A Robust Method for Multiple Face Tracking Using Kalman Filter

Zaheer Shaik and Vijayan Asari
Computational Intelligence and Machine Vision Laboratory
Department of Electrical and Computer Engineering,
Old Dominion University, Norfolk, VA 23529-0246, USA
Email: {zshai001, vasari} @odu.edu

Abstract

A robust method for tracking faces of multiple people moving in a scene using Kalman filter is proposed in this paper. To distinguish faces of people during partial occlusion the proposed method uses the non-parametric cloth distribution. To overcome the problem of total occlusion, faces are tracked using the values generated by Kalman prediction algorithm. The size, top-left coordinate and velocity of motion of the detected face being the parameters of the Kalman vector; the predicted values are used to locate faces in the next frame. The faces are redetected and the templates are updated at discrete time intervals when the similarity measures, between the faces detected and respective face templates, are less than a preset threshold. Skin segmentation based face detection makes the algorithm computationally simple, and updating the face template makes it invariant to pose changes. The proposed method is experimented to be invariant to lightning conditions, change of pose, and works well in the case of partial and total occlusion for a short period.

1. Introduction

The main objective of face tracking is to locate people in successive video frames. Face tracking has wide spread applications in multimedia analysis and human machine interface. It can also act as a pre-processing step for other high level applications. Face tracking is different from face detection, where temporal data is used to locate faces in the successive frames. The major problems encountered in face tracking are changes in illumination, pose of the faces being tracked and occlusion in the case of tracking multiple people.

The face tracking methods can be mainly classified into four categories. The first one being the *shape based* approach, where the elliptical contour of the face is like in [1]. This tracking method is not influenced by background color and illumination changes, but the assumption may break when a person is occluded by an object or another person. These conditions make it unsuitable for tracking multiple faces. The second type of method being *feature based* tracking, where the invariant structural features are extracted from the faces to classify the extracted faces. Like in [2] features extracted from Gabor filters are used as cues for face tracking. This approach is usually insensitive to illumination and pose changes. To deal with the feature variations during large pose variations correlation templates have been used in [3]. However, they are computationally expensive and are hard to implement for real time application. The third type of approach being the *model based* approach where face models are used for detection and tracking. To overcome the problem of pose variations, the face is modeled using a 3D model like in [4] and is used for pose estimation and for matching purposes. The model based approach is reliable and accurate, but the computational cost is high making it unsuitable for real time applications. The final category of face tracking method, the *color based* approach like in [5], is suitable for real time applications. These approaches rely upon the accuracy and the robustness of the skin color model used. The model fails when the faces are occluded by other faces or objects. These types of approaches can track multiple faces; however they could not track during total occlusion. Using multiple cues in face tracking, like in [6], considering the face and cloth models, makes the algorithm robust against the problems of partial and total occlusions. In this paper, we present an approach for tracking multiple faces for overcoming the problems of illumination and pose changes, velocity and trajectory changes of the faces, and also to recover from the problems of partial and total occlusions. The tracker is initialized by face detection using the Viola-Jones face detection method [7], where the face templates are initialized. The non-

1550-5219/07 $25.00 © 2007 IEEE
DOI 10.1109/AIPR.2007.21

parametric density distribution of cloth model is also initialized for each respective person. The Kalman filter is also initialized at this point for each face. The Kalman vector parameters are initialized to the top-left coordinates of the bounding face rectangle, the velocity of the face motion is initialized to 1, and the width parameter is initialized to the width of the face rectangle. The algorithm uses multiple cues to overcome the tracking problems like, the pose variations using template matching and updating, illumination variations by considering only the chromatic color spaces, and occlusion recovery by using occlusion detection, Kalman prediction, non-parametric cloth distributions of the cloth etc. The results demonstrate the capability of the algorithm to deal with the above mentioned problems.

In this paper, the tracking algorithm is presented in Section 2. Experimental results and the importance of each algorithmic step is explained and demonstrated in Section 3 followed by the conclusion in Section 4.

2. The Multiple Face Tracking Algorithm

Fig. 1 shows the process flow of the proposed algorithm. In the proposed algorithm, the face tracker is initialized using the Viola-Jones face detection [7] based on the Adaboost algorithm, provided in the OpenCV library [8]. The face detection results in bounding rectangles around each face detected. The number of faces to be tracked is equal to the number of faces detected during the first frame. The representation for the cloth region is derived by employing a convex and monotonically decreasing function like the Epanechnikov kernel, which assigns smaller weights to the locations of the pixels farther from the center of the region [6], increasing the robustness of the density estimation. Given the distribution of colors in the cloth region, x_{ij} being the pixel location inside the cloth region with origin at the center, the non-parametric distribution of the cloth region q_u is computed by:

$$q_u = C \sum_{i,j} k(\|x_{ij}\|^2)\, \delta[b(x_{ij}) - u] \qquad (1)$$

where, δ is the *Kronecker delta* function, k is the Epanechnikov kernel function and b is the function that associates the pixel at location x_{ij} to the index $b(x_{ij})$ in the quantized color space.

The linear Kalman filter [9], a well known algorithm for state prediction is used to predict the location and size of the face in the subsequent frames. The Kalman vector, with five parameters in its state vector is initialized. The coordinates of the top-left position of the face, the velocity in x and y directions (initialized to 1) and the width of the face rectangle are

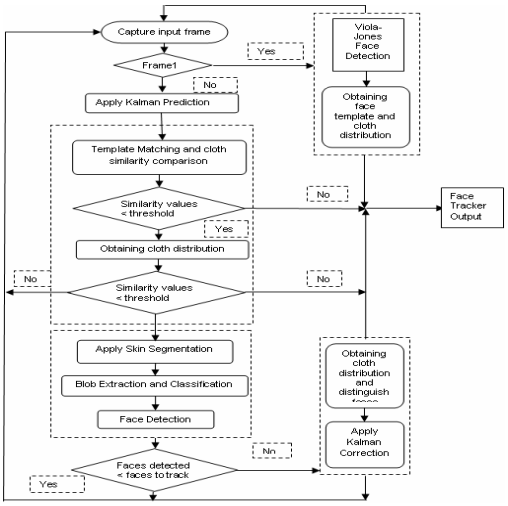

Fig.1. Process flow of the proposed algorithm.

the elements of the state vector. The height is excluded assuming that the width and height of the face are in the ratio of 1:1.25. The state-space representation of the tracker used in the Kalman filter is given below:

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \\ x'_{t+1} \\ y'_{t+1} \\ W_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_t \\ y_t \\ x'_t \\ y'_t \\ W_t \end{pmatrix} + w_t \qquad (2)$$

where, x_{t+1} and y_{t+1} are the predicted coordinates of the face and x'_t and y'_t are the velocities of the face in the respective direction, W_t represents the width of the face rectangle, Δt represents the time interval of state correction and w_t is the white Gaussian noise with diagonal variance Q.

The predicted coordinates and dimensions of the face are used to locate the face in the present frame. Using template matching technique, with normalized cross correlation coefficient as a similarity measure, the distance between the face template from the previous frame and the image region with the predicted values is calculated. If the value is observed to be less than a preset threshold (experimentally observed to be 0.9) a partial occlusion is assumed. Based on the prediction coordinates of the face, the non-parametric distribution of the cloth model, p_u, is modeled in the present frame. The similarity measure, using Bhattacharya metric, between two discrete distributions is given by:

$$\rho = \sum_u (p_u . q_u)^{1/2} \qquad (3)$$

If the similarity value is observed to be less than a preset threshold (experimentally observed to be 0.5), the models are assumed to be deviating from the correct values and need to be updated.

To update the face template and cloth models the faces need to be redetected in the frame. Skin segmentation in the quantized color space is used to classify pixels into a skin or non-skin class. The luma separated YCbCr color space is used for this purpose. Human skin color forms a relatively tight cluster, even when considering darker and brighter skins in a certain transformed 2-D color space (Cb and Cr only). Color allows fast processing and is robust to changes in pose and illumination. Also, excluding the luminance component makes the algorithm robust against illumination variations. Segmentation of the skin pixels can be obtained by setting a threshold in the Cb and Cr values (experimentally observed to be Cb ϵ [100, 135] and Cr ϵ [135, 170]). After extracting the skin pixels, the next step is to group all the skin pixels from the face region for face detection. The first stage is to erode the whole image region using a face structured element (elliptical element) to create separation between the faces. To fill the gaps as a result of erosion, the eroded image is dilated using a face structured element of less size than the eroding structuring element. The result of these morphological operations is a binary mask comprising of "blobs" in the face regions.

To detect the region surrounding each blob, the processed image is passed to a contour processing algorithm to extract the boundaries of the probable face regions in the binary mask obtained. Contour processing provided by OpenCV which is based on Freeman chain coding is used in this process to detect the boundaries. These detected blobs, with surrounding boundaries are to be further classified into face or non face regions depending upon the geometrical and statistical features. A statistical parameter called smoothness coefficient based on the second order moment, variance, is calculated for each blob region and is classified into a face and non-face region, if the smoothness coefficient is observed to be less than 0.5. Some of the "blobs" detected are eliminated using this process. Further classification is done based on the geometrical features of the "blobs" detected. A geometrical parameter, called *form factor*, which is the ratio of the area over the perimeter is calculated from the regions remaining in the binary mask. The form factor threshold, calculated from the minimum area and perimeter of the bounding rectangles obtained from the Viola-Jones face detection method in the first frame, is used to make further eliminations in the remaining "blobs". The left over regions are observed to be the face regions and the bounding rectangles are drawn by taking the extreme coordinate of each probable face region detected. Since the dimensions the face (width and height) are assumed to be in the ratio of 1:1.25, the face rectangle is drawn by taking only the width of the "blob" region detected into consideration and scaling its height with the above ratio.

If the number of faces detected is less than the number of faces being tracked, total occlusion is assumed and the program continues on the predicted values until occlusion ends. To distinguish the faces detected, a new face is registered if all the other faces are classified as to be the ones tracked in the previous frame based on the template matching and the cloth distribution model data obtained in the previous frame. When the faces are distinguished, the Kalman vector is updated using the measurement equation given below:

$$\begin{pmatrix} mx_{t+1} \\ my_{t+1} \\ mW_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{t-\Delta t} \\ y_{t-\Delta t} \\ x'_{t-\Delta t} \\ y'_{t-\Delta t} \\ W_{t-\Delta t} \end{pmatrix} + v_t \quad (4)$$

where, mx_{t+1} and my_{t+1} are the measured coordinates, mW_{t+1} is the measure width of the face at time t and v_t is white Gaussian noise with diagonal variance R. The position, velocity and acceleration are updated based on the values obtained in the present frame and the data from the previous frame. In this manner the algorithm is able to track multiple people in the case of occlusion, lighting variation and also during pose variance. The proposed algorithm is summarized depending upon the types of problems encountered:

2.1. Partial Occlusion

Partial occlusion is a case where the details (face or cloth) are lost partially due to overlap by a rigid object, a complete change in face orientation, or due to an overlap of faces being tracked. To overcome the problems of these kinds of situations, rather than running the program on predicted values given by the Kalman prediction algorithm, secondary details like the cloth distribution model is considered. This increases the robustness of the algorithm by including multiple cues to overcome the global face tracking problems.

2.2. Illumination and Pose Variations

The change in illumination is the major problem in the case of color based tracking techniques. Since, the proposed algorithm employs skin segmentation and

cloth models changes in illumination is a serious issue. To overcome this problem, the luminance component is separated from the transformed 2D color space chosen. Also, since the algorithm employs a template matching technique to distinguish faces any sudden change in pose will not match the templates obtained in the previous frame. So, the secondary details of cloth models are used to distinguish faces. In a special case, where the illumination and also pose changes simultaneously, the template and cloth models are updated. Any change in pose is adapted and assumed to stay for at least a short period or the algorithm is forced to run on predicted values for a longer period without correction.

2.3. Total Occlusion and Recovery

Total occlusion is a common problem encountered in tracking faces of multiple people. This might be due to complete overlap of a person by another person or due to a rigid object occluding the person completely. To overcome the problem of total occlusion, the proposed algorithm makes use of predicted values from the Kalman prediction algorithm. When the similarity values for both the face and cloth are less than threshold then a total occlusion is detected. After the occlusion ends, the faces are redetected and are distinguished based on template matching combined with cloth similarity comparison values. The templates are updated and can be used for further reference. Using multiple cues for occlusion recovery makes the algorithm robust against occlusion recovery problem.

3. Experimental Results

The proposed algorithm was tested in various lighting conditions with people overlapping and changing pose. The videos were collected in our lab and also in the apartment which have different lighting conditions. The video camera position was also moved slightly without zooming. The results are presented in this section and are also compared with a detection based tracking method, where faces are detected in each frame but not distinguished. The results presented demonstrate the robustness of the algorithm in various scenarios.

Fig. 2 demonstrates the face tracking algorithm, based on the skin segmentation algorithm in tracking two people. The face rectangles are drawn considering the face dimensions, assumed to be in the ratio of 1:1.25. The clips are collected at the frames 1, 57, 100, 109, 198, 213, 219, 230 and 263 respectively. In this type of tracking the faces are detected in each frame and the rectangles are bounded based on detection.

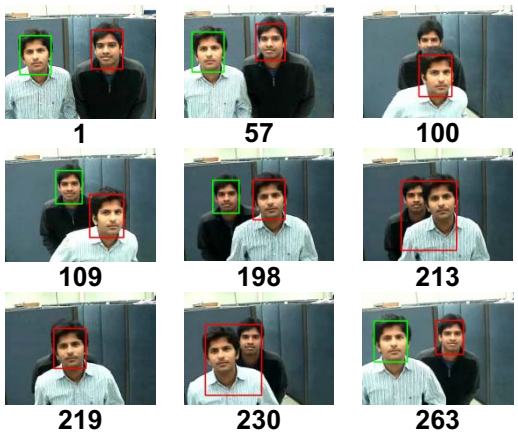

Fig.2. Detection based tracking.

Fig. 3 demonstrates face tracking based on the proposed algorithm. The video used to take the images shown in Fig. 2 and Fig. 3 is of the same, and is captured at the same time. The images shown concentrates mainly on the problems of face tracking. In frame 1, the face templates are extracted and the cloth models are extracted with the green rectangles below the face

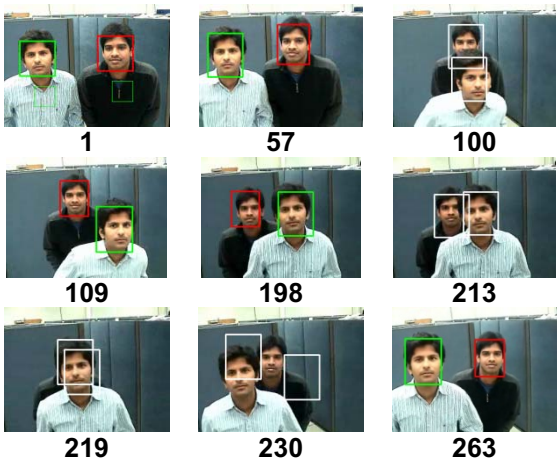

Fig.3. Proposed algorithm based tracking.

bounding the rectangles. Frame 100 demonstrates the problem of partial occlusion. The white bounding rectangles show that the algorithm uses the predicted values for tracking. In Fig. 2, at frame 100, it can be seen that only one face is detected, but the faces are tracked during partial occlusion with the proposed algorithm. In frame 109, 198, where the face positions are switched the color of the bounding rectangle is switched in Fig.2, but remains the same in Fig.3 demonstrating the partial occlusion recovery problem. In frame 213, the face regions are close enough which results in a single larger bounding rectangle in Fig.2

when compared to Fig. 3 which outputs the right values. This demonstrates that the algorithm works even when the face is occluded by a face with similar color values. Frame 219 demonstrates an example of complete occlusion. In Fig.2 it can be seen that the tracking algorithm based on face detection outputs only one rectangle, but the proposed algorithm gives the correct bounding rectangles. Frame 230 shows a case where both algorithms fail due to a long occlusion period. The predicted values deviate from the normal due to velocity changes during occlusion. Frame 263 demonstrates that the proposed algorithm recovers from total occlusion.

Fig. 4 demonstrates the x-coordinate location of the face being tracked, the person with the white shirt being considered in this case. The proposed algorithm is compared with the original values and also with detection based tracking. The blue colored line representing the detection based tracking outputs errors when the face rectangles are switched after occlusion.

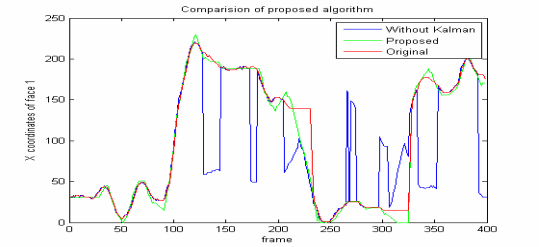

Fig.4. Comparison of x-coordinate location of face.

The proposed algorithm based tracking is efficient in tracking the respective person which yields a very small error during partial occlusions. Both algorithms output the same error value when the total occlusion is for a long period of time. Fig. 5 demonstrates the mean square error in tracking the face of the person wearing the white shirt in Fig. 3. Fig. 5 shows that the errors are reduced because the faces are distinguished after occlusion which is not the case in detection based tracking.

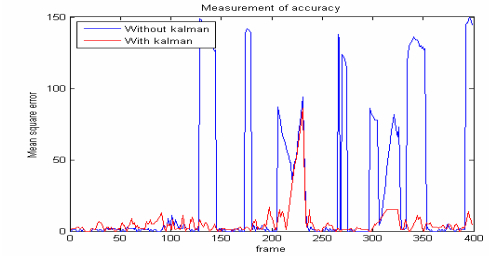

Fig.5. Comparison of mean square errors.

Fig. 6 and Fig. 7 show the comparison face similarity and cloth similarity values for the person wearing the black shirt in Fig. 3. The figures demonstrate that during intermediate frames and during partial occlusion where the face values are below the preset threshold, the cloth values are above the threshold value, which is used to overcome the problem of partial occlusion. The peak undershoots in detection based tracking (blue line) is a result of tracking the wrong person.

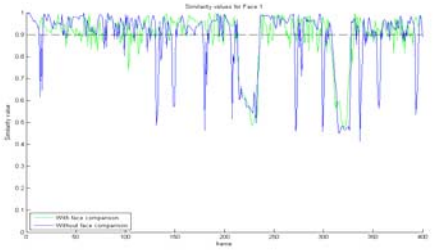

Fig.6. Comparison of face similarity values.

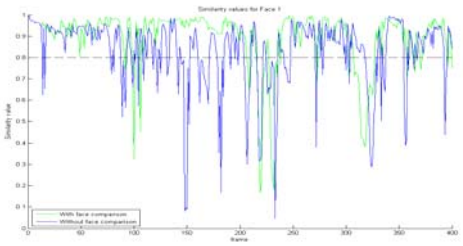

Fig.7. Comparison of cloth similarity values.

Fig. 8 demonstrates the robustness and accuracy of the proposed algorithm in the case of pose change and occlusion recovery and in the case of tracking three people.

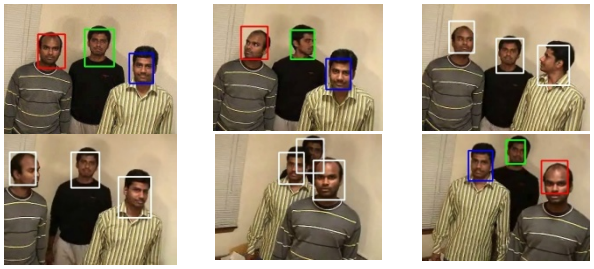

Fig.8. Proposed algorithm tracking three faces.

Table 1 shows the summary of a few face tracking techniques obtained from [6]. Many of the algorithms are incapable of tracking multiple people and occlusion recovery problems, which are overcome by the proposed algorithm. The algorithm was implemented using MS C++ using OpenCV library on a 2 GHz PC and can process a 320 × 240 video at 13 f/s.

Table 1: Summary of face tracking techniques.

Author	Yr	Tracking method	MT *	OR *
K. Schwerdt et al. [11]	2000	Track the skin region based on color histogram.	No	No
J. Ruiz-del- Solar et al [12]	2003	Repeat face detection on the image sequence.	Yes	No
V.Vezhnevets et al.[13]	2002	Locate the elliptical shape of the skin color region, verify by locating facial features in the skin area.	No	No
J. Yang et al [14]	1996	Predict search region of face from motion estimation, then apply face detection.	Yes	No
L. Wang et al. [15]	2002	Background subtraction of moving regions, locate face using template matching.	Yes	No

*MT- Multiple Tracking *OR- Occlusion Recovery

4. Conclusion

A robust algorithm to track multiple faces, which can overcome the problems of illumination variations, pose changes and, partial or total occlusion for a short period has been presented. The algorithm is also robust against change in the position of the camera, and the persons can move front and back while being tracked.

5. References

[1] A. Eleftheriadis and A. Jacquin, "Automatic face location detection and tracking for model-assisted coding of video teleconferencing sequences at low bit-rate", *Signal Processing: Image Communication*, vol. 7, no. 3, Sept. 1995, pp. 231-48.

[2] S. McKenna, S. Gong, R. Würtz, J. Tanner and D. Banin, "Tracking facial feature points with Gabor wavelets and shape models", *Audio- and Video-Based Biometric Person Authentication. First International Conference, AVBPA'97. Proceedings*, 1997, pp. 35-42.

[3] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition", *Proceedings 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994, pp. 84-91.

[4] Y. Wu and T.S. Huang, "Non-stationary color tracking for vision based human computer interaction," *IEEE Transactions on Neural Networks*, vol. 13, no. 4, July 2002, pp. 948-60.

[5] Y. Raja, S. McKenna and S. Gong, "Tracking and segmenting people in varying lighting conditions using color", *Proceedings Third IEEE International Conference on Automatic Face and Gesture*, 1998, pp. 228-33.

[6] C. Lerdsudwichai, M. Abdel-Mottaleb and A.-N. Ansari, "Tracking multiple people with recovery from partial and total occlusion", *Pattern Recognition*, vol. 38, no. 7, July 2005, pp. 1059-70.

[7] P. Viola and M. J. Jones, "Robust Real-Time Face Detection", *International Journal of Computer Vision*, vol. 57, no. 2, May 2004, pp. 137-54.

[8] Intel Corp., Opencv Library, World Wide Web, http://www.intel.com/technology/computing/opencv.

[9] R. Belaroussi, L. Prevost, and M. Milgram, "Combining model-based classifiers for face localization," *Ninth IAPR Conference on Machine Vision Applications*, 2005, pp. 290–293.

[10] Rein-Lien Hsu, Abdel-Mottaleb, M and Jain, A.K, "Face detection in color images", *Proceedings 2001 International Conference on Image Processing*, vol. 1, no. 1, 2001, pp. 1046-9.

[11] K. Schwerdt and J. Crowley, "Robust face tracking using color", *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 90-5.

[12] J.R. Solar, A. Shats and R. Verschae, "Real-time tracking of multiple persons", *Proceedings 12th International Conference on Image Analysis and Processing*, 2003, pp. 109-14.

[13] V. Vezhnevets, "Face and facial feature tracking for natural human computer interface," *Conference of Graphicon*, 2002.

[14] J. Yang and A. Waibel, "A real-time face tracker", *Proceeding. Third IEEE Workshop on Applications of Computer Vision. WACV'96* 1996, pp. 142-7.

[15] L. Wang, T. Tan and W. Hu, "Face tracking using motion-guided dynamic template matching", *Proceedings of the Fifth Asian Conference on Computer Vision*, vol. 2, no. 2, 2002, pp. 448-53.