A novel face and hands tracking in a complex background

* SANG HWAN LEE, ** SANG MIN YOON, * JI HOON JEONG
* National Security Research Institute, Daejeon, Korea
*Basic Technology Research Department
*National Security Research Institute
* 161 Gajeong-Dong, Yuseong, Daejeon, 305-700
** Imaging Media Research Center
** Korea Institute of Science and Technology
** 39-1 Hawolgok-dong, Seongbuk-gu, Seoul, 136-791
* ** KOREA

Abstract: - We present a novel ~~face and hands tracking~~ algorithm for solving problems of partial occlusion in complex backgrounds. Without background subtraction or training, we successfully extracted the features with incorporation of color and motion information to reject the static background objects that have similar skin color. Hybrid tracking algorithm based on MAWUPC algorithm and particle filter is robust to the multiple objects' movement and some kinds of distracters. This is the essential and necessary process of natural Human Computer Interaction, Video Surveillance and Monitoring.

Key-Words: - MAWUPC, Kalman filter, color transform, UPC, particle filter, HCI

1 Introduction

Tracking human body parts has been one of challenging research issues in computer vision areas for recognition and understanding of human behaviors. Especially, tracking face and hands in human body parts has been applied to many applications because face and hands are very important parts in HCI, recognition and surveillance.

Computer vision based on the tracking of human body parts has been categorized with used cues and models[1][2]. In the approaches with measurements, most tracking algorithms are relied on the information of color, motion, shape, edge, depth[3][4][5][6], or fusion of each cue[7][8]. In the tracking of human's face and hands, we encounter the problems of ambiguity and occlusion. Ambiguity and occlusion arises due to distracting noise, mismatching and overlapping the tracked objects, and a complex background that has similar color or motion information of the tracked objects. A common and robust approach for real-time tracking is to combine multiple visual cues. However, in the domain of cues such as color and motion, occlusion still brings problems because body parts such as hands have much activity and they are not virtually distinguished. Therefore, joint tracking of the body parts must be performed with an exclusion principle on observations [9][10].

To more efficient and robust tracking, particle filter and various methods are introduced. One of the typical various methods is Mean Shift. Particle filter performs random search guided by a stochastic observation model with color and motion to obtain an estimate of the posterior distribution describing the object's configuration. On the other hands, various methods such as Mean Shift localize an object based on minimizing a cost function [11][12]. In this paper, we propose the fusion of particle filter and novel observation model called MAWUPC algorithm[13]. This algorithm is probabilistic distribution that reflects the characteristic of color and motion and helps particle filter to track the objects.

Organization of this paper is as follows. In section 2, we develop a feature extraction method with color and motion information. We then presents particle filter in section 3 and the fusion of particle filter and

MAWUPC algorithm in section 4. By walking through computing the tracking of human's face and hands, we describe an approach to link the experimental results in section 5. Our concluding remark is presented in section 6.

2 MAWUPC Algorithm

To be used for robust observation model in particle filter, we convert the color input image sequence to the stochastic probability distribution image via MAWUPC algorithm. Samples are measured on the basis of our proposed probability model obtained by MAWUPC algorithm. In addition, the results of particle filter is used for measurement of the Kalman filter because they means the poison of object tracked.

2.1 Color transform

In color space models, RGB is one of representative color space. RGB color model has not only color component and but also brightness component. To reduce the brightness effect in color perception, we define the Normalized RGB color space that removes the brightness component with remaining the color component. Normalized process follows such as

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B} \tag{1}$$

where $r + g + b = 1$.

Color transform is defined to a gray image that emphasizes a specific color region with Gaussian distribution from normalized color image[14]. Transformed image has a high density as the color of the normalized input pixel is close to the center of the Gaussian distribution. The color transform is represented as follows.

$$C(x,y) = \frac{1}{2\pi\sigma_r\sigma_g}\exp[-\frac{1}{2}\{(\frac{r(x,y)-m_r}{\sigma_r})^2 + (\frac{g(x,y)-m_g}{\sigma_g})^2\}] \tag{2}$$

where (x, y) is a coordinate of a pixel , $g(x, y)$ and $r(x, y)$ are the normalized color values of green and red components, respectively. $C(x, y)$ is a 2D Gaussian function, and σ_r and σ_g are standard deviations of red and green components. m_r and m_g are also defined as mean of each component. Skin color that has stochastic probability distribution has been popularly used to tracking the typical color object, but it is difficult to avoid the confusion of tracking object and background that has similar color.

2.2 MAWUPC Algorithm

In this section, we explain MAWUPC algorithm that is the adaptive fusion of color and motion information. MAWUPC algorithm is extended to AWUPC by incorporating two methods to supply more motion information with more effective combination of color and motion information.

The first method is introducing a weighted search window around an object position that is estimated by the Kalman filter, using the spatial mask. The Kalman filter in this paper assumes the system state-equation of Newtonian dynamics. In Eq.(3), every state-variable of the state-equation is composed by position value[15]. It provides two advantages. Firstly, state-equation is not related to time-variable. Then state-equation is independent of time-interval of input sequence. Secondly, it is easy to make modeling of noise because state-variable is only one, position value. In general, the system state-equation of Newtonian dynamics has three state-variable that is acceleration, velocity, and position.

$$X_k = [\ p_k \ \ p_{k-1} \ \ p_{k-2}\ ]^T, \tag{3}$$

$$\begin{bmatrix} p_{k+1} \\ p_k \\ p_{k-1} \end{bmatrix} = \begin{bmatrix} 2.5 & -2 & 0.5 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} p_k \\ p_{k-1} \\ p_{k-2} \end{bmatrix} + \begin{bmatrix} \omega_k \\ \omega_{k-1} \\ \omega_{k-2} \end{bmatrix}. \tag{4}$$

where state-variable p_k is the position value at time k.

The measurements of the Kalman filter are obtained by the results of particle filter.

After the next position is estimated by the Kalman filter, a search window is placed at the estimated position, and the following process is restricted within the window. Additionally 2D Gaussian spatial mask is overlaid at the position of a color transformed image to get a weighted result such as

$$Z_S(x,y) = S(x,y)C(x,y)\ , \tag{5}$$

$$S(x,y) = N(p_t^-, \Sigma_p^2)\ . \tag{6}$$

where $S(x, y)$ represents the spatial mask with mean p_t^- and the covariance Σ_p^2. The mean p_t^- is the estimated position by the Kalman filter at time t. This operation improves the rejection operation of background objects having similar color to targets.

The second method is introducing a limiter function for the UPC(Unmatched Pixel Count) motion detection according to relative magnitude of motions. If a target object has a small motion and another closely located objects has a large motion, the target object is likely to have small result of the UPC operation. This effects badly the result of center obtaining process for the target object. This, in turn, results in a poor measurement for the Kalman filter. So, the estimation of the Kalman filter becomes worse and worse as frames go on. To overcome this problem, we limit the UPC operation's results using relative magnitude of motion vectors that are estimated by the Kalman filter when we can expect another object to appear within the search window for target object. The limiter operation can be formulated as follows,

$$DC_{O_1,O_2} = \begin{cases} 1, & if \quad D_{O_1,O_2} \le 2L_1 \\ 0, & otherwise \end{cases} , \tag{7}$$

$$U_{O_1} = \begin{cases} (2N+1)^2 \dfrac{|M_1|}{|M_2|}, & if \quad DC_{O_1,O_2} = 1 \ and \ |M_1| < |M_2| \\ (2N+1)^2, & otherwise \end{cases} . \tag{8}$$

In Eq. (7), DC_{O_1,O_2} represents whether another objects exists within the search window. In Eq. (7), D_{O_1,O_2} is te distance between object O_1 and object O_2, where O_1 is a target object. L_1 is the side length of the search window, which is related with size and speed of target object O_1. L_1 should be larger than the size of target object, and it should be increased as the speed is increased. Eq. (8) cuts out UPC operation's result. In Eq. (8), M_1 and M_2 are motion vectors of object O_1 and object O_2, respectively. This operation can be represented as a function $f_{limiter}(\cdot)$. Finally, the MAWUPC formulated by the following equations,

$$MAWUPC(x,y,t) = Z_S(x,y,t) \otimes f_{limiter}[UPC(x,y,t), U_0] , \tag{9}$$

$$UPC\ (x,y,t) = \sum_{i=x-N}^{x+N} \sum_{j=y-N}^{y+N} U\ (i,j,t) , \tag{10}$$

$$U(x,y,t) = \begin{cases} 1, & if \ |Z(x,y,t) - Z(x,y,t-1)| > Th(Z_S(x,y,t)) \\ 0, & otherwise \end{cases} , \tag{11}$$

$$Th(Z_S(x,y,t)) = \frac{255}{1+\exp\{\dfrac{Z_S(x,y,t)-255/2}{Q}\}} . \tag{12}$$

where $\otimes$ represents fuzzy AND operation, which can be replaced by a product, and $Z_S(x,y,t)$ is the one obtained in Eq. (5). In Eq. (10), $(2N+1)\times(2N+1)$ is the size of the window for UPC operation. The change in color transformed images is detected by using the UPC operation. The UPC operation in Eq. (10) is a kind of matching process. In this case the threshold is adjusted pixel to pixel using a sigmoid function, Eq. (12). The characteristics of sigmoid function play a role of a kind of sensitivity function. It increases the sensitivity for specific colored and located objects while reducing it for objects of no concern. This is the proposed method for more effective combination of color and motion information in this paper. Q is a characteristic parameter of the sigmoid function to determine the threshold. Finally, the result of the UPC operation's motion detection is multiplied to the color transform output to produce the MAWUPC algorithm output as in Eq. (9). This result is used as the observation model in the particle filter. Each sample is measured on the basis of our proposed probability model obtained by MAWUPC algorithm.

3 Particle Filter

The particle filter algorithm belongs to the filtering and data association class of tracking algorithms. We are interested in estimating the state of the objects at the current time-step k, given knowledge about the initial state and all measurements $Z^k = \{Z_k, i = 1,\dots,k\}$ up to the current time. Typically, we work with a state vector X. This estimation problem is an instance of the Bayesian filtering problem, where we are in constructing the posterior density $p(X_k \mid Z^k)$ of the current state conditioned on all measurements. In the Bayesian approach, this probability density function is taken to represent all knowledge we possess about the state X_k, and from it we can estimate the current position.

We need to recursively compute the density $p(X_k \mid Z^k)$ at each time-step. In prediction, we use a motion model to predict the current position of the objects in the form of a predictive PDF $p(X_k \mid Z^{k-1})$, taking only motion into account. We assume that the current state X_k is only dependent on the previous state X_{k-1} and a known control input u_{k-1}, and motion model is specified as a conditional density

$p(X_k \mid X_{k-1}, u_{k-1})$. The predictive density over X_k is then obtained by integration.

$$p(X_k \mid Z^{k-1}) = \int p(X_k \mid X_{k-1}, u_{k-1}) p(X_{k-1} \mid Z^{k-1}) dX_{k-1} \quad (13)$$

In the second phase, update, we use a measurement model to incorporate information from the sensors to obtain the posterior PDF $p(X_k \mid Z^k)$. We estimate that the measurement Z^k is conditionally independent of earlier measurements Z^{k-1} given X^k, and that the measurement model is given in terms of a likelihood $p(z_k \mid X_k)$. This term expresses the likelihood of the state X_k, given that Z^k was observed. The posterior density over X_k is obtained using Bayes' theorem:

$$p(X_k \mid Z^k) = \frac{p(z_k \mid X_k) p(X_k \mid Z^{k-1})}{p(z_k \mid Z^{k-1})} \quad (14)$$

After the update phase, the process is repeated recursively.

4 The fusion of particle filter and MAWUPC algorithm

In particle filter, each sample is measured on the basis of the probability model obtained by MAWUPC algorithm. In MAWUPC algorithm, the measurement of the Kalman filter is obtained by particle filter. This is the fusion of particle filter and MAWUPC algorithm.

5 Experiments

As the experiments, we tested our algorithm for robustness of occlusion. Color image sequences have the resolution of 320*240 pixels. Particles of each object in the tracking are 1000. Figure 1 shows the process of color transform and UPC. Fusion of each feature is used as a observation model in particle filter. Figure 2 shows the results of tracking of face and hands even if the hand occluded the face or another hand and when hands are crossed. In the figure 2, green blob is the center of face, Yellow blob is the center of right hand, and blue circle shows the center of left hands.

6 Conclusion

In this paper, we have proposed the robust face and hands tracking with the fusion of particle filter and novel observation model called MAWUPC algorithm. It provides the solution of occlusion and ambiguity problems in tracking multiple objects in a complex background that has static objects with similar skin color.

References:

[1] K.Imagawa, S.Lu, and S.Igi, "Color based hands tracking system for sign language recognition", IEEE International Conference on Face and Gesture Recognition, 1998.

[2] J.Martin, V.Devin, and J.Crowley, "Active hand tracking", IEEE International Conference on Face and Gesture Recognition, 1998.

[3] Y.Azoz, L.Davis, and R.Sharma, "Tracking hand dynamics in unconstrained environment", IEEE International Conference on Face and Gesture Recognition, 1998.

[4] C.Jennings, "Robust finger tracking with multiple cameras", International Workshop on Recognition, Analysis, and Tracking for Faces and Gesture in Real-Time systems, 1999.

[5] E. Ong and S. Gong, "A dynamic human model for hybrid 2D-3D representations in hierarchical PCA space", British Machine Vision Conference, 1999.

[6] C.Wren, B.Clarkson, and A.Pentland, "Understanding purposeful human motion", IEEE International Conference on Face and Gestures Recognition", 2000.

[7] K.Toyama and E.Horvitz, "Bayesian modality fusion: Probabilistic integration of multiple vision algorithms for hand tracking, Asian Conference on Computer Vision, 2000.

[8] J.Triesch and C.von der Malsburg, "Self-organized integration of adaptive visual cues for face tracking", IEEE International Conference on Face and Gesture Recognition, 2000.

[9] C.Rasmussen and G.Hager, "Joint probabilistic techniques for tracking multi-part object", IEEE Conference on Computer Vision and Pattern Recognition, 1998.

[10] J.Mac Cormick and A. Blake, "A probabilistic principle for tracking multiple objects", IEEE International Conference on Computer Vision, 1999.

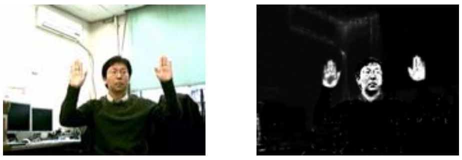
[11] Caifeng Shan, Yucheng Wei, Tieniu Tan, and Frederic Ojarias, "Real time hand tracking by combining particle filter and mean shift", IEEE International Conference on Automatic Face and Gesture Recognition, 2004.

[12] E.Maggio, and A.Cavallaro, “Hybrid particle filter and mean shift tracker with adaptive transition model”, IEEE International Conference of Acoustics, Speech, Signal Processing, 2005.

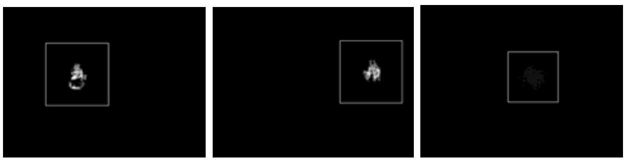
[13] S.Hwan Lee, Sang Chul Ahn, Hyoung-Gon Kim, "MAWUPC algorithm for tracking face and hands in complex background", ISPA2001, pp. 295-300, June 2001.

[14] H.G.Kim, N.H.Kim, and S.C.Ahn, ”Skin region extraction using moving color technique”, IEEE International Workshop on ISPACS, 1998.

[15] Ig-Jae Kim, SHwan Lee, Sang Chul Ahn, Yong-Moo Kwon and Hyoung-Gon Kim, “3D Tracking of Multi-objects using Color and Stereo for HCI”, ICIP 2001, pp. 278-281, October 2001.



(a)input color image (b)color transform



(c) UPC of each target objects in the search window

Figure 1. Process of color transform and UPC

Frame # 147 Frame # 156 Frame # 164

Frame # 186 Frame # 195 Frame # 204

Figure 2. Occlusion of hand and face