Fuzzy Based Foreground Background Discrimination for Probabilistic Color Based Object Tracking

Heydar Toossian Shandiz †, Seyed Mostafa MIRHASSANI††, Bardia YOUSEFI†† , M.J. RASTEGAR FATEMI†††

† EE Department, Shahrood University of Technology, Shahrood, IRAN
†† Vision lab, PIV Co., Tehran, IRAN
††† Islamic Azad University Saveh Branch/ EE Department, Saveh, 39187-366, IRAN

Summary
In the most of object tracking tasks dealing with partial occlusion is a challenging issue. Recently the use of color cue based on Monte Carlo tracking method and particle filtering is mostly considered to overcome the problem of partial occlusion and non-rigid motion. The proposed approach in this paper is based on using of sequential Monte Carlo and particle filtering for tracking. But in this method a special fuzzy based color model for object is employed. Then comparison of mean value of reference and candidate window in the proposed color space is utilized for tracking of objects. Some of the morphological operation is also used to provide a unit region for object location in the fuzzy based color space. Experimental results indicate that the algorithm is efficient in dealing with partial occlusion.

Key words:
Object tracking, Fuzzy decision, color cue, sequential Monte Carlo.

1. Introduction

In the past few decades, detection and tracking of people, vehicles and generally objects in multifarious environment become one of the most attractive topics of scientific researches and papers. Various applications of these kind such as human–computer interaction [1], ambient intelligent systems [2],[18] , robotics [3] ,video communication/compression [4],[5],and etc is the reason for growing interest in people and object tracking. Generally, several methods have been proposed for the solving the problem of object tracking. Some of them [6] employ a cost function minimization to locate the object in subsequent frames. In [7] color and edge of object are utilized as features for tracking. In some methods [6],[9] shape and skin color are used for the tracking task. Using of stereo information for people tracking is presented in [10] in which, a particle filter employs depth, color, gradient information of object for tracking.

Pattern recognition algorithms for tracking presented in [11], [12]. In these methods, each frame is searched for tracking of the trained pattern's object. In [13] combination of sum-of-squared differences (SSD) and color-based mean-shift (MS) trackers is used for tracking of objects in which, partial likelihood ratio help SSD tracker to deal with occlusion.

For estimating the state of a dynamic system from sequential observations, particle filtering techniques are used in [14]. Dealing with systems in which both observation and posterior density are non-Gaussian, is the reason of attraction of particle filtering for utilizing in such tracking tasks. For solving the problem of multiple target tracking, multiple particle filters [15] are employed. In which, for each person independent trackers are used. In [16] particle filter is designed base on multiple cues which are learnt from the first frame and adaptive parameters such as color, texture and edge.

Having low computational complexity and robustness to appearance change, color-based tracking methods using kernel density estimation [17],[3] have recently acquired attractiveness among researchers. In all of particle filtering based approaches, when multiple cues are employed [19] in comparison with the single cue [20], the penalty of computation complexity is paid. However, in this way, the advantages of robustness of algorithm in cluttered background, dealing with noise and changes in illumination have far outweighed the disadvantage of computational cost.

In our method particle filter with the aid of a special fuzzy based model for color space is used for tracking of objects. This color model makes discrimination between background and the object by enlightening the dominant color elements of the object in the scene. Moreover, a method for color model updating is proposed to deal with the changes in aspect and illumination condition.

2. Probabilistic Tracking

2.1 Tables and Figures *Sequential Monte Carlo tracking*

This technique tries to estimate the posterior probability density function (pdf) $p(x_k | Z_k)$ of the state vector $x_k \in R_{nx}$, where nx is the dimension of x_k.
Using a set $Z_k = \{ Z_1, \dots, Z_k \}$ of observations, given up to time k in Monte Carlo method for representation of the state pdf, samples in each state have been utilized.

Manuscript received January 5, 2010
Manuscript revised January 20, 2010

In each state, multiple particles (sampling windows) are generated. After that, each one in accompanying with a weight W_k is considered. This weight describes the quality of a particular particle L, L = 1,2, . . . ,N.
If we suppose that $\sum_{k=0}^{M} W_k = 1$, Sum of weighted particles will be the evaluation of Monte Carlo technique for our favorable variable.
This algorithm could be discerned in two main stages which are prediction and update.
In the prediction stage, with considering the favorable region in the video frame and its state model, the particles (sampling window) are modified. For simulation of noise effect in each state, additional random noise is also employed. In the next stage, known as update, based on new data, weight of each particle is re -evaluated.
The success of particle filter (PF), in comparison with the Kalman filter, could be clarified by its potential to handle non-linear observation models. Dealing with multi modal densities is other preference of PF.
Generally, in visual tracking tasks, because of the existence of elements in scene which are similar to the object, multi-modality of the measurement density is so regular [6]. Moreover, since image data are very redundant, the observation model, which relates the state vector to the measurements, is non-linear. As a consequence the Kalman filter, which employs Gaussian noise for modeling of state of the object, is usually not suitable for visual tracking.
For better conception, in a standard state space model suppose that x_k is the location of object in the kth frame. In a Bayesian framework, between prediction and update steps, a recursive relationship could be founded as follows:
Prediction step:

$$p(x_{k+1} | Z_k) = \int_{\mathbb{R}} p(x_{k+1} | x_k)\, p(x_k | Z_k) dx_k \quad (1)$$

And the update step:

$$p(x_{k+1} | Z_{k+1}) = \frac{p(z_{k+1} | x_{k+1})\, p(x_{k+1} | Z_k)}{p(z_{k+1} | Z_k)} \quad (2)$$

Where $p(z_{k+1} | Z_k)$ is a constant for normalization. As (2) indicates, following proportion for update recursion exist:

$$p(x_{k+1} | Z_{k+1}) \propto p(z_{k+1} | x_{k+1})\, p(x_{k+1} | Z_k) \quad (3)$$

Since for propagating $p(x_{k+1} | Z_{k+1})$ there is not any simple analytical expression, numerical methods are utilized.
To evaluate the posterior probability of $p(x_{k+1} | Z_{k+1})$ in particle filter, the integral is mapped to discrete sums by set of M weighted particles as follow:

$$(x_{k+1} | Z_k) \approx \sum_{m=1}^{M} W_{k+1}^{[m]} \delta(x_{k+1} - x_{k+1}^{[m]}) \quad (4)$$

Where $W_{k+1}^{[m]}$, which is normalized $(\sum_{k=0}^{M} W_{k+1}^{[m]} = 1)$, indicates importance of weights for particles.

2.2 Dynamic of the state space

The purpose of the algorithm is to track a part of image which is favorable. The object which would be tracked is located in the center of the target window. Useful information for tacking of the object is obtained from the first frame in which the target is manually specified by user. Generally, in tracking tasks, affinity to the prior location of the window and color properties of the object is two significant cues. The former is efficient when the velocity of the target could be approximated. While the latter make the tracker robust to rotation of the object since the color property is less sensitive to changing the aspect ratio of the object.
Considering the location f = (x, y) in the image coordinate system and the scale S as the hidden variables to be evaluated, the second-order auto-regressive dynamics is given as follows:

$$x_{k+1} = A\, x_k + B\, x_{k-1} + C v_k,\ v_k \sim N(0, \Sigma) \quad (5)$$

Where x_k is defined as $x_k = (f_k, f_{k-1}, s_k, s_{k-1})$. A, B, C and Σ are matrices which could be trained using the information which is obtained by manually tracking of the object in the first sequences and v_k is the noise of the state model.

2.3 Color space model

In this paper a special fuzzy based model for color is employed including a preprocessing on the frame which enlighten the color element of the object and reduce the background color elements. First from the manually tracked region of the target, the object and background color information is obtained. Then the first frame is converted to four color space including H, S, r and g where, the normalized red and green layers are $r = R/(R+G+B)$ and $g = G/(R+G+B)$ in RGB space and H, S are layers of HSV color space. The median values of the object's pixels in each color space are calculated as follows:

$$\mu_{oH} = median\ (I^m{}_H | w)$$
$$\mu_{oS} = median\ (I^m{}_S | w)$$
$$\mu_{og} = median\ (I^m{}_g | w)$$
$$\mu_{or} = median\ (I^m{}_r | w) \quad (6)$$

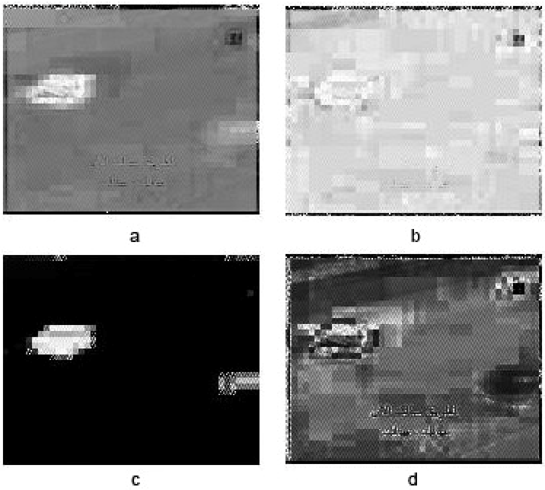

Fig.1: Fuzzified layers of first frame including (a) M_r, (b) M_g, (c) M_H and(d) M_S

Where $I^m{}_H$, $I^m{}_S$, $I^m{}_g$ and $I^m{}_r$ are the median filtered version of the original image in H, S, g and r space color, respectively. W is the object window and μ_{oH}, μ_{oS}, μ_{og}, μ_{or} are their median value. The reason of applying the median filter in each layer of the original image before calculating median value is to make the common color element of the object dominant in the image. Similarly, μ_{bH}, μ_{bS}, μ_{bg} and μ_{br}, the median value for the background in each layer of the first frame, are calculated.

After that, for fuzzification, each layer of the color space is fed to the fuzzy membership function which is given as follows:

$$M_H = \exp\left(-\left|\frac{I^m{}_H - \mu_{oH}}{\lambda * \sigma_H}\right|\right)$$

$$M_S = \exp\left(-\left|\frac{I^m{}_S - \mu_{oS}}{\lambda * \sigma_S}\right|\right)$$

$$M_g = \exp\left(-\left|\frac{I^m{}_g - \mu_{og}}{\lambda * \sigma_g}\right|\right)$$

$$M_r = \exp\left(-\left|\frac{I^m{}_r - \mu_{or}}{\lambda * \sigma_r}\right|\right) \qquad (7)$$

Where M_H, M_S, M_g, M_r are memberships of the object in each layer and $\sigma_H = \mu_{bH} - \mu_{oH}, \sigma_S = \mu_{bS} - \mu_{oS}$ $\sigma_r = \mu_{br} - \mu_{or}, \sigma_g = \mu_{bg} - \mu_{og}$. ê and ĉ are height and width of the object window and λ is a coefficient which is equal to 0.1. For smaller value of λ, we have a more discrimination for object and background. However, for very small A the object elements are partially eliminated in M images. Figure1 indicates images of memberships (M_H, M_S, M_g and M_r).

As Eq. 7 indicates, the maximum value of the M_H, M_S, M_r or M_g is equal to 1. Fuzzy decision function is given as follow: #

$$M_{total} = \min(M_H, M_S, M_r, M_g, M_{mean}) \qquad (8)$$

Where :

$$M_{mean} = \left(\frac{1}{\hat{e} \times \hat{c}}\right) \sum_{\hat{e}} \sum_{\hat{c}} (M_H + M_S + M_r + M_g)$$

(9)

Alpha cut is also utilized for defuzzification. So we have:

$$M_F = M_{total} > \alpha \qquad (10)$$

$$I^m{}_{mean} = \left(\frac{1}{4}\right) \times (I^m{}_H + I^m{}_S + I^m{}_r + I^m{}_g) \qquad (11)$$

$$M_F|_{I^m{}_{mean} > \beta} = 0 \qquad (12)$$

Where β is a large value equal to 0.9 (if we consider the maximum of $I^m{}_{mean}$ as 1). Using of this part of the inequality eliminates the regions of the frame which have low saturation. α is a value which could be obtained manually from the first frame.

At frame k, the M_F image, using the fuzzy decision theory from the color spaces and μ_{oH}, μ_{oS}, μ_{or}, μ_{og} and other predefined parameters is calculated.

The reason of using four color spaces is that the color elements of the object in each color space have a special value. If a color element of object change due to change in lighting condition, this change would be less effective in performance of the algorithm. In comparison with other similar methods which utilize the histograms of current window with the reference histogram, the proposed algorithm uses the mean value of particle window in the M_F image for the calculation of particle filter weights after the calculation of M_F image in each frame.

In state of $x_k{}^m$ for $W_{m=1:M}$ particle windows M value of φ could be calculated from following equation:

$$\varphi_k{}^m = 1 / |\, \mathrm{Mean}(M_F|_{W_{m=1:M}}) - \mathrm{Mean}(M_F|_{W_{ref}}) \,|$$

(13)

According to the particles with the greater weights and using of second order dynamic equation, next M particles ($x_{k+1}{}^m$) for frame of k+1 would be calculated. However to avoid degeneracy, a condition in which the weights of two particle are dominant and other particles have negligible weights, the best particle's weight is compared with a threshold and if it was less than the threshold value the particles of this frame are made again.

Fig. 2. Original image (up), M_{total} (middle), M_F (down)

In the most of object tracking algorithms due to occlusion, changes of aspect, illumination condition, etc the color model of object would be varied. As a result, M_H, M_S, M_r or M_g matrices change. For modification of the color model and updating of μ_{oH}, μ_{oS}, μ_{or} and μ_{og} following steps are accomplished.

Firstly, the object window with the best weight is filtered using median filter. Then the median value of the window in each color space is calculated. With this assumption that the object has been correctly tracked, the calculated median values indicate the dominant color of object. This values are updated version of μ_{oH}, μ_{oS}, μ_{or} and μ_{og}.

3. Experimental Results

In a test of algorithm, x_k and y_k are considered as 1 pixel / frame. And S_k is equal to 0.2 pixel / frame. The number of particles for each frame, M, is 90. Figure 2 shows a typical M_F image. As it indicates, the regions of image with higher intensity level show the object region. It is obvious that the fuzzy based color space is able to discriminate the object and background. The employed value for threshold, α, is sufficient to eliminate the background components. To reduce the computational costs the background is considered as a frame which is surrounding the object's window instead of the whole of image. Height and width of this rectangular area which surrounds the object's window is 10 pixels greater than that of object window.
Figure 3 illustrates the result of the algorithm in comparison with the mean shift method.

In each frame after calculation of M_F, some of morphological operations such as dilation and closing are used to provide a unit region for object and to improve the performance of the algorithm to deal with partial occlusion. For greater occlusion, larger structure element is required. The shape of structure element could be considered corresponding to the direction of motion. As figure 3 indicates, the algorithm is robust to the partial occlusion. This is due to using of the fuzzy color space and morphological operation.

4. Conclusion

A semiautomatic method for tracking of objects in video sequences presented. In this approach, sequential Monte Carlo in companying with the particle filtering is utilized. For weightening the candidate windows (particles) a special fuzzy based color model for object is used. This color model makes discrimination between background and object color elements. Experimental results indicated the efficiency of the proposed approach in dealing with the partial occlusion.

Fig.3. the object is the car with blue color. Red rectangle is the result of the proposed method and darken blue rectangle is that of mean shift method.

References

[1] C. Colombo, AD. Bimbo, A. Valli, "Visual capture and understanding of hand pointing actions in a 3-D environment", *IEEE Transactions on Systems, Man and Cybernetics*—Part B 33 (2003) 677–686.

[2] T. Darrell, D. Demirdjian, N. Checka, P. Felzenszwalb, "Plan-view trajectory estimation with dense stereo background models", *Eighth IEEE International Conference on Computer Vision* (ICCV 2001), vol. 2, 2001, pp. 628–635.

[3] G. Bradski, "Computer vision face tracking as a component of a perceptual user interface", *Workshop on Application of Computer Vision, Princeton,* NJ, 1998, pp. 214–219.

[4] B. Menser, M. Brunig, "Face detection and tracking for video coding applications", *Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers*, 2000, pp. 49–53.

[5] W.E. Vieux, K. Schwerdt, J.L. Crowley." Face-tracking and coding for video compression", *International Conference on Computer Vision Systems*, 1999, pp. 151–160.

[6] B. Lucas, T. Kanade, "An iterative image registration technique with an application to stereo vision", *international Joint Conference on Artificial Intelligence*, 1981, pp. 674–679.

[7] M. Isard, A. Blake, "Contour tracking by stochastic propagation of conditional density", *European Conference on Computer Vision,* 1996, pp. 343–356.

[8] J. Yang, A. Waibel, "A real-time face tracker", *WACV*, 1996, pp. 142–147.

[9] G. Cheung, S. Baker, T. Kanade, "Shape-from silhouette of articulated objects and its use for human body kinematics estimation and motion capture", *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2003, pp. 77–84.

[10] Rafael Mun˜oz-Salinas , Miguel Garcı a-Silvente and Rafael Medina Carnicer, " Adaptive multi-modal stereo people tracking without background modelling ", *international journal of visual communication and image representation*,19 (2008) 75–91

[11] M.J. Black, A.D. Jepson, Eigentracking:" robust matching and tracking of articulated objects using a view-based representation", *Int. J. Comput. Vis*. 26 (1) (1998) 63–84.

[12] O. Williams, A. Blake, R. Cipolla, "Sparse bayesian learning for efficient visual tracking", *IEEE Trans. Pattern Anal. Mach. Intell*. 27 (8) (2005) 1292–1304.

[13] T. Darrell, G. Gordon, M. Harville, J. Woodfill, "Integrated person tracking using stereo, color, and pattern detection", *International Journal of Computer Vision* 37 (2000) 175–185.

[14] N. Gordon, D. Salmand, "Bayesian state estimation for tracking and guidance using the bootstrap filter", *Journal of Guidance, Control and Dynamics* 18 (1995) 1434–1443.

[15] W. Du, J. Piater, "Tracking by cluster analysis of feature points and multiple particle filters", *IEEE Conference on Advanced Video and Signal Based Surveillance*, 2005, pp. 165–170.

[16] Paul Brasnett a, Lyudmila Mihaylova b and David Bull a, Nishan Canagarajah, " Sequential Monte Carlo tracking by fusing multiple cues in video sequences ",*international journal of Image and Vision Computing* 25 (2007) 1217–1227

[17] D. Comaniciu, V. Ramesh, P. Meer, Real-time tracking of non-rigid objects using mean shift, *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2000, pp. 142–149.

[18] G. Hager, P. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination", *IEEE Trans. Pattern Anal.*

[19] Mach. Intell. 20 (10) (1998) 1025–1039.D. Comaniciu, V. Ramesh, P. Meer, "Kernel-based object tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (5) (2003) 564–577.

[20] K. Nummiaro, E. Koller-Meier, L.V. Gool, "An adaptive color-based particle filter", *Image and Vision Computing* 21 (1) (2003) 99–110.

[21] C. Rasmussen, G. Hager, "Probabilistic data association methods for tracking complex visual objects", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (6) (2001) 560–576.

[22] P. P´erez1 C. Hue2, J. Vermaak1 and M.Gangnet1, "Color-Based Probabilistic Tracking", *Springer-Verlag Berlin Heidelberg* 2002, pp. 661–675, 2002.

Heydar Toossian Shandiz has received Bsc and Msc degree in electrical engineering from Ferdowsi Mashad University in IRAN. He has graduated PhD in Instrumentation from UMIST, Manchester UK in 2000. He has been assistant professor in Shahrood University of Technology, IRAN. His fields of research are identification systems, adaptive control, Image and signal processing, neural networks and Fuzzy systems.

Taro Denshi received the B.S. and M.S. degrees in Electrical Engineering from Shibaura Institute of Technology in 1997 and 1999, respectively. During 1997-1999, he stayed in Communications Research Laboratory (CRL), Ministry of Posts and Telecommunications of Japan to study digital beamforming antennas, mobile satellite communication systems, and wireless access network using stratospheric platforms. He now with DDI Tokyo Pocket Telephone, Inc.

Seyed Mostafa Mirhassani received the degree in electrical engineering from Shahrood University of technology, Shahrood, Iran, in 2005. Since 2007 he has been a researcher in PIV Company in Iran. He is especially involved in team working on classification of remotely sensed images from urban area in PIV Laboratory. He is currently working toward the MS. degree from IA University science and research in Tehran, Iran. His research interests are remote sensing, pattern recognition, computer vision and biomedical image processing.

Bardia YOUSEFI received the B.S. degree in Electrical-Electronic Engineering from Shahrood University of Technology in 2007. During 2006-2008,
he stayed in Image Processing and Computer Vision Research team in EE department of Shahrood University of Technology and worked on face detection and tile defects detection. He works currently in PIV Image Processing & Computer Vision Laboratory. He is interested in working on classification urban areas remote sensing images, Object tracking, and medical images.

Mohammad Jalal Rastegar Fatemi was born in Saveh (Iran). He received the BSc degree in power electrical engineering from Azad University of Saveh (Iran) in 2003 and the MSc degree in electronic engineering from Azad University of Arak 2006. He is currently working toward the PhD degree in the Department of Electrical Engineering, University of Research and science of Tehran in electronic engineering. He joined the Azad University of Saveh in 2006 as lecturer. His research interests focus on power electronics and semiconductor device modeling and he does some researches on image processing and computer vision.