Multiple People Tracking Using Non-iterative Tensor Voting Method

Shipra Ojha, Dr. S.R. Sakhare

Abstract— Human tracking is an emerging research field in computer vision and video surveillance. Lightning changes and occlusion are the main challenges faced during tracking. This paper focuses on people tracking both in indoor and outdoor scenes. Our method is based on color modeling using tensor voting framework, which can handle these challenges. The lightning conditions are controlled by homomorphic filtering and occlusion is handled by modelling the color clothing of each individual. Tensor voting is non-iterative method in which only the scale has to be specified at beginning and is able to build number of clusters automatically. The experimental work shows that our method can track the person successfully and also handle occlusion by finding the centroid of a person.

Index Terms— Background subtraction, Color modeling, Filtering, Tensor voting, Tracking, Video surveillance, 2D Gaussian.

1 INTRODUCTION

A video is represented as smallest unit of hierarchical unit such as scene, shot and frame. We can use single or multiple cameras to shot a video frame. A video is divided into frames that represent continuous actions, a camera is a device that stores and records the scene, popularly known as sensing element. Manual object detection and tracking becomes tedious task due to number of factors such as humans cannot keep a continuous watch on object, chances of manual errors, etc. which results in automation requirement [1]. Tracking means following a particular object that is moving to find trajectories of an image. A tracker assigns labels to the object so that it is easily recognized in different frames. Depending upon which method or algorithm is used, a tracker provides information such as area, orientation, shape and size of an object. The ultimate objective of tracking is to connect the target object into continuous frames in order to track. The enhancement in tracking is motivated by many real-time applications such as security purpose, human-computer interaction, sports, tracking people in public places, etc.

Image processing is a research area which deals with issues like filtering, compression, enhancement of images, etc. It makes a study of description of image and video sequence. Computer vision is an area that deals with acquiring, analyzing and understanding behaviour of images. It is an artificial intelligence that extracts useful information from raw data. In order to understand image behaviour, study of various factors like camera model, lightning, color, texture, shape and motion taken from video are important [1]. Our research topic will overcome manual limitation to keep a continuous watch on an object. Tracking strategy can be completed in three stages: object detection, object recognition and object tracking. Detection means to find out the object in motion e.g. we are interested in human or car or animal tracking. Different techniques are available for object detection viz. background subtraction, temporal differencing and optical flow method of which first method is most widely used. In background subtraction [2] we subtract the background and extract only foreground moving object. In this method the object that is in motion is recognized as foreground object and static image is the background frame. As shown in Fig. 1 the background image is subtracted and foreground human image is extracted. Temporal differencing [3] is based on frame differencing and optical flow [4] is a vector-based method that matches the point on objects over multiple frames. Objects can be classified using shape-based, color-based, motion-based and texture-based classification as suggested in [5]. The next stage is the object recognition that states the entity of interest to track. That means which moving object we need to track has to be recognized from multiple objects. And finally the last stage is to track the target object.

Fig. 1. Input frame and foreground extracted image

In this paper our main focus is to track a person using a single camera, and a static background image. The video captured contains impurities like noise, lightning changes and fluctuations. Various clues such as size, shape, color, intensity can be used to differentiate between persons. Multiple methodologies discussed in literature survey section for tracking person, hand motion of people, vehicle in traffic monitoring, etc. There are many times when people cross each other resulting in partial or complete occlusion. Few methods discussed herewith that overcomes occlusion problem. Tensor voting is a framework used for color modelling of an image which is non-iterative method

- *Shipra Ojha is currently pursuing masters degree in Computer Science from Vishwakarma Institute of Information Technology in Pune, India E-mail: shipraojha4@gmail.com*
- *Dr. S.R. Sakhare, Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune, India E-mail: sakharesachin7@gmail.com*

ISSN 2229-5518

and requires no prior knowledge of data and thresholding. Here each input is encoded as tensor and voting is a field where tensors communicate with each other. After the voting process is done clusters are formed based on maximum color that gives maximum separation in between foreground and background. This color modelling is used for tracking purpose.

This paper focuses on object tracking in continuous video frames. This is organized as follows, section II focus on study of related work, section III describes tensor voting methodology, section IV covers the result set and discussion.

2 Related Work

In 2003 [6], a research continued in the field of automatic tracking and intruder-detection in a surveillance system. A tracking system includes background generation and subtraction method, can be affected by global intensity variations using a PTZ camera. Object merging and splitting can be handled by a method proposed in [7] by H. Galoogahi, where the main focus is on tracking a group by following events such as splitting, merging and leaving. Temporal differencing method is based on frame differencing to find out the target object. Due to lack of object trajectories, the fragmented trajectories cannot be combined into a real trajectory, to overcome this tensor voting method proposed in [8] provides efficient way to handle discontinuous trajectories. In [9] to automatically improve the image quality and enhance the binary image tensor voting framework is well suitable. The noisy input data contains some corrupted regions, these regions are restored and image quality is improved. Tracking of moving objects using a static camera, with severe occlusion handling, merging/ splitting of objects can be tracked easily. To detect the moving object P. Kornprobst and G. Medioni use the partial differential equation (PDE) based approach [10]. Smoothness constraint is used along with temporal dimension to correct detection defects due to occlusion and noisy data. Kalman filter [11] works well for ballistic trajectory but tensor voting works for generic approach and solves large class of tracking problems. As explained in [12] Modified Kalman filter method is used to track multiple similar objects such as vehicles using a moving camera. The M2 Tracker [13] is multiple-camera based human tracking system, uses presence probability and color features to classify pixels between objects. Object tracking can also be done using kernel based method [14] instead of using a single kernel. It uses multiple kernels to provide robust tracking performance, to make possible multiple kernels work well projected gradient method is used.

3 Tensor Voting Methodology

The main aim is to track the objects in motion. To detect the moving regions PDE method [15] is used that restores the binary image and maintains the sequence. In order to correct the defects in occlusion and noise it uses smoothness constraint with temporal dimension. This is done using tensor voting approach which is a perceptual grouping approach.

The previous work over inference is broadly classified into curves, regions and surfaces, these inputs are suitable for areas such as dot clustering in 2-D, curve and contour interface in 2-D, and surface reconstruction in 3-D [16]. Guy and Medioni [17] developed a non-iterative method that is able to handle both curve inference from 2-D and surface inference from 3-D form a point. Unlike the previous work this method is robust to noise, no limitation over surface topology, detects orientation discontinuity and is non-iterative. Tensor voting method is based on tensoral description of moving regions that encodes the aspects like: the number of neighbours involved in this estimation, the velocity and their coherence (called the saliency) [10], [15].

An input is in the form of tokens, a second order, symmetric, non-negative definite tensor is an nXn matrix and an ellipsoid in an N-D space [18]. Types of tensor are: Stick tensor or oriented tensor, Ball tensor or un-oriented tensor and Generic tensor. Tensors can be produced by the summation of eigenvectors $\overrightarrow{n}\,\overrightarrow{n}^{T}$. The tensors can be computed as [18]:

$$T = \sum_{i=1}^{d} \overrightarrow{n}\,\overrightarrow{n}^{T} \qquad (1)$$

Where di→ dimensionality

$\overrightarrow{ni}$ → unit vector

The methodology is grounded for two elements: tensor calculus for data representation and voting means communication. The only parameter that needs to be set is scale range where voting process has to be carried out.

The activity diagram in Fig. 2 explains color modeling process starting from background removal till color clustering.

4 Experimental Results

The experiment was performed over several videos taken from different datasets such as PETS 2009, Visor, etc. we found that our algorithm works good for both indoor and outdoor scene's. Once at the beginning we need to set a static camera and fix its position. Then we have to set a background/location and capture the video, at first we need to manually mark the person whom we have to track and then our tracking process is automatically done. Our first step is to extract the foreground object that is any person in motion and subtract the background frame using appropriate thresholding usually in the range of 30-40. Then filtering is done to remove noisy and edge pixel using appropriate filtering method and finally color modeling is done using tensor voting. Herewith we represent a person using single color. Fig. 3 shows initial step of experimental study representing foreground extraction of multiple people.

Based on color clothing of individual each dominant color is modeled by 2D-Gaussian in an appropriate feature space. As tensor voting framework is robust to lightning variation, we use a mixture of R+G, G+B color feature and form clusters as proposed in [19]. Tokens communicate with each other under a predefined scale known as σ, usually it lies in the range of 6 to 12 to form color clusters as shown in Fig. 4(c). Using these clusters we form a 2D Gaussian and obtain the mean and covariance matrix in the feature space. For each cluster we compute the mean and covariance matrix as;

$$\mu_k = (1/Nk) \sum_{i=1}^{Nk} xi \qquad (2)$$

ISSN 2229-5518

$$C_k = (1/Nk)\sum_{i=1}^{Nk}(xi-\mu k)(xi-\mu k)^{T} \quad (3)$$

Where Nk is the number of feature vector in k^{th} cluster.

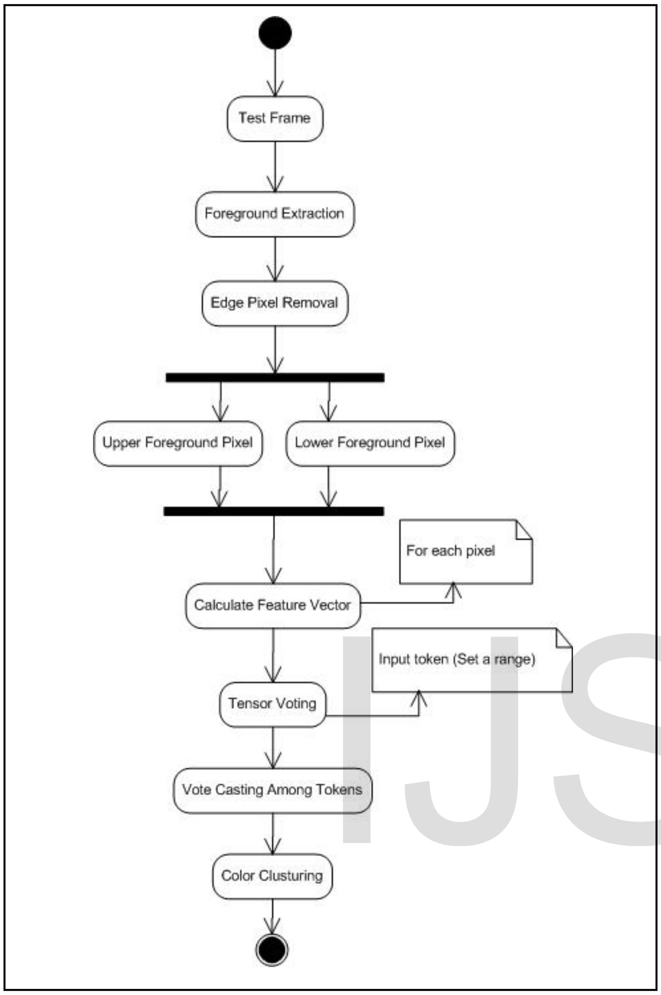

Fig. 2. Activity flow diagram for color modeling

Fig. 3. (a) Input frame from PETS-2009 dataset (b) Foreground extraction

Gaussian formation of color clusters is as shown in Fig. 5. Here x-axis indicates the combination R, G color and y-axis indicate combination of B, G color. Fig. 6 shows the result of tracking taken from Visor dataset. At first we model the person which is to be tracked, and then a single color is given for each individual, so that it is easily recognized and does not results in misclassification. It is seen that our approach effectively keeps track of a person even under occlusion. Our result successfully handles both partial and complete occlusion.

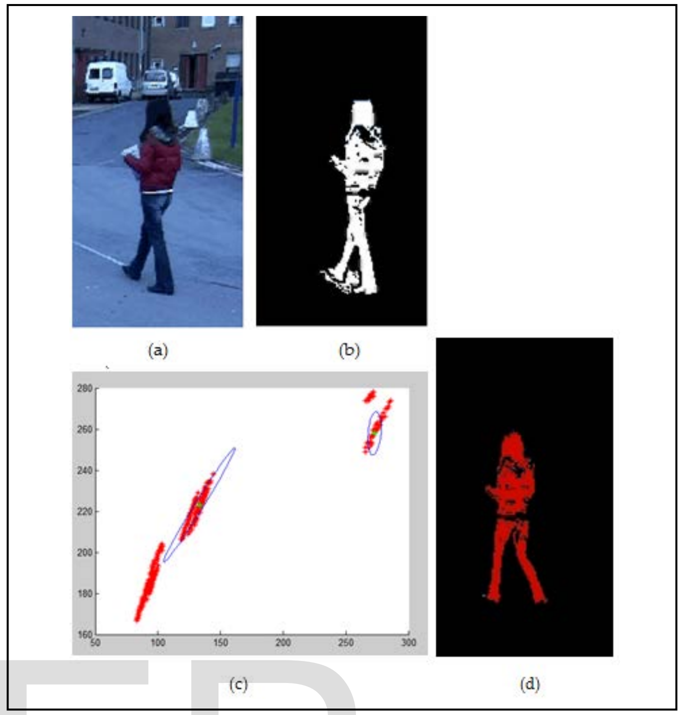

Fig. 4. (a) Input frame from PETS 2009 dataset, (b) foreground extracted image without noisy data (c) Color cluster formation (d) Color modelling

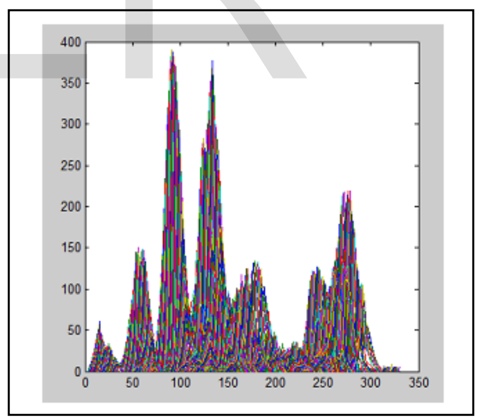

Fig. 5. Gaussian formation of color clusters

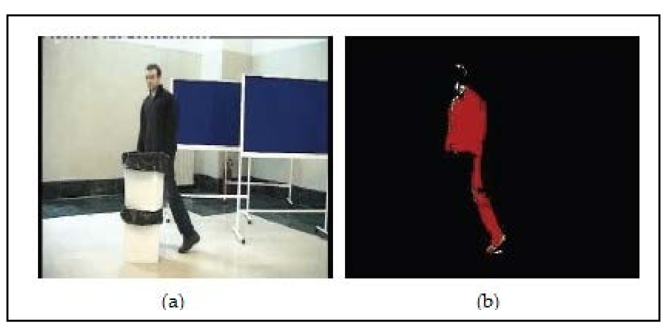

Fig. 6. Tracking- occlusion handling (a) Input frame (b) color modeling

Fig. 7(b) shows multiple people are labled under R, G, B color pixels. This helps in handling occlusion. As seen in the fig below blue and green labled persons occlude each other, still tracking continue successfully.

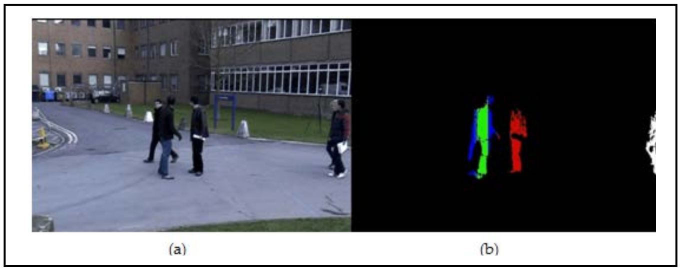

Fig. 7. People occluding each other (a) Input frame (b) color mode

Fig. 8 shows a series of actual input frame taken from video then foreground extracted image and finally person is modeled by color using tensor voting framework. To validate the performance of our method we performed our experiment on several videos that contains single or multiple persons. Table 1 shows the quantitative study of number of pixels assigned to person. Fig 8 (a, d) is the actual input image taken from video, Fig 8 (b, e) is the foreground extracted image and Fig 8 (c, f) is the color modeled image. Table 1 show that the difference between the actual foreground pixels and number of pixels modeled using color is less.

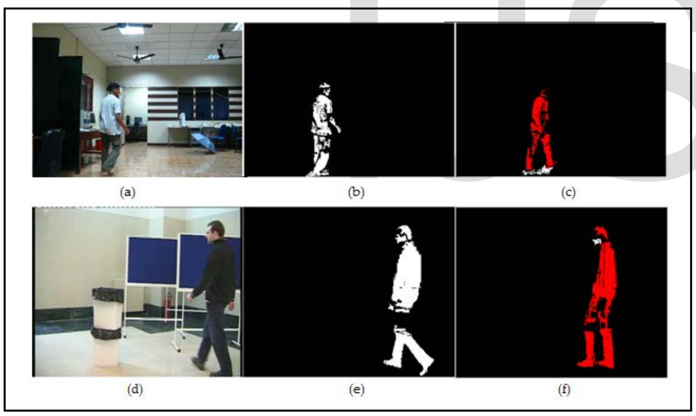

Fig. 8. (a, d) Input frames, (b, e) Foreground extracted image, (c, f) Color modeling

TABLE 1
QUANTITATIVE ANALYSIS OF COLOR MODELING

Frame	Foreground pixels	Color-modeled pixels
Fig. 4 (b, d)	12124	10519
Fig. 8 (b, c)	2129	1897
Fig. 8 (e, f)	4696	4560

5 Conclusion

We propose a robust method that can track a person using color modeling. For color modeling tensor voting method is used, each dominant color is modeled by a 2D Gaussian and color clusters are formed in feature space. Our approach is computationally efficient that runs on 30 frames per second. From the experimental work it is clear that our approach can handle occlusion and person is recognized by a single color. During experimentation we observe that classification accuracy is better for RGB color as compared to HSI color model. This algorithm is tested on several videos to evaluate the performance of our method.

Acknowledgment

I am especially grateful to my guide Dr. S. R. Sakhare who has provided guidance and encouragement. I would like to put forward my heartfelt acknowledgement to all my classmates, friends and all those who have directly or indirectly provided their overwhelming support during my project work and the development of this paper.

References

[1] B. Karasulu and S. Korukoglu, "Moving Object Detection and Tracking in Videos," *Performance Evaluation Software, SpringerBriefs in Computer Science*, 2013

[2] R.S. Rakibe and B.D. Patil, "Background Subtraction Algorithm Based Human Motion Detection," *Int. J. of Scientific and Research Publications*, vol. 3, no.5, 2013

[3] W. Shuigen, C. Zhen, L. Ming and Z. Liang, "An Improved Method of Motion Detection Based on Temporal Difference," *IEEE Int. Workshop on Intelligent Systems and Applications*, Wuhan, pp. 1-4, 2009

[4] C. Braillon, C. Pradalier, J.L. Crowley and C. Laugier, "Real-Time Moving Obstacle Detection Using Optical Flow Models," *IEEE Conf. on Intelligent Vehicles Symposium*, pp. 466-471, 2006

[5] S. Ojha and S.R. Sakhare, "Image Processing Techniques for Object Tracking in Video Surveillance- A Survey," *IEEE Int. Conf. on Pervasive Computing (ICPC)*, pp. 1-6, 2015

[6] S. Kang, J. Paik, A. Koschan, B. Abidi and M.A. Abidi, "Real-Time Video Tracking Using PTZ Cameras," *Proc. of SPIE 6th Int. Conf. on Quality Control by Artificial Vision*, vol. 5132, pp. 103-111, Gatlinburg, TN, 2003

[7] H.K. Galoogahi, "Tracking Groups of People in Presence of Occlusion," *IEEE, Fourth Pacific-Rim Symposium on Image and Video Technology*, Singapore, pp. 438-443, 2010

[8] J. Kang, I. Cohen and G. Medioni, "Continuous Multi-View Tracking Using Tensor Voting," *IEEE Workshop on Motion and Video Computing*, Florida, Dec.5-6, 2002

[9] T.D. Nguyen, J. Park, S. Kim, H. Park and G. Lee "Automatically Improving Image Quality Using Tensor Voting" *Springer, Neural Computing and Application*, 2010

[10] P. Kornprobst and G. Medioni, "A 2D+t Tensor Voting Based Approach for Tracking," *IEEE 15th Int. Conf. on Pattern Recognition*, vol. 3, pp. 1092-1095, 2000

[11] X. Li, K. Wang, W. Wang and Y. Li, "A Multiple Object Tracking Method Using Kalman Filter," *IEEE Int. Conf. on Information and Automation (ICIA)*, pp. 1862-1866, 2010

[12] C. Miller, B. Allik, M. Ilg and R. Zurakowski, "Kalman Filter-Based Tracking of Multiple Similar Objects From a Moving Camera Platform," *IEEE 51st Annual Conf. on Decision and Control (CDC)*, pp. 5679-5684, 2012

[13] A. Mittal and L.S. Davis,"M2tracker: A Multi-View Approach to Segmenting and Tracking People in a Crowded Scene," *Int. J. of Computer Vision*, vol. 51, pp. 189–203, 2003

[14] C.T. Chu, J.N. Hwang, H.I. Pai and K.M. Lan, "Tracking Human Under Occlusion Based on Adaptive Multiple Kernels with Projected Gradients," *IEEE*

ISSN 2229-5518
Trans. on Multimedia, vol. 15, no. 7, pp. 1602-1615, 2013

[15] P. Kornprobst and G. Medioni, "Tracking Segmented Objects Using Tensor Voting," *IEEE Conf. on Computer Vision and Pattern Recognision*, vol. 2, pp. 118-125, 2000

[16] G. Medioni, C.K. Tang and M.S. Lee, "Tensor Voting: Theory and Applications," *Int. J. of compu. Inf.*, Sci-5, pp 1-10

[17] G. Guy and G. Medioni, "Inference of Surfaces, 3D Curves, and Junctions from Sparse, Noisy, 3D Data," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 11, pp. 1265-1277, 1997

[18] P. Mordohai and G. Medioni, "Dimensionality Estimation, Manifold Learning and Function Approximation using Tensor Voting," *Journal of Machine Learning Research 11*, pp. 411-450, 2010

[19] M. Kulkarni and A.N. Rajagopalan, "Tensor Voting Based Foreground Object Extraction," *IEEE 3rd National Conf. on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, pp. 86-89, 2011.