Fast Mean Shift and Particle Filtering based Visual Human Tracking

Muhammad Farooq, Saeed Anwar and Sharif Khan
Faculty of Electronic Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology,
Topi,Swabi,23640,Pakistan

Abstract— **This paper presents a technique to detect and track multiple humans in a video sequence. Human detection is done using Fast Mean Shift approach. Multiple humans in a video sequence are differentiated on the basis of their color and appearance. The occlusion issue is also addressed in the paper and is removed using the particle filtering and appearance matching techniques. The technique presented for detecting and tracking are efficient and computationally inexpensive. Experimental work shows that in case of low image quality and occlusion, this approach works properly.**

I. INTRODUCTION

Human detection and tracking is a critical step in many computer vision applications such as surveillance, security systems and human computer interaction etc. In recent times human detection and tracking in a scene, has gained impetus and many algorithms have been developed, but the problem with most of these algorithms is that they are specialized i.e. they are limited in their scope and hence lack robustness e.g. in crowded scenes the large number of people close to each other forms a group which results in partial or complete occlusion [1] [2]. In such cases human detection and tracking becomes a complex task e.g. blob detection based algorithms [1] [2] segment groups as a single object due to the underlaying connected components labeling. The algorithms using segmentation on the basis of color only [3] do not provide good results since color is mostly not a good representative of individual humans. The algorithms employing silhouette analysis [2] [4] and stochastic segmentation from binary images [5] are good for detecting individuals in groups. In order to find some landmarks such as heads, these approaches need good motion segmentation quality. Methods based on the solution space of the possible human configurations [5] are computationally inefficient.

In this paper we propose a system to detect humans in a scene and to track their path. This method operates directly on the difference image obtained through the background subtraction. The difference image obtained after manipulation is a two dimensional probability distribution function, where the modes represent high probabilities of human presence. Fast mean shift approach is used to detect human and differentiate them from other objects. Tracking of humans is done using the color and appearance matching technique. The major issue related to human tracking is occlusion. Occlusion occurs when more than one human cross each other and exactly overlap each other. For occlusion removal particle filtering [15] is used. Particle filtering is recursive in nature and it involves two stages i.e. prediction and update. The flow of the technique developed is given in Fig.1.

II. HUMAN DETECTION

This stage involves Background subtraction and Fast Mean Shift approach.

A. Background Subtraction

The first step in human detection is the "Background subtraction" where the foreground image is obtained by subtracting the current image from a background image. A background image is the initial frame from the video which is used to obtain the foreground objects. This background image is subtracted from each and every frame of the video. A frame is nothing but a matrix containing mathematical values. Once the background matrix is subtracted, the resultant matrix is obtained and converted to a binary matrix having only 1s and 0s. 1s show the possible pixels having any object whereas 0s mean no pixel change. This binary matrix is achieved after thresholding foreground image. The Foreground image is obtained as follows.

$$ForegroundImag = CurrentImag - BackgroundImag \quad (1)$$

B. Mean Shift

Mean shift is a nonparametric statistical method which seeks the mode of a density distribution in an iterative procedure. It is popular and successful method for object tracking. Mean shift is an iterative process to find out a new location $x' = x + \Delta x$ from current location x. Δx and it is given as

$$\Delta x = \frac{\sum_{i=1}^{M} a_i w(a_i) k(\| \frac{a_i - k}{h} \|^2)}{\sum_{i=1}^{M} w(a_i) k(\| \frac{a_i - k}{h} \|^2)} - x \quad (2)$$

where M is the number of data points and h is the size of the kernel, $w(a_i)$ is the weight of each pixel a_i.

C. Fast Mean Shift

Human detection is performed by Fast Mean Shift procedure. The humans are detected by the subtraction of foreground image from the background image. The foreground image is a two dimensional probability density function

978-1-4244-4873-9/09/$25.00 ©2009 IEEE

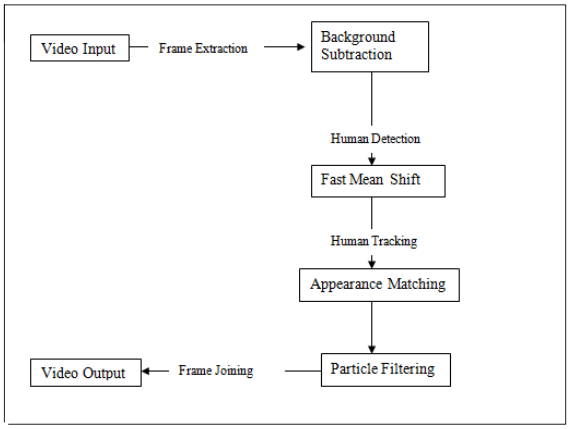

Fig. 1. Algorithm Flow

where high intensities imply high probability of moving objects. The number of high intensity peaks increases as the number of moving objects increases in the scene. Our aim is to find out individual high intensity peaks in order to identify humans in the scene. Human size model $H(y), W(y)$ is used for search. Where H is height and W is width. The information is obtained through calibration of the scene.

The main steps involved in Mean Shift are the same three steps as given in [6]

1) The difference image intensity maximum is mapped to unit intensity and its entire range is scaled proportionally.
2) A sample set of n points $X_1 ... X_n$ is defined by locating local maxima - above a very low threshold T_1 - in the difference image. The final result does not depend critically on T_1. A very low value just increases the run time and generates more outliers which can be eliminated during the mode tracking step.
3) The fast mean shift procedure is applied to the points of the sample set with a window size of $(H(X_i), W(X_i))$ according to the local size model. The mean shift procedure converges to the nearest mode typically within 3-4 iterations.

Zeroth and first moments are required for mean shift computation, and zeroth moment is considered as a probability density sampled at different location of x and y. The probability density is given as

$$p(x,y) = \frac{1}{W(y)H(y)} \sum_{i=1}^{n} I(x_i, y_i) \tag{3}$$

Where $p(x,y)$ are the points of convergence and when all the points are linked together, it forms the center of detected cluster. The process of linking is similar to [6] by merging all points in x and y direction which are near than the local window size $(W(y); H(y))$. Weighted mean of convergence points is used to find the center of the cluster.

The above method used for fast mean shift is slow and computing intensive. The fast variant is given by [7], by using integral image also known as summed area table [8]. Fast computation of the integral image can be performed in a single pass [9]. Using the integral image, the area sum of a rectangular region within the original image can be efficiently computed by the following step,

$$S_{area} = I_{int}(x-1, y-1) + I_{int}(x+W-1, y+H-1) \tag{4}$$
$$-I_{int}(x-1, y+H-1) - I_{int}(x+W-1, y-1) \tag{5}$$

where S_{area} is the sum of the rectangle starting with x,y at upper left corner having parameters of width(W) and height(H) and $I_{int}(x;y)$ is the integral image[7]. Mean shift vectors for a two dimensional probability density can be obtained by the following equations.

$$I^x_{int}(x,y) = \sum_{x' \leqslant x, y \leqslant y'} x' I(x', y'), \tag{6}$$

$$I^y_{int}(x,y) = \sum_{x' \leqslant x, y \leqslant y'} y' I(x', y'), \tag{7}$$

The mean shift vector $(m_x; m_y)$ computation using integral images and uniform kernel can be defined as

$$m_x = \frac{S_{area}(I^x_{int})}{S_{area}(I_{int})} - x \tag{8}$$

$$m_y = \frac{S_{area}(I^y_{int})}{S_{area}(I_{int})} - y \tag{9}$$

The above equations use the area sum computed over the window rectangle by using the integral images and their zeroth and first moment. This method is faster as it requires three arithmetic operations and four array accesses while this is not the case with the previous one, which requires nine arithmetic operations and twelve array accesses for a single pass.

1) Fast Mean Shift Implementation: In order to implement Fast Mean Shift Algorithm we used MATLAB [10]. We use a rectangular window of predefined height (H) and width (W). H and W are selected such that they represent the approximate height and width of humans as seen by the camera. Note that the values of H and W depend on the position of the camera. This rectangular window slides over the foreground image starting from the high value pixels. The center of this window keeps on changing until it converges to one point. After convergence the center exactly matches the center of human and the window encloses the human completely. The process is illustrated in Fig.2.

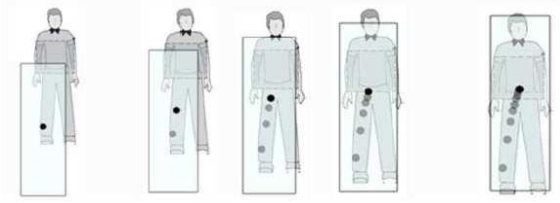

Fig. 2. Fast Mean Shift Implementation

III. Human Tracking

A. Appearance Matching

Once the human is detected, the next step is to track them. Tracking is done by appearance Matching. In this technique color information is extracted after human detection and then on the basis of this information the humans are compared. The comparison is done between current frame and the previous frame. A threshold value is used to compare two humans appearing between two successive frames. Appearance Matching technique is implemented in MATLAB [10]. The video camera used in our project produced 15 frames per second. In order to save the MATLAB computation time we only compared those humans which were very close to each other in a given frame. For instance in first frame three humans at three different locations are detected which are far from each other. Now in the succeeding frame humans will only be compared within a certain region of their previous location so as to save the time which could have otherwise been wasted in useless comparisons. In order to further enhance the efficiency of our system we reduced the frame rate by 50%. The speed was increased without any significant loss of information.

B. Particle Filtering

Variable of interest that evolves with time are tracked by particle filtering [11], typically with a non-Gaussian and potentially multimodal pdf. Entire pdf in particle filtering is represented in sample based form. A series of actions are taken each one modifying the state of variable of interest according to some model. Multiple copies of variable of interest are used, each one associated with a weight that signifies the quality of the specific particle. An estimate of variable of interest is obtained by weighted sum of all particles. Particle filtering is composed of two phases; prediction and update. Prediction phase uses a model in order to simulate the effects an action has on the set of particles with apposite noise added. Update phase uses the information from sensing device and update particle to get correct estimation of new pdf using Sampling-Importance Re-sampling (*SIR*) technique[12][13][14]. This process of particle filtering is repeated because of the its recursive nature. The basic idea of Particle filtering is shown in Fig.3.

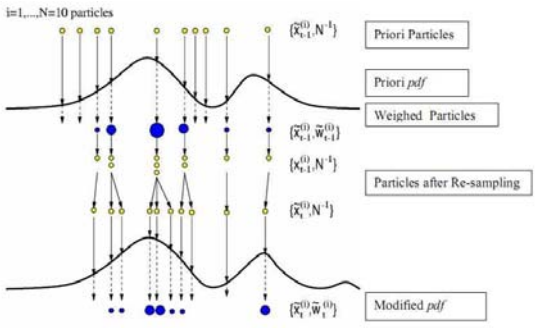

Fig. 3. Particle Filtering Process

Particle filters have high impact usage in tracking and navigation systems due to their ability to estimate future with high accuracy. In our system, we deploy particle filter to predict the position of individuals in next frame.

1) Occlusion Removal Using Particle Filtering: Occlusion is removed by the use of particle filtering. Previous position and priori pdf of the human is inserted in the human motion model for the prediction of the future position. After that the original position of the same human is determined in next frame. Both the predicted values and original values are compared, importance weights are calculated. Importance weights are normalized and importance resampling is applied in order to decrease the effect of erroneous samples and increase the effect of healthy samples which are closer to original values from observed frame. The priori pdf is updated to posteriori pdf after all the evaluation.

IV. Results

The primary purpose of the system was to track single human in a video scene. The results shows that the system is equally good for tracking of multiple human beings in different scenarios. The main issue related to multiple human tracking is occlusion. Occlusion is successfully removed using the particle filtering approach. Fig.4 shows scenario involving single human detection. These frames have only

Fig. 4. Human Detection

one person moving around randomly. After successful detection the path followed by human is also tracked. Human tracking is shown in Fig.5

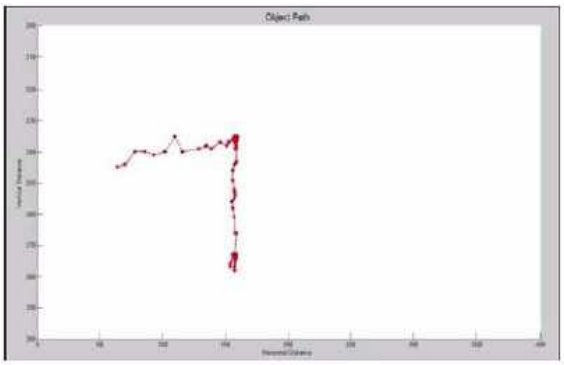

Fig. 5. Tracking of human (top view)

The scenario shown in Fig.6 involves two persons moving randomly. They are tracked and differentiated successfully.

Fig. 6. Human Detection

The results shows the use of color matching to track two different humans moving casually in an environment. Paths followed by two humans are shown in Fig.7

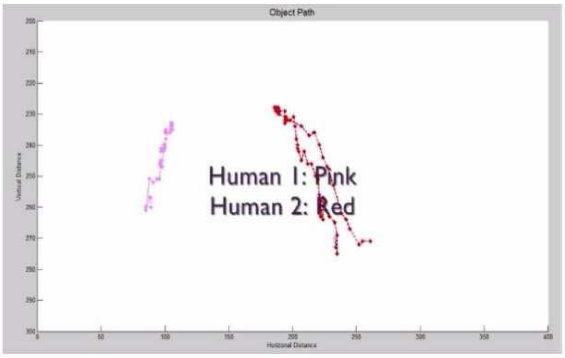

Fig. 7. Tracking of human (top view)

Fig.8 shows two humans moving in a video and they cross each other thus introducing occlusion in the frame. Particle filtering is used to avoid occlusion and results clearly depict it, as shown in Fig.9.

Fig. 8. Human Occlusion

V. Conclusion/ Future Enhancements

Humans are detected, tracked and differentiated from each other on the basis of Fast Mean Shift, Particle Filtering and Appearance Matching. Appearance matching gave us promising results in the tracking of multiple humans. Particle filtering tackled occlusion successfully. Real time evaluation shows improved performance and promising results. This system can be used for surveillance applications as well as indoor positioning system. Experimental results shows that in case of low image quality and reduced frame rate, this approach works properly.

Fig. 9. Occlusion Removed

After human tracking and detection, the task at hand would be to make the system more efficient in terms of computational cost and speed. In this research the algorithms are implemented in MATLAB but computational time taken by MATLAB makes the processing very slow. Therefore, it is suggested that the algorithms should be implemented in a faster environment such as C/C++ so that an embedded implementation such as DSP processor, can be achieved. Going a step further the "Gesture recognition" feature can be incorporated with the existing system, thus increasing the overall level of acuity of the system's senses. Using this new feature, the human motions can be extracted and various actions can be differentiated. On the gesture recognition human carrying out suspicious activities could be tracked hence making it a better surveillance system.

VI. Acknowledgements

The authors would like to acknowledge K. Qasim Maqbool for his contribution to the paper especially for providing the videos used and Dr. Abdul Bais, Sarhad University of Science and Information Technolgoy, for his guidance and support to this work. We would also like to thank Farhan Aslam Qazi for reviewing this paper.

References

[1] R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, and O. Hasegawa, " A system for video surveillance and monitoring: VSAM final report," in Technical Report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, 2000.

[2] I. Haritaoglu, D. Harwood, and L. S. Davis, W4: "Real time surveillance of people and their activities,"IEEE Trans. on PAMI, vol. 22, no. 8, pp. 809830, 2000.

[3] A.Elgammal, R. Duraiswami, and L. S. Davis, "Efficient nonparametric adaptive color modeling using fast gauss transform," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, December 2001, vol. 2, pp. 563570.

[4] Y. Kuno, T. Watanabe, Y. Shimosakoda, and S. Nakagawa, "Automated detection of human for visual surveillance system, in Int. Conf. on Pattern Recognition, Vienna, Austria, August 1996, p. C92.2.
[5] T. Zhao and R. Nevatia, "Bayesian human segmentation in crowded situations,"in IEEE Conference on Computer Vision and Pattern Recognition, Madison, USA, June 2003, pp. 459466.
[6] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," IEEE Trans. Pattern Anal. Mach. Intell., 24(5), pp. 603-619, 2002.
[7] Csaba Beleznai, Bernhard Fruhstuck and Horst Bischof, "Human detection in groups using a fast mean shift procedure" International Conference on Image Processing : (ICIP 2004) (Singapore, 24-27 october 2004)
[8] F. Crow, "Summed-area tables for texture mapping," in Proceedings of SIGGRAPH, 1984, vol. 18, pp. 207,212
[9] A.E.C. Pece, "Tracking by cluster analysis of image differences," in Proceedings of the 8th Int. Symposium on Intelligent Robotic Systems, Reading, UK, July 2000.
[10] www.mathswork.com
[11] Matthias Muhlich," Particle Filters an overview", Institut fur Angewandte Physik J.W.Goethe-Universitat Frankfurt, Filter-Workshop Bucures ti 2003
[12] Kim-Hung Li, "Pool size selection for the Sampling/Importance resampling algorithm", The Chinese University of Hong Kong, 2007.
[13] Li, K.-H. (2004). "The sampling/importance resampling algorithm. In Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives." (Edited by A. Gelman and X.-L.Meng), 265-276. Wiley, London.
[14] Rubin, D. B. (1988). Using the SIR algorithm to simulate posterior distributions. In Bayesian Statistics 3 (Edited by J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M.Smith), 395-402. Oxford University Press, Oxford.
[15] Fredrik Gustafsson, Fredrik Gunnarsson, Niclas Bergman, Urban Forssell, Jonas Jansson, Rickard Karlsson, Per-Johan Nordlund, "Particle Filters for Positioning, Navigation and Tracking", Final version for IEEE Transactions on Signal Processing. Special issue on Monte Carlo methods for statistical signal processing.