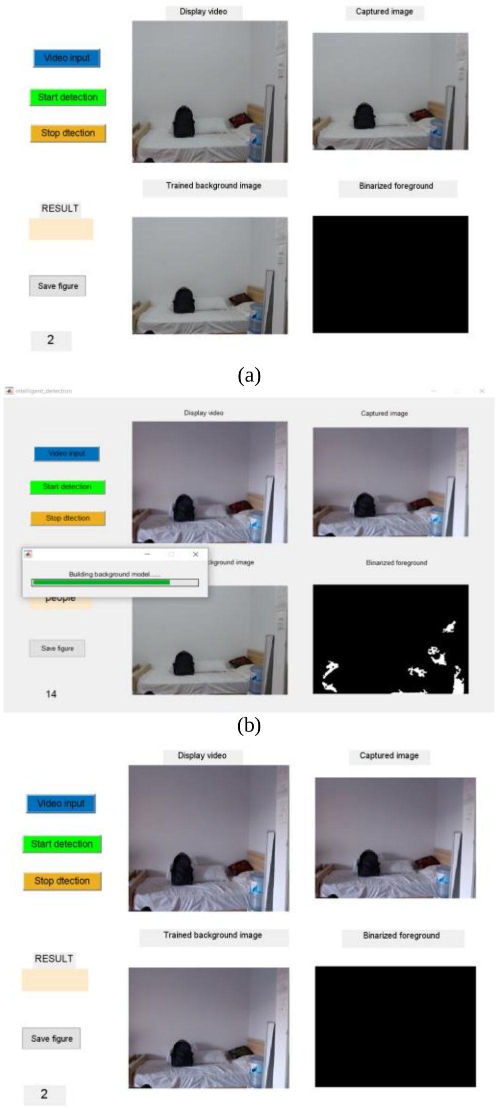


(a)



(b)

Fig. 8. Object detection and identification (a) Human detected (b) Pet detected.



(a)

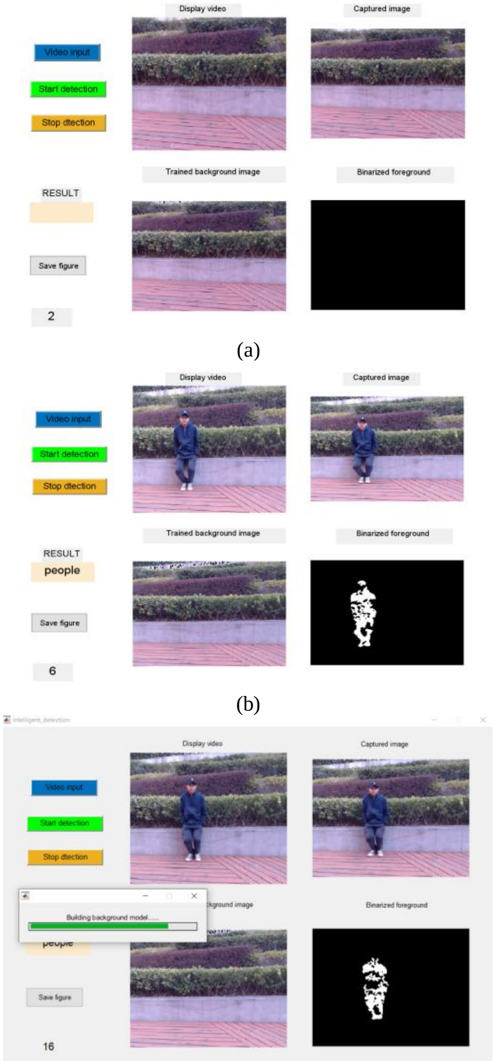
(b)

(c)

Fig. 9. Light mutation (a) Trained background (b) re-establish background (c) normal after background building.

To analyze and check the performance of the proposed algorithm, we have done several steps experiments as shown below. The Fig. 8(a) has shown the detected human which we have done in the school park. The background was very dynamic as trees were swaying because of wind, but it didn’t effect on the results. Fig. 8(b) has shown the detected pet in indoor where there was lighting variation. The Fig. 9 has shown the sudden light changed condition which we have done inside the room, where Fig. 9(a) has shown the normal background. In Fig. 9(b), we turned off the light then the algorithm judged that there is a sudden change in illumination which area has more than 80% in the scene, and if it occurs till 10 consecutive frames then it has jugged that the lighting condition changed permanently. Then the system reestablishes the background automatically. Fig. 9(c) has shown the background returned at normal condition. Fig. 10 has shown when the moving objects become stationary and become background. Fig. 10(a) is the original background, in Fig. 10(b) the moving object become background. In Fig. 10(c) the same moving object remains stationary for 10 consecutive frames in the scene then the system reestablish the background automatically and Fig. 10(d) has shown the background returned at normal condition. Fig. 11 has shown when the background objects turned in to motion. Fig. 11(a) has shown the original background, in Fig. 11(b) the background object start to move. In Fig. 11(c) the same background object remains stationary for 10 consecutive frames in the scene then the system reestablish the background automatically and Fig. 11(d) has shown background returned at normal condition after background rebuilding.

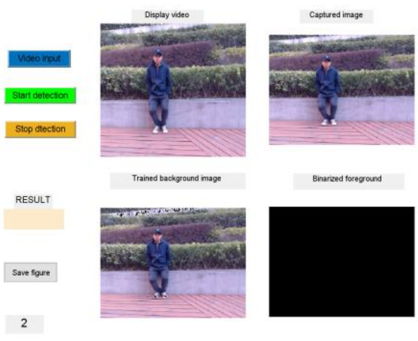
Therefore, the several steps of experiments have shown that the proposed algorithm has given satisfactory result in natural environment and it hasn’t influenced the results.



(a)

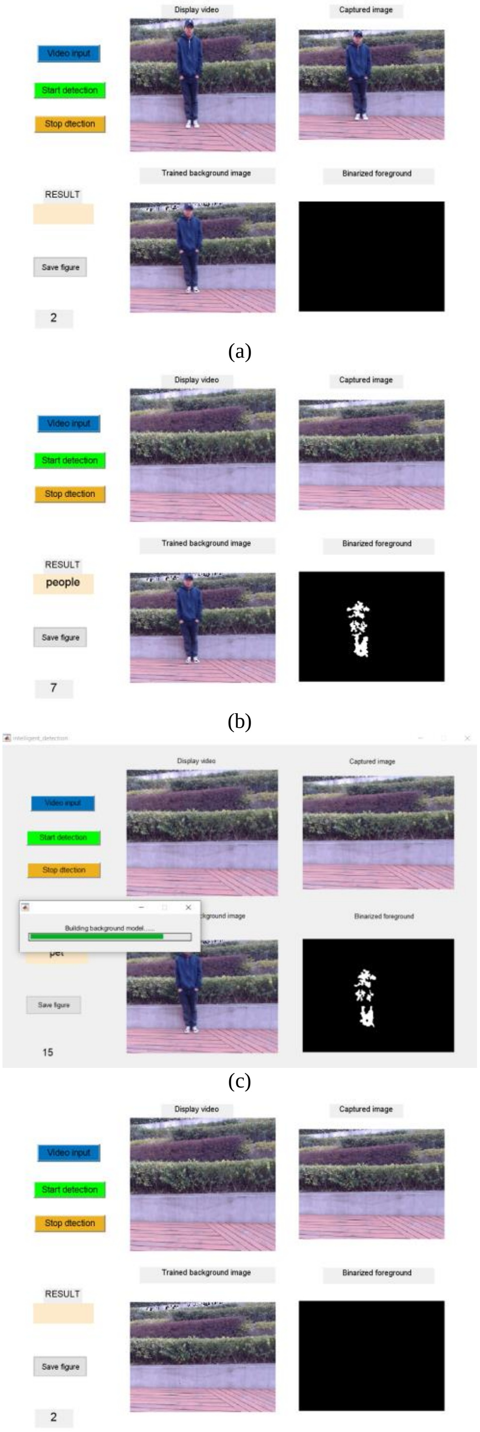
(b)

(c)



(d)

Fig. 10. (a) Trained background (b) moving object stop to move (c) re-establish the background (d) normal after re-established the background.



(a)

(b)

(c)

(d)

Fig. 11. (a) Trained background with still object (b) background object start to move (c) re-establish background (d) normal after re-established background.

V. Conclusion and Discussion

The proposed hybrid algorithm has conducted under the consideration of the natural environment based on Gaussian mixture model and the HSV color model through the background analysis, foreground extraction, shadow removal and background update steps to detect the moving objects completely and accurately. Intelligent recognition of objects, based on Hu invariant moments and BP neural network methods through the feature extraction of 7 moment invariants of moving objects, and proper training BP neural network and other steps accurately determine the category of moving objects. After the several steps of experiments in dynamic environment like wind swaying trees and different lighting condition, the proposed algorithm has shown satisfactory result in real-time performance in natural environment which has shown in the Section (IV)(B).

However, always there is room for improvement. It can be improved in future in several steps in our algorithm. In our algorithm we have used several thresholds value as it has adjusted manually. If possible to set all the thresholds values automatically then the result will be more robust and less time consuming, like thresholds for shadow removal. Also, our experiment has done with fixed single camera, in future if it is possible to use multiple camera detection method, then it will make complete detection and also solve the occlusion problem, although it is a challenging task till now.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

The authors Bhupendra Kumar Yadav and Jian Xiaogang have designed and conducted the main idea of this research paper. Bhupendra Kumar Yadav has written and developed all the theoretical, computational and experimental parts of this research under the supervision of the Jian Xiaogang. The data and results analyzed and verified by the Jian Xiaogang. Finally, we discussed and agreed to submit the final version of the research work.

Acknowledgment

I would like to thanks Mr. Jian Xiaogang for his support and guidance for this project. He is an associate professor in Tongji Uniersity and supervisor of this project. Also, I would like to thank my friends for helping me doing experiments.

References

[1] M. Paul, S. M. E. Haque, and S. Chakraborty, “Human detection in surveillance videos and its applications - A review,” *Eurasip Journal on Advances in Signal Processing*, no. 1, pp. 176, Nov. 2013.

[2] D. Farcas, C. Marghes, and T. Bouwmans, “Background subtraction via incremental maximum margin criterion: A discriminative subspace approach,” *Machine Vision & Applications*, vol. 23, no. 6, pp. 1083-1101, March 2012.

[3] K. H. Yang, Z. M. Cai, and L. L. Zhao, “Algorithm research on moving object detection of surveillance video sequence,” *Optics and Photonics Journal*, vol. 3, no. 2B, p. 5, June 2013.

[4] X. Han, G. Yuan, L. Zheng, Z. Zhang, and D. Niu, “Research on moving object detection algorithm based on improved three frame difference method and optical flow,” in *Proc. 2015 Fifth International Conference*

on Instrumentation and Measurement, Computer, Communication and Control (IMCCC), 2015.

[5] T. Bouwmans, F. E. Baf, and B. Vachon, "Background modeling using mixture of gaussians for foreground detection — A survey," *Recent Patents on Computer Science*, vol. 1, no. 3, pp. 219-237, Nov. 2008.

[6] Y. Xu, J. Dong, B. Zhang, and D. Xu, "Background modeling methods in video analysis: A review and comparative evaluation," *CAAI Transactions on Intelligence Technology*, vol. 1, no. 1, pp. 43-60, Jan. 2016.

[7] K. Sehairi, F. Chouireb, and J. Meunier, "Comparative study of motion detection methods for video surveillance systems," *Journal of Electronic Imaging*, vol. 26, no. 2, p. 023025, April 2017.

[8] A. Shimada, H. Nagahara, and R.-I. Taniguchi, "Background modeling based on bidirectional analysis," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 2013, pp. 1979-1986.

[9] S. A. Medjahed, "A comparative study of feature extraction methods in images classification," *International Journal of Image, Graphics and Signal Processing*, vol. 7, no. 3, p. 16, Feb. 2015.

[10] U. Rajanna, A. Erol, and G. Bebis, "A comparative study on feature extraction for fingerprint classification and performance improvements using rank-level fusion," *Pattern Analysis and Applications*, vol. 13, no. 3, pp. 263-272, Aug. 2010.

[11] M. Favorskaya, D. Pyankov, and A. Popov, "Motion estimations based on invariant moments for frames interpolation in stereovision," *Procedia Computer Science*, vol. 22, pp. 1102-1111, 2013.

[12] P. Bharathi and P. Subashini, "Optimization of image processing techniques using neural networks–A review," *WSEAS Transactions on Information Science and Applications*, vol. 8, no. 8, pp. 300-328, Aug. 2011.

[13] Y. Huang, "Advances in artificial neural networks–methodological development and application," *Algorithms*, vol. 2, no. 3, pp. 973-1007, Aug. 2009.

[14] M. Yazdi and T. Bouwmans, "New trends on moving object detection in video images captured by a moving camera: A survey," *Computer Science Review*, vol. 28, pp. 157-177, 2018.

[15] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.(CVR 1999)*, 1999, pp. 246-252.

[16] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. ICPR '04 the Pattern Recognition*, 2004, pp. 28-31.

[17] D.-S. Lee, "Effective Gaussian mixture learning for video background subtraction," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 5, pp. 827-832, May 2005.

[18] P. L. M. Bouttefroy, A. Bouzerdoum, S. L. Phung, and A. Beghdadi, "On the analysis of background subtraction techniques using Gaussian mixture models," in *Proc. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, June 2010, pp. 4042-4045.

[19] Q. Zang and R. Klette, "Parameter analysis for mixture of gaussians model," *Communication and Information Technology Research CITR*, The University of Auckland, New Zealand, Report 188, 2006.

[20] D. Tripathy and K. G. R. Reddy, "Adaptive threshold background subtraction for detecting moving object on conveyor belt," *Intl. Journal of Indestructible Mathematics and Computing*, vol. 1, no. 1, pp. 41-46, Jan. 2017.

[21] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features," *Computer Vision and Image Understanding*, vol. 95, no. 2, pp. 238-259, Aug. 2004.

[22] A. M. Hamad and N. Tsumura, "Background updating and shadow detection based on spatial, color, and texture information of detected objects," *Optical Review*, vol. 19, no. 3, pp. 182-197, May 2012.

[23] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, "Improving shadow suppression in moving object detection with HSV color information," in *Proc. IEEE Intelligent Transportation Systems*, Aug. 2001, pp. 334-339.

[24] P. Gorur and B. Amrutur, "Speeded up Gaussian mixture model algorithm for background subtraction," in *Proc. 2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2011, pp. 386-391.

[25] L. Wang, Y. Zhang, and J. Feng, "On the Euclidean distance of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1334-1339, Aug. 2005.

[26] D. ping Tian, "A review on image feature extraction and representation techniques," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 8, no. 4, pp. 385-396, July 2013.

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

[28] J. Flusser, "On the independence of rotation moment invariants," *Pattern Recognition*, vol. 33, no. 9, pp. 1405-1410, Sep. 2000.

[29] F. Al-Azzo, A. M. Taqi, and M. Milanova, "3D human action recognition using Hu moment invariants and Euclidean distance classifier," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 4, pp. 13-21, May 2017.

[30] Z. Huang and J. Leng, "Analysis of Hu's moment invariants on image scaling and rotation," *Computer Engineering and Technology*, pp. V7-476-V7-480, April 2010.

[31] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning Internal Representations by Error Propagation*, California Univ San Diego La Jolla Inst for Cognitive Science, pp. 1-33, 1985.

[32] Y. Xu, F. Liang, G. Zhang, and H. Xu, "Image intelligent detection based on the Gabor wavelet and the neural network," *Symmetry*, vol. 8, no. 11, pp. 130, Nov. 2016.

[33] B. Xu, H. Zhang, Z. Wang, H. Wang, and Y. Zhang, "Model and algorithm of BP neural network based on expanded multichain quantum optimization," *Mathematical Problems in Engineering*, Oct. 2015.

[34] S. C. Joshi and A. Cheeran, "MATLAB based back-propagation neural network for automatic speech recognition," *Int. J. Adv. Res. Electr. Electron. Instrum. Eng*, vol. 3, no. 7, pp. 10498-10504, July 2014.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).

Bhupendra Kumar Yadav is from Nepal. He is graduated with bachelor in mechanical engineering, and now he is currently pursuing master's degree in mechanical design in Tongji University Shanghai, China. His research interest projects are image processing and machine learning.

Jian Xiaogang is from China. He is graduated with PhD and currently working as an associate professor in mechanical engineering department in Tongji University Shanghai, China. His main research interests are image processing and machine learning, fatigue damage and life prediction, innovative design of construction machinery.