PAKISTAN'S MULTI DISCIPLINARY JOURNAL FOR ARTS & SCIENCE https://pmdjas.com

ISSN 2789-2611

| December. 2020, VOL. 1 (Issue, 03) Page. 40 – 56 |

MULTIPLE HUMAN TRACKING IN MULTIPLE CAMERA DISTRIBUTED ENVIRONMENT

Mehwash Razzaq[1] Niaz Muhammad[1]

[1]Department of Computer Science and Information Technology, University of Balochistan Quetta

Received on: December 20/2020. Published on: December: 31/2020.

Abstract

We propose a human tacking method using multiple-viewpoint images and videos. In visionbased human tracking, human-human occlusions and self occlusions are a part of the more significant problems. Employing multiple viewpoints and a viewpoint selection technique, however, can minimum these problems. The vision system in this case should select the best viewpoints for extracting human tracking information; the" best" selections can be replaced through different kinds of target information. We route the issue of tracking human. We divide the task into two primitive sub-tasks (human tracking and object color matching). Each subtask has a different criterion for selecting viewpoints and a valuation output of one sub-task can help another sub-task. We describe the criteria for accomplishing the individual sub-tasks and the relationships between sub-tasks. We have built an experimental based system on a small number of reliable image features and performed fundamental observation on the key point selection approach.

KEYWORDS: Camera, Human Tracking, Vision, Analysis, Digital, Multiple-viewpoint.

1. INTRODUCTION

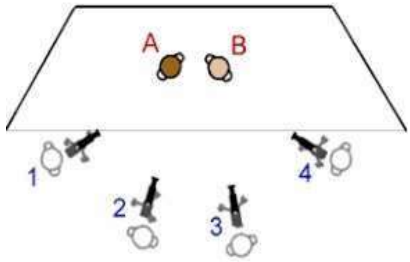
Tracking humans in multiple cameras is interest for a variety of applications such as surveillance, activity monitoring and gait analysis. With the limited field of view (FOV) of video cameras, it is necessary to use multiple, distributed cameras to completely monitor a site. Typically, surveillance applications have multiple video feeds offered to a human spectator for analysis. However, the capability of humans to focus on multiple videos simultaneously is partial. Therefore, there has been an interest in developing computer vision methods that can analyze data from multiple cameras concurrently and possibly present it in a compact symbolic fashion to the user.

The growing demand of multiple camera systems in some applications such as public security, transportation control, military, urban planning, and business information has involved increasing attention, and a large number of networked multiple camera methods are getting installed in public places, for instance, airports, subways, railway stations, highways, parking lots, banks, schools, shopping malls, and military areas. These multiple camera systems not only effectively protect the security of public facilities and citizens, but also seamlessly help to transform to smart city. However, human tracking is the most challenging since human may differ greatly in look on account of changes in brightness and viewpoint, background mess, occlusion, non-rigid deformations, intra-class variability in shape and pose. Human tracking includes human tracking with in a camera and human tracking across multiple cameras. When a person enters into the field of view (FOV) of a camera, human tracking within a camera is needed. However, when he/she leaves the FOV, the human information is no longer available, thus the limited FOV of a camera cannot meet the needs of wide-area human tracking. In order to widen the FOV, human tracking across multiple cameras has to be used since video streams across multiple cameras covering a wider range of areas, which helps to analyze global activities in the real world.

To cover an area of interest, it is reasonable to use cameras with overlapping FOVs. Overlapping FOVs are normally used in computer vision for the purpose of extracting 3D information. The use of overlapping FOVs, however, creates an uncertainty in monitoring people. A single person present in the region of overlap will be seen in multiple camera observations. There is need to identify the multiple projections of this person as the same 3D object, and to label them regularly across cameras for security and monitoring application.

In the non-overlapping camera views' scenario, there is not a common FOV area between two cameras' views, every camera's view is completely disjointed, and human cannot be seen in the so-called blind area Compared with human tracking across overlapping cameras, human tracking across nonoverlapping cameras will be more challenging and practical. As a result, human tracking over camera networks is necessary and quite challenging in the intelligent video surveillance

Fig. 1. Multiple human tracking.



2 RELATED WORKS

In related work, M. Yokoyama and T. Poggio, (2005) presents an approach of dealing with the hand off problem based on 3Denvironment model and calibrated cameras. The 3D coordinates of the person are established using the calibration information to find the location of the person in the environment model. At the time of hand off, only the 3D voxel-occupancy information is compared to achieve hand off, because multiple views of the same person will map to the same voxel in 3D.

In, D.-N. Ta, W.-C. Chen, et. al., (2009). only relative calibration between cameras is used, and the correspondence is established using a set of feature points in a Bayesian probability framework. The intensity features used are taken from the center line of the upper body in each projection to reduce the difference between perspectives. Geometric features such as the height of the person are also used. The system is able to predict when a person is about the exit the current view and picks the best next view for Tracking.

A different approach is described in H. Zhou, et. al., (2009). that does not require calibrated cameras. The camera calibration information is recovered by observing motion trajectories in the scene. The motion trajectories in different views are randomly matched against one another and plane homographs computed for each match. The correct homograph is the one that is statistically most frequent, because even though there are more incorrect homographs than the correct one, they lie in scattered orientations. Once describe approaches which try to establish time correspondences between non-overlapping FOVs. The idea there is not to completely cover the area of interest, but to have motion constrained along a few paths, and to correspond objects based on time from one camera to another. Typical applications are cameras installed at intervals along a corridor or on a freeway Y. Wei and L. Tao, (2010) and D. Comaniciu, et. at., (2003).

The luxury of calibrated cameras or environment models is not available in most situations. We therefore tend to prefer approaches that can discover a sufficient amount of information about the environment to solve the hand-off problem. We contend that camera calibration is unnecessary and an overkill for this problem, since the only place where handoff is required is when a person enters or leaves the FOV of any camera. By building a model of the relationship between FOV lines of various cameras can provide us sufficient information to solve the hand-off problem.

In S. Nigam and A. Khare, (2012) Researcher proposed an appearance-based technique for tracking multiple people's motions by an asynchronous multiple camera system. In the proposed method, dynamically modeled the head appearance of every person is using multiple camera-based remarks. Observed color information and the associated reliability features are kept in the head model as a set of color squares concentrated in 3-D space. Authors applied the proposed technique to a human communication scene which have minor inter-person space where the motion-based method not pretty good work due to occlusions. The matching steps between the produced appearance and the input image can be correctly executed by reliability value of each patch.

In P. Viola and M. Jones, (2001), presented an approach for object detection which minimizes computation time while achieving high detection accuracy. The approach was used to construct a face detection system which is approximately 15 fasters than any previous approach. This paper brings together new algorithms, representations, and insights which are quite generic and may have broader application in computer vision and image processing.

In S. Kushwaha, et. at., (2012) Detect the human in the video and then track human in the subsequent video frames. for training purpose, they used binary adaptive boosting and for tacking particular filter has been used but this purposed method is more efficient than A swine, Ashoke and Rama method. J. Wang, et. at., (2009) proposed a technique for multi-object tracking in individual frames and then linking direction across frames, authors used givens steps in proposed method: particle filtering, linking short tracks generated using Kaman filtering, or via greedy Dynamic Programming in which paths are predictable one after another and shortest path algorithm is used for multiple objects tracking optimization complexity, proposed method deal with two main problems. The 1st is time-independent detection, in which a prediction structure assumes the number and positions of objects from the existing signal at every time step individually and the second depend on modeling detection errors and object motions to link detections into the most likely trajectories in a multiple-object tracking domain.

In H. Liu, et. at., (2009). proposed a method which detect and track Multiple Human in Dynamic Environment, authors have used two type of systems 1. laser based system, 2. Vision-based system to detect object and identify its human or not. Bayes filter, Particle filter, Ad boost algorithms were used to track the human in indoor environment.

The J. Liu and X. Zhou,(2019) proposed a method for multi-commodity network flow for tracking multiple people which purpose was solve the problem of multiple human which paths will intersect continuously for the long period of time. Authors used minimum cost as maximum flow for the multi object tracking in the proposed method which steps are given below

1) Recursively track from frame to frame.

2) Dynamic programming for global trajectory optimization.

3) Frame by frame difference is solved using K-shortest path.

In J. Segen, (1996) describes a method for real-time tracking of human in video sequences. The input to the method is live or recorded video data of camera acquired by a monocular camera in an environment where the primary moving objects are people. The output consists of trajectories which give the s spatiotemporal coordinates of individual persons as they move in the environment. The system uses a new model-based approach to object tracking. It identifies feature points in each video frame, matches feature points across frames to produce feature" paths", then groups short-lived and partially overlapping feature paths into longer living trajectories representing motion of individual persons. The path grouping is based on a novel model-based algorithm for motion clustering. The system runs on an SGI Indy's workstation at an average rate of 14 frames a second.

The Q. Cai, *et. at.*, (1995) presents an approach to tracking human motion in a sequence of monocular images. The process consists of detecting motion, segmenting moving subjects by recovering the background and, finally, tracking the subject of interest. The usual assumptions of small image motion, a fixed viewing system and constant velocity are systematically relaxed. Two cases are studied:(1) a viewing system with negligible motion, and (2) a viewing system with non-negligible motion.

C. Rasmussen and G. D. Hager, (2001) propose to track complex visual objects based on the JPDAF algorithm, where a related technique called Joint Likelihood Filter (JLF), i.e., relating the exclusion principle at the heart of the JPDAF to the method of masking out image data, is used to deal with occlusions between tracked objects. However, this method calls for very high computational requirements with the number of associated objects increasing. To take full advantage of more available information to further improve the tracking performance.

D. Schulz, *et. at.,* (2003) propose sample-based JPDAF for tracking multiple moving human objects using a mobile robot, where the JPDAF algorithm is directly applied to the sample sets of the individual particle filter to determine the correspondence between the individual object and measurement. Moreover, the proposed approach adopts different features extracted from consecutive sensor measurements to explicitly deal with occlusions.

The S.-K. Weng, *et. at.,* (2006) propose a real-time and robust human tracking algorithm in a real-world environment, such as occlusion, lighting changes, fast moving human object, etc., based on adaptive KF, which allows the parameter estimations of KF to adjust automatically.

In order to track the objects more effectively, some constraints among kernels need be considered in the MKT. While for the discriminative trackers, all the human locations in each video frame are first obtained through a human detection algorithm M. Paul, (2013) and then the tracker jointly establishes these human objects' correspondences across frames through a target association technique.

In W. Li and X. Wang, (2013) propose a new approach for matching images observed in different camera views with complex cross-view transforms and apply it to person re identification. It jointly partitions the image spaces of two camera views into different configurations according to the similarity of cross-view transforms. The visual features of an image pair from different views are first locally aligned by being projected to a common feature space and then matched with softly assigned metrics which are locally optimized.

In Y. Yang, (2014) propose a new approach for Color naming, which relates colors with color names, can help people with a semantic analysis of images in many computer vision applications. In this paper, we propose a novel salient color names-based color descriptor (SCNCD) to describe colors. SCNCD utilizes salient color names to guarantee that a higher probability will be assigned to the color name which is nearer to the color. Based on SCNCD, color distributions over color names in different color spaces are then obtained and fused to generate a feature representation.

The S. Liao, et. al., (2015) propose an effective feature representation of human appearance called Local Maximal Occurrence (LOMO) for human re-id, where the LOMO analyzes local color and texture features' horizontal occurrence and maximizes the occurrence so as to obtain a robust feature representation against viewpoint changes, based on HSV color histogram and scale invariant local ternary pattern (SILTP) texture descriptor.

S. Wu, et. at., (2016) propose Feature Fusion Net (FFN) to describe human appearance for human reid, where the FFN combines conventional neural network (CNN) deep feature with handcrafted features, including color histogram computed in five different color spaces, i.e., RGB, HSV, YCbCr, Lab and YIQ, and Gabor texture descriptors with multi-scale and multi-orientation. The CNN deep feature is constrained by the handcrafted features through back propagation to form a more discriminative feature fusion deep neural network.

Automated human tracking over camera networks L. Hou, et. at., (2017) is getting essential for video surveillance. The tasks of tracking human over camera networks are not only inherently challenging due to changing human appearance, but also have enormous potentials for a wide range of practical applications, ranging from security surveillance to retail and health care. In this paper human tracking over camera networks two important functional modules for the human tracking over camera networks are addressed, including human tracking within a camera and human tracking across non-overlapping cameras.

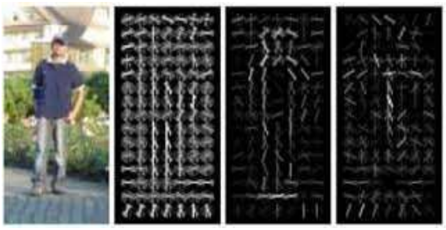
3. METHODOLOGY

The proposed method purposes to extract and track multiple humans from video streams. Two important issues of system performance are Efficiency and accuracy. In the following we define the laboring features including our proposed oriented Patterns feature and the HOG feature. This section also describes the descriptor which we deployed in the sliding window based system.

A. Histogram of Oriented Gradients (HOG)

Histogram of Oriented Gradients (HOG) Another feature for silhouette and edge information encoding, called Histogram of Orientation Gradients, is proposed by Dalal and Triggs in Dalal (2005) for people detection.

Fig. 2. Multiple human tracking using HoG



The feature extraction is more complex than in Edge Orientation Histograms, increasing the discriminative power of the descriptor while ensuring a certain degree of invariance. As described in Dalal, (2005), steps:

1) A global image normalization equalization, using a gamma compression, is performed to reduce the effects of local shadowing and influence of illumination variation effects.

2) Computation of first order image gradients.

3) The image window is divided into small spatial regions, called “cells,” and a local 1D histogram of edge orientations with K orientation bins over all the pixels in the cell is accumulated. Each edge pixel contributes to each orientation bin with a value proportional to the magnitude of its orientation.

4) A normalization step is carried out by accumulating a measure of local histogram “energy” over local groups of cells called “blocks”. Each cell is normalized with respect to the block which it belongs.

5) The final HOG descriptor of the whole detection window is obtained by concatenating all HOG descriptors of all blocks of a dense overlapping grid. The HOG feature extraction is depicted in Figure 2.4 taken from [Dalal 2005]. Four variants of the HOG descriptor have been presented by the authors. The difference between them lies in the shape of considered cells. These four variants are: Rectangular HOG(R-HOG), which is the original one, Circular HOG(C-HOG) where the cells are defined into grids of log-polar shape. Bar HOG where the descriptors are computed similar to the RHOG, but use oriented second derivative filters rather than first derivatives and Center-Surround HOG which use a center-surround style cell normalization scheme. Many other approaches for people detection, using HOG descriptors, have been proposed

B. Support Vector Machine SVM

The algorithm is composed of HOG feature extraction of both positive samples (target object) and background samples, these features are used to train SVM and a trained SVM classifies the samples in every frame so that the coordinates of the detected object is updated. The tracker is coupled with the RGB Mean filter for the detection of object when the SVM prediction scores go low.

After every 20 frames, the algorithm trains the SVM using new background samples so that the tracking capability increases and makes the algorithm adaptive. In each frame, localized search around region of interest is done for detections which significantly reduce the CPU time. During processing of every frame, the region of interest is updated using the previous detections. Fig.1 illustrates the working of the proposed method. SVM Classifier is trained online using the samples taken from previous detections by sliding a window. The older samples are always preserved and produced in the training phase to avoid the probability of bias in classification.

A. Target Representation and Tracking Target representation is essential during object tracking. Object features like contours, corners, global statistics descriptors and so on are used for representation. The HOG descriptor has advantages when compared to other descriptors. Its performance will not

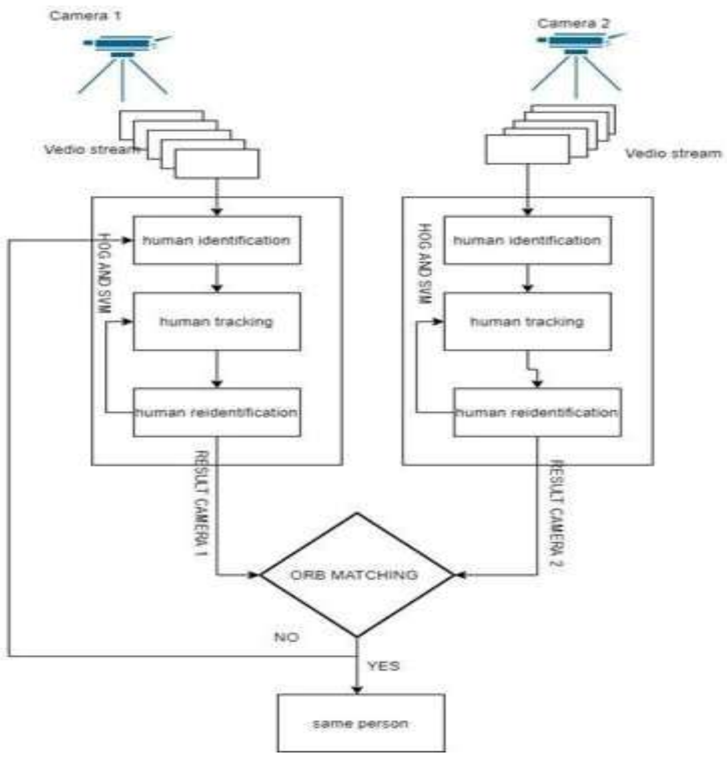


Fig. 3. The proposed model flow diagram.

get degrade when there is a change in objects illumination and shadowing and also it is invariance to geometric and photometric transformations but not for object orientation. The HOG and linear SVM method of object tracking is adopted which provides better result. SVM is a statistical learning technique which is used to find the best score in binary classification case. Practically SVM trains using a smaller number of samples and the time of training is less when compared to many other classifiers. The training of the SVM is done using the HOG features and the approach is composed of several steps: HOG features are extracted for positive training image as well as a random number of negative training samples are also considered. With this training set, the SVM classifier training is done which is according to the SVM light technique defined in. Since SVM light produces the support vectors vi and the corresponding i, and the weight vector w can be calculated as shown in (1)

$$W!‘ =!‘V`1 \qquad (1)$$

The training samples are taken by sliding window over the previous frame based on the previous detection. In a frame t1, a rectangular area which includes target object is taken and put into a subset S1 and treated as of positive samples. The remaining portion of the frame is considered as background and treated as negative samples. The dimensions of the rectangular area in the current frame t2 are the same as the previous window, and the assumption is stated as there is no considerable size change in tracked object in subsequent frames. After every t20 frames, background samples are collected into another subset S2 of samples. This subset includes old samples in the training; thus, it avoids over fitting.

B. RGB Mean Filter Color-based representation of objects is one of the successful approaches for object tracking which is found in the literature. When the object is moving away from the camera or the prediction scores of the SVM classifier is going low (in our case ¡ -0.9), the tracking results, points to False positives. In order to improve the tracking accuracy, in this case, we introduced RGB mean filter, which calculates the mean of individual color channels (R,G,B) of the positive sample taken in the first frame and the samples in the region of interest and small portion around the region of interest. The minimum value generated by subtracting the mean of positive sample and the mean of samples which is under test determines the True Positives. The minimum mean calculation is shown using (2) and (3)

$$= mean\ (Rp)\ -mean(R!\ `) = mean(Gp)\ -mean(G!`) \quad (2)$$

$$= mean\ (Bp)\ -mean(B!`)Min = r + g + b \quad (3)$$

Where r, g, b, is the difference of the mean of Red channel, green channel and blue channel values from positive sample and the sample under test respectively. Rp, Gp and Bp are the red, green and blue channel values of positive sample. Rt , Gt and Bt are the red, green and blue channel values of test samples in the current frame. Min () should be a minimum value which indicates both the samples are similar.

C. *ORB MATCHING*

In this section, we briefly introduce the ORB descriptor.

The ORB is a very fast binary descriptor which combines the FAST key point detector and the recently-developed BRIEF descriptor That’s the reason why it’s called ORB, the abbreviation of oriented FAST and rotated BRIEF.

A. FAST key point Orientation by Intensity Centroid FAST is a high-speed corner key points detector and widely used because of its efficient computation. But it does not have multiscale and orientation components. a scale pyramid scheme is employed to produce multiscale features. Since FAST has strong responses along edges, Harris corner filter is used to reject edges. There are many methods to describe the orientation of a key points, such as histogram of gradient and block patterns. However, most of them need large computation cost. Intensity centroid, a simple but effective measure of corner orientation, is adopted to Describe the orientation of FAST key points. The intensity centroid assumes that the centroid of corner’s intensity is offset from its center, and this shift vector can be used to calculate an orientation. The moments of a patch are defined as tracking sensor via an extended Kalman filter. Detailed simulation results demonstrate that the tracking algorithm substantially reduces the relative position estimation error introduced by noisy color region tracking.

The algorithm thus enables target pursuit based on an extremely noisy but simple and low cost sensor, a monocular camera with color region tracking.

4. EXPERIMENTS AND RESULTS

The experimental result of the proposed method is discussed in this section. The pet’s data set, INRIA person data set and real world environment video is given as an input to the system. Frames are generated, features are extracted and human are recognized. The output video detects maximum fifteen to sixteen human objects and track them shown in ***table 1.***

Table I: Multiple Human Tracking Model Results

video	Total human	True Positive	True Negative
1	6	4	2
2	10	5	2
3	21	16	5
4	18	9	9

Human tracking across non-overlapping cameras establishes detected/tracked human objects' correspondence between two non-overlapping cameras so as to successfully perform label hand off. Based on the approaches used for target matching, human tracking across cameras can be divided into two main categories, human re-id, HOG tracking, and ORB color matching. We first tested the Hog and SVM over those images. The model predicted 91 images correctly and made incorrect prediction for 9 images. The accuracy rate for the base line model was 91.44% whereas validation accuracy of 89.11%.

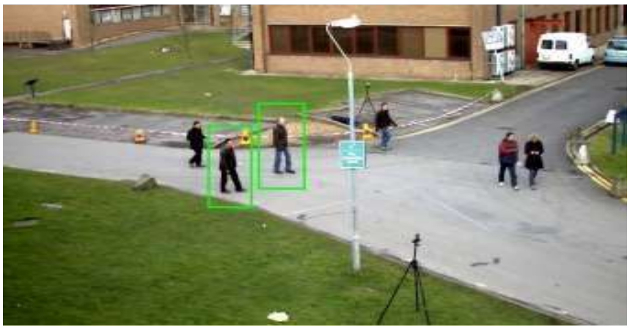


Fig 4. Frame 2

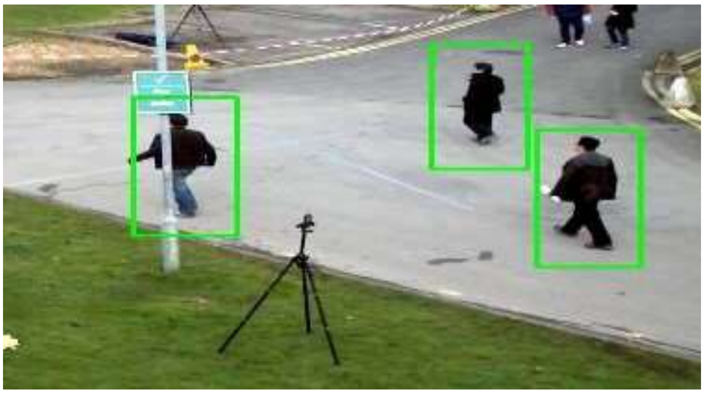
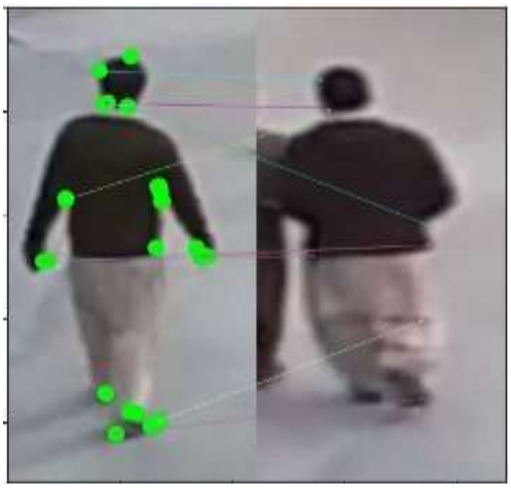


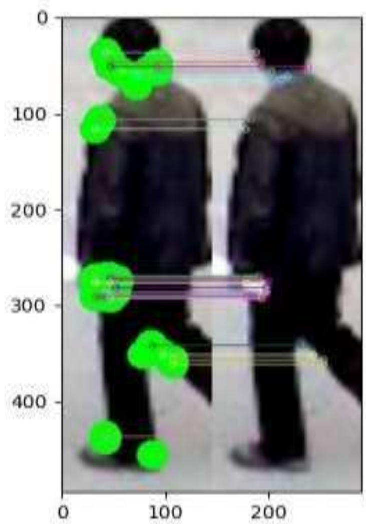
Fig 5. Frame 17

Fig 6. Frame 44

In our research, while developing the proposed method, we come across a model. We considered this model as second baseline model and name it (HOG AVG) Model. The reason to add this ((HOG AVG)) model to test vigorously the robustness of the proposed method. In the proposed model, we have used HoG and SVM for training and testing purpose for human tracking and re identification, first we used 2 data set for training and testing purpose we get different results in experiment 2 we apply orb matching for multiple cameras out we used first experiment output as input in 2nd experiment. Their 91 images from camera 1 and 75 image frame camera 2 and apply orb matching for same person detection its predicate 61 images correctly and 14 image incorrect. Finally, we tested another dataset of hundred images on our customized model that predicted 92 images correctly out of hundred images. The results are shown in the figure and loss of the model was 0.02%, validation loss of 0.03% and accuracy of 93.89% and validation accuracy of 94.03%on thirty iterations,



(a)



(b)

Fig. 7. frame 9 dataset of Peshawar University.

A. *Performance evaluation*

In the experiment, we collected the data set of various human and carry experiments to test the robustness of the proposed method. We collected the various images, videos of human online and real world applied the proposed method. The method correctly tracking the human in different view using the different cameras shown in graph 1.

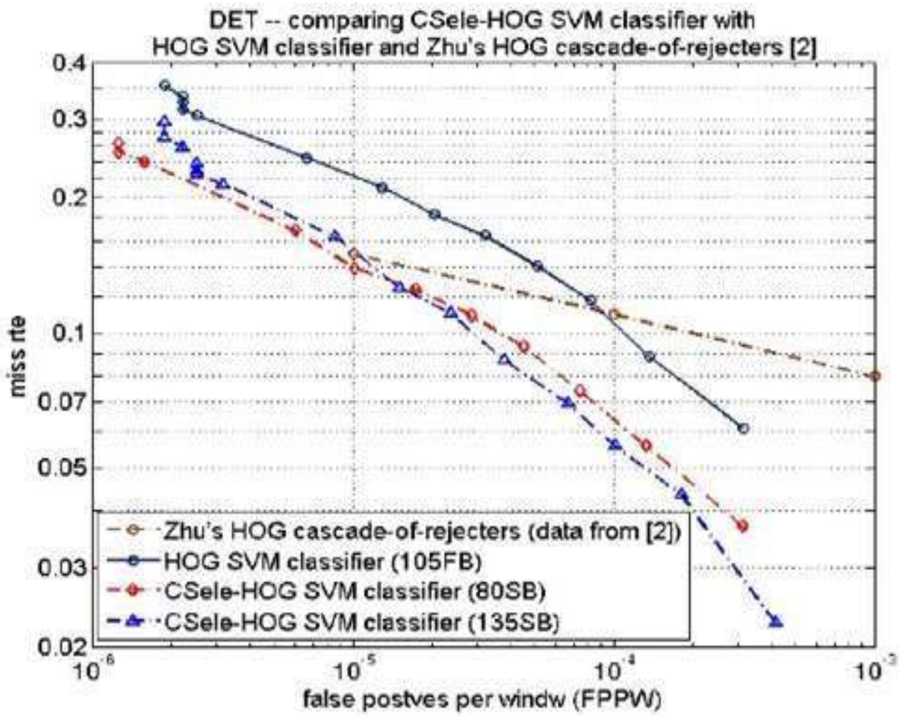


Fig. 8. proposed algorithm comparison

B. *Accuracy test*

We also carried experiments to test the accuracy of the proposed method. We tested over the multiple different video streams and applied different techniques to compare proposed method shown in table 1.

5. CONCLUSION

and Multiple human tracking in non over lapping field of view of cameras is an extremely useful technology for surveillance applications. However, to our knowledge there is no existing efficient ground-based system able to track and pursue multiple humans autonomously in complex real-world environment.

Table II. Multiple human tracking model results

Frame	total human	True Positive	True Negative
1	1	1	0
16	7	5	2
24	21	15	6
32	35	28	7
40	45	30	15
63	45	25	20
71	45	41	4
90	30	21	9

In this paper, we proposed a visual-based model to detect and track the multiple humans in the image scene and real-world environment. We first locate various possible technique to track the human and re-identification of human. We used the feature base method to find the possible object. We used histogram of Oriented Gradient (HoG) for locating the human in the image scene. Later, we used orb algorithm for the human recognition and tracking. Which is a faster algorithm for tracking human.

6. ACKNOWLEDGMENTS

We are thankful to government of Pakistan for supporting and funding Mehwash Razzzaq research work.

7. REFERENCES

"Automatic multiple human detection and tracking for visual surveillance system," in 13, no. 5, p. 1729881416657746, 2016.

2012 International Conference on Informatics, Electronics & Vision (ICIEV). IEEE, 2012, pp. 326–331. 271–276, 2005.

A. K. S. Kushwaha, C. M. Sharma, M. Khare, R. K. Srivastava, and A. Khare,

C. Rasmussen and G. D. Hager, "Probabilistic data association methods for tracking complex visual objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 560–576, 2001.

D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 25, no. 5, pp. 564–577, 2003.

D. Schulz, W. Burgard, D. Fox, and A. B. Cremers, "People tracking with mobile robots using sample-based joint probabilistic data association filters," *The International Journal of Robotics Research*, vol. 22, no. 2, pp. 99–116, 2003.

D.-N. Ta, W.-C. Chen, N. Gelfand, and K. Pulli, "Surftrac: Efficient tracking and continuous object recognition using local feature descriptors," in *2009 IEEE*

Conference on Computer Vision and Pattern Recognition. IEEE, 2009, pp. 2937–2944.

H. Liu, J. Luo, P. Wu, S. Xie, and H. Li, "People detection and tracking using rgb-d cameras for mobile robots," *International Journal of Advanced Robotic Systems*, vol.

H. Zhou, Y. Yuan, and C. Shi, "Object tracking using sift features and mean shift," *Computer vision and image understanding*, vol. 113, no. 3, pp. 345–352, 2009.

IEEE, 1995, pp. 215–218.

J. Liu and X. Zhou, "Observability quantification of public transportation systems with heterogeneous data sources: An information-space projection approach based on discretized space-time network flow models," *Transportation Research Part B:*

J. Segen, "A camera-based system for tracking people in real time," in *Proceedings of 13th International Conference on Pattern Recognition*, vol. 3. IEEE, 1996, pp. 63–67.

J. Wang, Y. Ma, C. Li, H. Wang, and J. Liu, "An efficient multiobject tracking method using multiple particle filters," in *2009 WRI World Congress on Computer Science andviola2001rapid Information Engineering*, vol. 6. IEEE, 2009, pp. 568–572.

L. Hou, W. Wan, J.-N. Hwang, R. Muhammad, M. Yang, and K. Han, "Human tracking over camera networks: a review," *EURASIP Journal on Advances in Signal Processing*, vol. 2017, no. 1, pp. 1–20, 2017.

M. Paul, S. M. Haque, and S. Chakraborty, "Human detection in surveillance videos and its applications-a review," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1, pp. 1–16, 2013.

M. Yokoyama and T. Poggio, "A contour-based moving object detection and tracking," *Methodological*, vol. 128, pp. 302–323, 2019.

P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1. IEEE, 2001, pp. I–I.

Q. Cai, A. Mitiche, and J. K. Aggarwal, "Tracking human motion in an indoor environment," in *Proceedings., International Conference on Image Processing*, vol. 1.

S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2197–2206.

S. Nigam and A. Khare, "Curvelet transform-based technique for tracking of moving objects," *IET computer Vision*, vol. 6, no. 3, pp. 231–251, 2012.

S. Wu, Y.-C. Chen, X. Li, A.-C. Wu, J.-J. You, and W.-S. Zheng, "An enhanced deep feature representation for person re-identification," in *2016 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2016, pp. 1–8.

S.-K. Weng, C.-M. Kuo, and S.-K. Tu, "Video object tracking using adaptive kalman filter," *Journal of Visual Communication and Image Representation*, vol. 17, no. 6, pp. 1190–1208, 2006. *Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp.

W. Li and X. Wang, "Locally aligned feature transforms across views," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3594–3601.

Y. Wei and L. Tao, "Efficient histogram-based sliding window," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 3003–3010.

Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li, "Salient color names for person re-identification," in *European conference on computer vision*. Springer, 2014, pp. 536–551.