Histogram Specification and Kalman Filter with Euclidian distance Technique for identification of Human in the video Surveillance systems

Raviprakash M L, C S Pillai
Assistant Professor, Department of CSE, KIT, Tiptur
Professor, Department of CSE, ACS College of Engineering, Bangalore

Abstract— In Real time video Surveillance systems, the primary and most significant step is target object detection. Video Surveillance systems provides us continuous monitoring of the objects for the enhancement of security and control. Even though there is lot of progress in this field, human detection still remains one of the challenging tasks.This paper presents novel approach for detection of human using Histogram Specification with the Kalman filtering approach. The main objective of this paper is providing precise identification of human and estimation of their location from an unknown scene. Whenever the human is detected from extracted frames of the input video the background subtraction will be applied with Blob Analysis technique for modelling then the morphological operations are applied to remove the noise in the frame by choosing the size and shape of the neighborhood frame, is applied to find the frequency components of a signal buried in noise, Finally Kalman filter technique is applied where different positions of humans are identified based on the Calculation of Euclidian distance using the correlation matrix. The results of this work are drawn in the MATLAB tool by considering the input video dataset taken from various sources and extracting the frames from the input video for the detection then the efficiency of the proposed techniques will be measured.

Keywords— Histogram Specification, Background Subtraction, Fourier transformation; Kalman Filter, Morphological operation, Euclidian Distance.

1. INTRODUCTION

With increasing concerns for public safety and security nowadays, the need for automatic surveillance systems based on real time videos of public places is realized. The highly crowded and sensitive places like markets, shopping space, famous restaurants, railway stations etc. must be equipped with these surveillance systems. Without constraining their application to security under public spaces, they are also most demanded for the purposes of traffic control and examination, activity recognition and tracking, fault detection in industrial applications and semantic video indexing. In order to achieve the high level tasks of classification or tracking a target from video stream, the strategy used is to detect the target of interest in individual frames in the first place.

The technique used in many of the works for surveillance is the Background subtraction method. This technique performs extraction of the foreground object i.e. the target under motion by separating the background and foreground pixels in the frame under processing. The advantages of this method namely the performance in the presence of a non-mobile video camera and illumination invariance are well exploited by many researchers. The building of a background model of the captured video frame is a key point to be considered.

Video surveillance is a trending research topic within automation domain. The prime purpose of surveillance of video footages is identification and subsequent tracking of human motion possibly for monitoring suspicious activity, gait analysis etc. This paper aims to continually track multiple human targets through stationary cameras with the possibility of each human occluding the other in the scene while in motion. There exist numerous methodologies and frameworks in computer vision literature for video surveillance at present, still there is a lot of scope and many issues need to be overcome to devise smart video surveillance systems. The influence of lightning variations, interfering noise in the capturing device and medium, occlusions including inter-object and object-background cases etc provide a scope for lot of improvement. Human motion tracking in videos also is very popular in crime detection, forensics, crowd density estimation, biometric security and so on. All these applications involve recognition of the target object in each video frame and its continuous tracking across the video length to produce the trajectory.

Mean shift methodology, Camshaft methodology, Kalman filter [3], Foreground segmentation [2] are some of the various strategies at present being utilized for trailing the target object within frames of the video. Mean shift methodology is associate degree reiterative method within which the mean value of pixels over the image is compared with the color histogram of the target object and the tracking window is shifted iteratively to cover the target object.

The Camshift methodology [1], an alternative algorithm for tracking moving targets is a modified, adaptive form of mean shift algorithm. The sole distinction lies in the size of scan window being fixed in case of Mean shift whereas in Camshift, the scan window size is adapted dynamically based on the target motion. But, the fastest and easiest object recognition technique is the background subtraction, which detects the target object by separating the foreground pixels from background non-target pixels and this is employed in our paper.

The background subtraction methodology [5] [6] is very popular and straightforward technique with less computational load that achieves detection of every mobile target in the video frames which can be further tracked using Kalman Filter. It predicts the next state of the target given the current state of the target. In this paper, detection of human in frames and tracking its motion are discussed along with the simulation results.

1.1 Tracking human beings in videos

Human tracking is implemented on a video sequence, which involves the analysis of the presence, size, shape and position of human beings. Human tracking is generally used in applications like video surveillance, crowd monitoring. Videos are series of images, called frames, which are displayed in faster frequency, such that human eyes will not be able to percept the continuous contents. Usually, the information of the contents present in two consecutive frames will be closely related. The first step is to identify the regions of interest. A general human detection algorithm is desired, but sometimes it will be difficult to handle the humans properly with considerable amount of variations. Hence, many practical systems prefer stationary camera, which makes human detection procedure more straightforward.

2. Literature Review

In this section literature review for the research work is carried out where various

2.1 Afef Salhi et.al in this paper authors present an implementation of a human identification system in the video Surveillance systems. This identification of human in the video is an important task in many of the computer vision applications. The main steps in video analysis are two: detection of interesting moving objects and tracking of such objects from frame to frame. In a similar vein, most tracking algorithms use pre-specified methods for preprocessing. In our work, we have implemented several object tracking algorithms (Meanshift, Camshift, Kalman filter) with different preprocessing methods

2.2 Super pixel tracking using Kalman Filter [16]: In this paper, the authors present an algorithm for Human identification in video streams. To locate the target object approximately, Kalman filter is applied. The real time finding of target location surrounding the predicted location is achieved by employing super pixel based tracking methodology. The principle used here is to design a Kalman filter with an assumption that constant acceleration is maintained throughout the video sequence. The dynamics of motion and mechanics are applicable. The advantage of the Kalman filter lies in handling occlusions in the video captured. In the case of long time occlusion appearance too, this algorithm tracks the target well.

2.3 Current Advancements in Feature Extraction and Description Algorithms [17]: This paper describes the modern advances in feature detection and description algorithms developed for video surveillance purposes. The fundamental concepts have been discussed initially. A comparison of these algorithms for their functional abilities, performance, advantages and disadvantages has been taken up. The Maximally Stable External Regions algorithm and the Scale Invariant Feature Transform algorithms are used for reporting the algorithmic complexity and derivatives**.**

2.4 Dariusz et.al [18]: In this paper, a "Smart Monitor" system has been proposed with the system architecture, methods employed and algorithms. Experimental findings and results for tracking the video data is summarized. Monitoring the routine of ill persons is the central objective.

2.5 Systems for face detection and tracking [19] Human detection and tracking can be accomplished by different algorithms as well as specially designed architectures exploiting silhouettes of human body. The occlusions in the frames of video pose a difficulty in detection and tracking human. During occlusion, a part of the target object is hidden by the occluding object for a time either totally or partially. The features of the target object thus cannot be extracted in that case continuously. Re-identification of the target in multiple cameras, or during total or partial occlusion is to be ensured. The structure and shape of the target persons cannot be recognized for that duration.

The linking of a priori information for an object to be tracked in multi-Human identification system is very significant unpredictable trajectory is presented. It employs the omega-shaped descriptor. To locate the next position of the moving person, the methodology uses particle filter system along with linear filter.

The error between consecutive frames of the video incurred while applying particle filter is handled by Viola-Jones and HOG based SVM classifier. The filter fails to detect and track the person momentarily if it is occluded or collided by neighboring objects. The target objects are found in the form of moving blobs. Color based histogram helps in establishing relationship between lost and reappeared target.

3. Proposed Methodology

The proposed methodology of this research work is implemented in MATLAB tool where the video input is processed to extract individually each of the frames for human object detection in need to background segmentation to be applied. The main objective of this research work is frame-wise tracking of the target in presence object-object occlusions as well as object-background occlusions till the target mobility is observed in the video.

The procedure for video processing normally consists of the following steps-

Step1: Firstly, reading the input video data; a continuous signal to extract the individual frames it is comprised of is performed.

Step2: The input video is usually color image in RGB color space. It is easy to apply image processing techniques.

Step3: Thus, next step is to convert each RGB video frame into its grayscale form.

Step4: Blob analysis technique involving background subtraction can be applied with Thresholding and morphological operations as sub-steps.

Step 5: In the next step tracking using Kalman filter is done which predicts the next state of the target's movement, given its previous state.

Step 6: The task of background subtraction is a challenge as it is subject to variations in the lighting conditions or occlusions in the scene.

Step 7: Morphological processing is carried out on the target detected frames where a number of operations perform noise removal.

Step 8: As a last step, Fourier transform is applied on the morphologically processed frame to extract the lost frequencies of the target video frame in the noise signal so that no data is compromised.

3.1 System Architecture

The proposed system architecture shown in the figure 1, describes the how the human can be identified from input video dataset. The algorithm measures the detection of human. The given input video will be is initialized by reading the video file using video reader command and preprocessing of the video will be applied. Morphological operations are applied to these frames which involve noise removal and filling holes.

To carry out this filling operation, identifying the connected components in the video frames is necessary. These connected components are defined as the blobs which are identified by using the adjacency of the neighbor frame pixels that are similar in nature. The present location of the identification stage is predicted by the features of the human being.

Centroid is defined as the average of all the pixels in the image. It is output in the form of a set of x and y coordinates. In Matlab, the centroid of objects in an image can be computed together and the output is given as a matrix of size M×2, where M is the number of rows that correspond to the blobs of objects in the image and the two columns are the [x, y] coordinates of that object's centroid. For illustration, if two objects are detected in an image in the form of blobs, then the centroid output will be as below-

$$[x1\ y1;\ x2\ y2]$$

The first row contains centroid of first blob [x1, y1] and second row is the centroid of the other blob [x2, y2]. The boundary box position gets shifted based on the location prediction and then it will be updated. The calculation of centroid itself is used to locate the position of boundary box around the moving object blob, which undergoes updating with every frame. The boundary box has three attributes – its location coordinates, height and width. The image containing blobs will have as many boundary boxes as the number of blobs itself. So, each of these can be accessed through the N×4 matrix denoted as [x y width height]. Each row corresponds to the boundary box of one particular blob in the image with the columns representing its location and size.

Next step is tracking the detected objects using Kalman filter. Here, the boundary box is updated frame-by-frame as the blobs go on moving in the image. Initially the centroid of object blobs is recorded. As the video progresses, the previous identified scene is deleted because it will be replaced by the new values of the human detection; once the previous track is deleted the new scene track will be created using the Kalman filtering approach. This loads all the information about the human identification in the array. The distance between the connected components is calculated using the correlation matrix by using Euclidian distance which is mainly used in connecting the identified components. The Euclidean distance is calculated to find the minimum distance between the centroid vector of one object with x and y coordinates of the object's centroid and a collection of column vectors corresponding to centroid of other objects. The algorithm produces the output using the below equation-

$$\sqrt{(p_1-q_1)^2+(p_2-q_2)^2+\cdots+(p_n-q_n)^2}$$

In one dimension, the Euclidean distance between a pair of points, x1 and x2, lying on the same line is the absolute difference between the two points as:

$$\sqrt{(X_2-X_1)^2}=|X_2-X_1|$$

In two dimensions, the distance between P = (p1, p2) and q = (q1, q2) as:

$$\sqrt{(p_1-q_1)^2+(p_2-q_2)^2} \quad = \sqrt{\sum_{i=1}^{n}(p_i-q_i)^2}$$

Finally the accuracy can be determined by generating the graph based on the number of frames vs. number of frame detected for the identification based on human features.

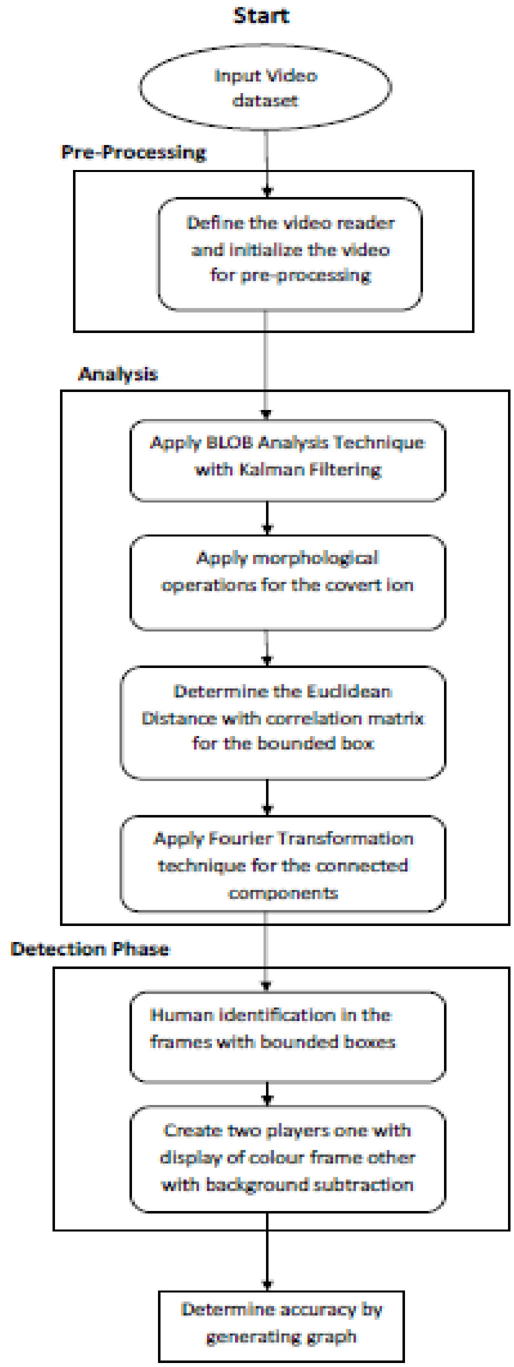

Figure 1: Proposed system Architecture

3.2 Pre-processing and Background Subtraction Technique

Video is described as a sequence of frames closely related to each other but defined by specific time instants. The video processing is performed by carrying out the various processing operations on individual frames or images that make up the video. Thus, first step involves conversion of video signal into consecutive frames. The pre-processing stage involves filtering the frame using one of the filters like mean filter, convolution filter and median filter.

The Background subtraction also known as foreground detection is a technique to distinguish the target object pixels that form the foreground scene of an image or frame leaving the rest of the image as background. There are many ways to extract the initial background model of the image. The first frame of the video with no presence of any of the foreground objects itself can be considered as the background model. Next, each frame in sequence is compared with the background frame and the foreground pixels are extracted that belong to the target.

$$P[F(f)] = P[I(f)] - P[D] \ldots\ldots (1)$$

$$|P[B(f)]| - |P[B(f+1)]| > \text{Threshold value} \ldots (2)$$

P [B (f)] is the background video frame, P [I (f)] is input frame at each iteration and P [D] is fixed reference background video frame. P [B (f+1) is the next input video frame.

3.3 Histogram Specification with the Kalman filtering

The Histogram Specification is mainly applied for multiple human identification and Blob detection is carried out using Background modelling. The blob detection is mainly used to carry out the filing operation which is mainly needed to identify the connected components in the video frames.

These connected components are defined as the blobs which are identified by using the adjacency of the neighbor frame pixels that are similar in nature. The present location of the identification stage is predicted by the features of the human being. These types of features are mainly obtained using the centroid of the blobs for the each blob that should be tracked.

The common plan of background segmentation is to involuntarily produce a binary mask that splits the cluster of pixels into 2 sets, foreground and background pixels. A stationary background frame are often matched to current frame uncomplicated cases. Pixels with high pixels area unit are known as foreground. This straightforward technique may add few specialised situations. Each constituent is characterised by its intensity within the RGB house. Next is Kalman Filtering technique is applied, which is primarily based technique to seek out the regions of the article in forthcoming frame. Centre of object is set 1st then Kalman filtering is formed use for predicting position of a similar within the next frame. Kalman filtering is most popular to cipher the state of a linear system. Kalman filtering consists of 2 steps, prediction and correction.

It provides optimum estimate to come up with the position for the motion model for a moving object though the video contains some quantity of dynamic noise and uproarious observations concerning the position at on every occasion step. For Gaussian noises, Kalman filtering can give an optimum resolution. The filter reduces the mean sq. errors of the parameters like position, velocity, etc. The Kalman filter is essentially an internet method, which implies the new observations are going to be processed as and once they arrive. A distinct time dynamic linear system having AN additive racket that may model surprising disturbances are going to be needed to formulate a Kalman filter is the problem.

The Kalman filter approach describes the creation of new object tracks for n number of frames x ∈ Rn of a discrete-time controlled process which are functioned by

$$x_k = Ax_{k-1} + Bu_k + w_{k-1} \quad \text{..... (1)}$$

With a measurement m y ∈Rn that is

$$y_k = Hx_k + v_k \quad \text{........ (2)}$$

The random variables w k and k v represent the process and measurement noise respectively. The new tracks are identified based on the equation (3) and (4)

$$p(w) \approx N(0, Q) \quad \text{....... (3)}$$

$$p(v) \approx N(0, R) \quad \text{....... (4)}$$

The process noise covariance Q and measurement noise R covariance matrices might change with each frame measurement, however here we assume they are constant.

4. Results and Discussion

The results of the proposed approach are drawn in the MATLAB tool which provides us the effective solution for Multiple Human detection in the video. Even though there are various works has been carried out so far the proposed research work strands unique with its desire to provide an excellent results for the different kinds of the videos considered. In this research work first the BLOB analysis is applied then applied the Kalman filtering along with Euclidean distance provides more effective results with minute number of false detections.

The main advantages of this proposed work is provide an efficient algorithm which is easy to implement and contains less complexity in detecting the human presents in the video.

The first input video under the consideration is a good quality video of the duration 15 seconds and the memory size is 25MB. The video taken is from jewelry shop which consists of 3 persons with sitting in the shop watching the jewels. The identification of human result its grey scale image representation of a frame of video is as shown in figure 2 with the background subtraction. The figure 2 describes the identification of persons based on the hand movements. Figure 3 describes the human identification based on the face and the leg movements of the person.

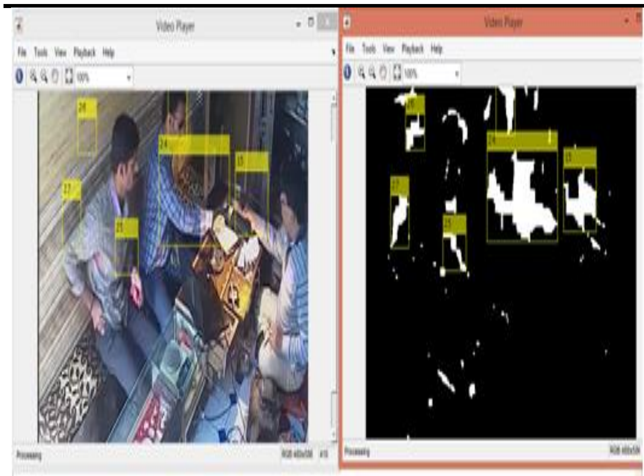

Figure 2: Human identification based on Hand Movements

Figure 3: Human identification based on Face and the legs of the human

The accuracy of the proposed technique is analyzed with total number of percentage of frames detected with human vs. the total number of frames. Here graph is generated by taking four input video from various sources with different sizes and different background, for each video we could able to get an accuracy of 90%. Where frames for each video vary from 100 to 1000 and accuracy for each video varies from 80% to 95% in detecting the human.

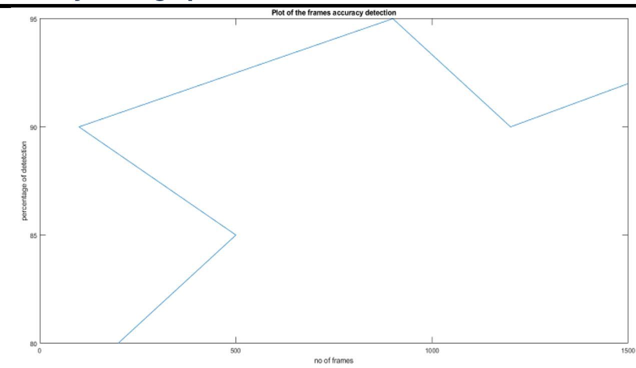

Figure 4: Accuracy of the proposed number of frames vs. percentage of detection

The graph plot(X, Y) creates a two dimensional representation of the data with number of frames on the horizontal axis against percentage of object blob detections made on the vertical axis.

- When X and Y are vectors, they both must be equal in length. The graph is plotted as Y versus X.
- When X and Y are matrices, they must have same dimensions. The graph in this case is plotted with columns of second matrix Y versus columns of first matrix X.
- In case one among X, Y being vector and other being a matrix, then the size of the matrix is restricted so that either the row or column count must be same as the length of the vector. If the number of rows in the matrix is equal to the length of the vector, then the graph is plotted with matrix column on x-axis against the vector on y-axis.
- Number of frames detected is the matrix that has been taken from the blob analysis and kalman filter that gives the percentage of frames that detected the blob.

5. Conclusion

This paper concludes the human identification task with Histogram Specification and Kalman filtering with Euclidian distance measure. For human identification process, initially the pre-processing techniques are carried out on the input frames. The pre-processing steps remove the noise and enhance the quality of input video data for detection of target. Background subtraction is performed to eliminate the background unmatched objects. Background subtraction or foreground detection separates out the target pixels of interest from unwanted background pixels while the target objects are moving on the foreground. Then the Histogram Specification with the Kalman filtering approach is applied, where Blob Analysis technique for modelling then followed by Kalman-filter technique is applied where different positions of humans are identified based on the Calculation of Euclidian distance calculation using the correlation matrix. In future the scope is to implement the Human identification for theft activity tracking using multiple cameras to monitor the footage for suspicious criminal or illegal activity and sending alarms in real time.

References

[1] Afef Salhi and Ameni Yengui Jammoussi , “Object tracking system using Camshift, Meanshift and Kalman filter”,W orld Academy of Science, Engineering and Technology, 2012

[2] Alok K. Watve ,Indian Institue of Technology, Kharagpur , seminar on“Object tracking in video scenes”, 2005.

[3] Amir Sala pour and Arezoo Salarpour and Mahmoud Fathi and mirhossein Dezfulian , “Vehicle tracking using Kalman filter and features”, Signal & Image Processing : An International Journal (SIPIJ) Vol.2, No.2, June 2011.

[4] C. Lakshmi Devasena, R. Revathi, “ Video surveillance system-A survey”, IJCSI International journal of computer science Issues, vol 8, issue 4, no.1, July 2011

[5] Flavio B. Vidal and Victor H. Casanova Alcalde (2010). “Object Visual Tracking Using Window-Matching Techniques and Kalman Filtering”, Kalman Filter, Vedran Nordic (Ed.), ISBN: 978-953-307-094-0.

[6] rite P. Kuralkar, Prof. V.T.Gaikwad, “Human Object Tracking using Background Subtraction and Shadow Removal Techniques”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 3, March 2012.

[7] P. Perez,C. Hue, J.Vermaak1, and M. Gangnet, “Color-Based Probabilistic Tracking” Proc. 7th Eur. Conf. Compute. Vis., 2002, pp. 661–675.

[8] A.Mittal and L.S.Davis, “A multi-view approach to segmenting and tracking people in a cluttered scene” int.J. Comput. Vis., vol. 51, no. 3, pp. 189–203, 2003.

[9] Zoran Zivkovic, Ferdinand van der Heijden, “Efficient adaptive density estimation per image pixel for the task of background

subtraction" Pattern Recognit. Lett. vol. 27, no. 7, pp. 773–780, May 2006.

[10] L. Wang, T. Tan, W. Hu, and H. Ning, "Automatic gait recognition based on statistical shape analysis," IEEE Trans. Image Process., vol.12, no. 9, pp. 1120–1131, Sep.2003.

[11] R. Tanawongsuwan and A. Bobick, "Modelling the effects of walking speed on appearance-based gait recognition," in Proc. 2004 IEEE Computer Society Conf. Computer Vision and Pattern Recognition, 2004, vol. 2, no. 2, pp. 783–792.

[12] H. Fujiyoshi, A. Lipton, and T. Kanade, "Real-time human motion analysis by image skeletonization," IEICE Trans. Inf. Syst., vol. 87 no. 1, pp. 113–120, 2004.

[13] A. Senior, A. Hampapur, Y. L. Tian, L. Brown, S. Pankanti, and R. Bolle, "Appearance models for occlusion handling," Image Vision Comput., vol. 24, no. 11, pp. 1233–1243, 2006.

[14] R. Collins, R. Gross, and J. Shi, "Silhouette-based human identification from body shape and gait," in Proc. Int. Conf. Automatic Face and Gesture Recognition, Washington, DC, 2002, pp. 366–371.

[15] Shuai Zhang, Member, Chong Wang, Member, Shing-Chow Chan, Xiguang Wei, and Check-Hei Ho "New Object Detection, Tracking, and Recognition Approaches for Video Surveillance Over Camera Network" IEEE sensors journal, vol. 15, no. 5, may 2015.

[16] Mohammad Faghihi, Mehran Yazdi , Sara Dianat Superpixel tracking using Kalman filter "3rd International Conference on Pattern Recognition and Image Analysis (IPRIA)",2017

[17] Ehab Salahat, Member, IEEE, and Murad Qasaimeh, Member, IEEE "Recent Advances in Features Extraction and Description Algorithms: A Comprehensive Survey", 2017

[18] Dariusz Frejlichowski, Katarzyna Gosciewska, Paweł Forczmanski Radosław Hofman "SmartMonitor" — An Intelligent Security System for the Protection of Individuals and Small Properties with the Possibility of Home Automation" Sensors 2014, 14, 9922-9948; doi:10.3390/s140609922

[19] Cancela, B., Ortega, M., Penedo, M.: Multiple human tracking system for unpredictable trajectories. Machine Vision and Applications, 25(2), 511-527, 2014.

[20] Tathe, S., Narote, S.: Real-time human detection and tracking. 2013 Annual IEEE India Conference (INDICON), pp. 1-5, 2013.

[21] Kushwaha, A., Sharma, C., Khare, M., Srivastava, R., Khare, A.: Automatic multiple human detection and tracking for visual surveillance system. 2012 International Conference on Informatics, Electronics Vision (ICIEV), pp. 326-331.