

Complex Adaptive Systems, Volume 1
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University University of Science and Technology
2011-Chicago, IL

Tracking and Recognizing Multiple Faces Using Kalman Filter and ModularPCA

Jacob Foytik^a, Praveen Sankaran^b, Vijayan Asari^a

^aUniversity of Dayton, Dayton, Ohio

^bNational Institute of Technology Calicut, Kerala, India

Abstract

Real-time tracking and recognizing multiple faces in complex environments has the ability to provide efficient security automation to large areas. Previous research has shown that Kalman filter techniques paired with the traditional face detection methods can be used to track one or more faces in a viewing region, but prove unreliable under variant conditions due to the inability to reliably distinguish between multiple trackers. A real-time face tracking and recognition system is presented that is capable of processing multiple faces simultaneously. The proposed system utilizes the Kalman filter for tracking and uses a low-level recognition system to properly distinguish between the many trackers. This low-level system is created using a face database of twenty unrelated people trained using Modular Principal Component Analysis (MPCA) and classification is performed using a feature correlation metric. After tracking the faces, they are then analyzed by a high-level face recognition subspace which is created using a large database of people and processed using Adaptive MPCA. The overall system is shown to provide reliable tracking of more than one person and to allow a more accurate recognition rate due to the ability to create a time-average of the recognized faces.

Keywords: Kalman Filter; Modular PCA; Face Recognition; Face Tracking

1. Introduction

Advanced security applications that can effectively distinguish and track multiple people simultaneously have the ability to provide significant aide in the surveillance of large areas. Creating an effective real-time automated security application capable of tracking and recognizing multiple people simultaneously would prove useful. Ideally, all that is needed to create such a system is an entirely accurate face recognition system that reliably recognizes people in view and maintains a log of their locations. However, the current state of face recognition techniques does not provide enough accuracy with very large face databases, variant lighting and face pose angle. Furthermore, various face recognition methods created to mitigate the variances in the scene, such as that presented in [1], are too computationally complex to track individuals in real-time. To obtain higher recognition accuracy, it is beneficial to analyze faces over a period of time through a tracking process, creating a time-average of the recognized faces. The face tracking system must consistently distinguish the differences between multiple faces in the scene with varying environmental conditions and face occlusions. In previous related works, specifically [2], Kalman filter techniques and basic face template matching were employed to successfully track multiple faces simultaneously. In this method, a secondary cloth similarity test was also used when the face template methods fail. This ensured that two trackers were never confused, but was based on the assumption that two people with similar face templates would be wearing very different clothes. However, the Kalman face tracking algorithm proved to be capable of multiple tracking and recovering the face region after experiencing occlusions [2]. This paper maintains the general concept of this previous research, but aims to further enhance the tracking methods by replacing the face template matching technique with more capable methods which eliminate the need of the cloth similarity process. A new Kalman tracking system is developed which distinguishes between tracker units using Principal Component Analysis (PCA) based techniques.

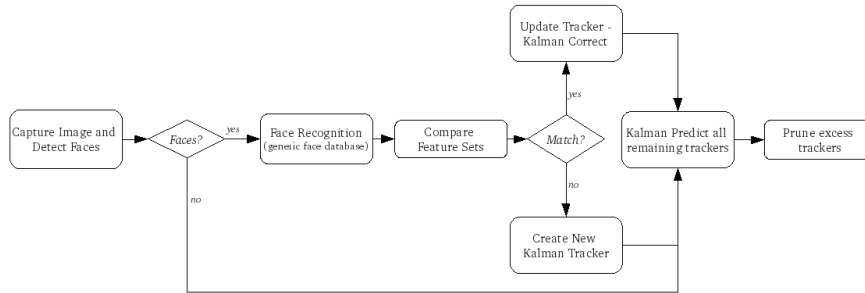


Figure 1: Flow chart of multiple face tracking system.

2. Multiple Face Tracking

An overall view of the multiple face tracking system is shown in figure 1. As the diagram shows, after a frame from the live video stream is captured, we first apply haar-based face detection methods [3] to quickly locate all regions of the image containing features pertaining to a human face. If any faces are detected, they are then analyzed by the low-level face recognition system to form a feature set, which is used to create a similarity score with all of the existing trackers. The tracker unit with the closest similarity to the input face is then compared with a threshold value to determine if a *match* is found. If a suitable *match* is obtained between the input face and a currently tracked face, then the corresponding tracker is updated with measurement data. If no tracker is significantly similar to the input face, or no tracker has been created yet, a new tracker is created for the person. After performing any measurement updates or creating trackers as necessary, we then apply the Kalman prediction equations to all remaining trackers that were not updated. This allows faces that were not detected, possibly due to face occlusions or poor pose angle, to be effectively tracked by analyzing the trackers motion history. Finally, after all trackers are updated, they are then pruned for any duplicates.

2.1. Kalman Tracker

The process of tracking any object involves using past and present information to estimate future changes of the object. In the particular case of face tracking, we are calculating and predicting changes of the x-y coordinates and the overall size of a face in the image region. The well known linear tracking methods first introduced by Kalman [4], consist of a set of mathematical equations that calculate the state of a process, while minimizing the mean of the squared error [5]. These equations are separated into two categories, *measurement update* and *time update* equations. The *measurement update* equations correspond to the equations used to correct the tracker in the event that the face being tracked is detected and its location in the image is definitive. Conversely, the *time update* equations defined by Kalman [4] are used to predict the present location of the tracked face when it has not been currently detected in the image space. In the case of our face tracking system, the Kalman state vector is initialized with five parameters consisting of the x-y coordinates of the upper left bounding point of the face region, the velocity in the x and y directions, and the width of the face bounded region. The height of the face bounded region is always assumed to be 1.25 times the size of the calculated width. The x-y coordinates and width of the face region are initially set to the values given by the face detection process, while the velocity values of the state vector are set to 1. The state-space representation of each tracker is given as

$$\begin{bmatrix} \hat{x}_t \\ \hat{y}_t \\ \hat{x}'_t \\ \hat{y}'_t \\ \hat{W}_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ x'_{t-1} \\ y'_{t-1} \\ W_{t-1} \end{bmatrix} + w_t \quad (1)$$

where \hat{x}_t and \hat{y}_t are the Cartesian coordinate values, \hat{x}'_t and \hat{y}'_t are the velocities, and \hat{W}_t is the width of the newly predicted face region. The 5×5 matrix is the transition matrix, A [4], which relates the current state to the predicted state after the time interval Δt . Furthermore, w_t is defined as the measurement noise of the process and is assumed to be white Gaussian noise. From this equation set, we see that future values are logically calculated using previous

knowledge of the tracker's motion. This prediction method is required whenever a face is not detected for the given tracker. Thus in the instance of a face occlusion or poor pose angle, the tracker predicts the motion.

On the other hand, when faces are detected and matched with a tracker system, they are then given as measurement input. The Kalman tracker then uses this data to effectively correct the system and obtain a probable state vector. The measurement correction equation is given as,

$$\begin{bmatrix} x_t \\ y_t \\ x'_t \\ y'_t \\ W_t \end{bmatrix} = \begin{bmatrix} \hat{x}_t \\ \hat{y}_t \\ \hat{x}'_t \\ \hat{y}'_t \\ \hat{W}_t \end{bmatrix} + K_t \left\{ \begin{bmatrix} mx_t \\ my_t \\ mW_t \end{bmatrix} - \begin{bmatrix} 1 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}_t \\ \hat{y}_t \\ \hat{x}'_t \\ \hat{y}'_t \\ \hat{W}_t \end{bmatrix} \right\} \quad (2)$$

where the state vector on the left side of the equation is the calculated posterior state of the tracker, and the vector to the right of the equation represents the prior state calculated using the Kalman prediction equation stated above. Also, K_t is the Kalman gain factor which is based on the error of the tracker, and mx_t , my_t , and mW_t form the measurement vector where the x-y coordinate and width of the input face region is measured. These equations allow the tracking system to be properly corrected so that the process remains seamless.

Using the prediction and correction methods described in equations 1 and 2 as a basis for the tracking unit, the multiple face tracking system defines an independent unit to each distinct person being tracked. Each Kalman tracker maintains a record of its person's recognized name, interpreted facial features, calculated state vector, and the length of time since a measurement update. Storing the person's recognition name allows a time-average process to take place. Furthermore, the facial features give the ability to analyze similarities between other trackers and input face images. Storing the state vector is required in order to define the present and future position of the tracked face. Lastly, keeping a record of the length of time since a measurement update occurred allows the overall face tracking system to have a definitive metric for pruning inactive trackers.

2.2. Low-level Face Matching

The Kalman tracking system, though adequate in following a single person at a time, is not equipped to track multiple people simultaneously. This process requires the addition of a system that can distinguish between many different people to appropriately link active trackers. As stated before, previous methods relied on simple template matching techniques, which do not provide reliable results in variant conditions. Thus, we explore an alternative method for fast and accurate face matching. Face recognition research has shown that applying Principal Component Analysis (PCA) to extract key features from a face database can provide very accurate results. Additionally PCA performs all vigorous calculations before the real-time process (off-line), requiring only small calculations to be performed during the actual recognition process. Recognition methods based on PCA are known to function well when no scene variations are present, such as pose angle, lighting, and face expression variations [6]. However, later advancements by Gottumukkal and Asari [7] formed the modular approach to PCA face recognition which significantly improved accuracies for variances in lighting and face expression. Notably, Modular PCA provides excellent recognition rates for particularly small face databases, which serves well for our tracking purposes. Since we are creating a generic tracking system that must be able to function for any person, we cannot train the matching system off-line with known faces. One solution to this problem is to collect the face images and train during on-line processing. This would allow an accurate feature space to be created with the correct face images for each person, but would invoke complex calculations during the tracking process. On-line training would drastically slow the system down and would not be able to perform in real-time. However, an alternative method involves training the face matching system with several unrelated faces to create a generic feature space. Using this trained feature space, the input face images of the tracking system are analyzed and a similarity score is given according to each unrelated person in the database. This array of similarity scores form a subsequent feature vector for the tracked person which can be used for comparison with other trackers in the system.

The recognition database was populated with people having very different appearances, such as skin color, eyes, nose, and hair, in order to maximize the feature space. Also faces for the database were selected to have the exact same face expression and pose angle to provide consistent similarity results. The training set used in the implementation of the system consisted of 20 different people chosen from the face database created for the 2003 Face Recognition Grand Challenge (FRGC) [8].

Training the matching system followed the same process as described in [7]. First, the face images are masked to remove any non-facial features. Then, the face images are broken into N smaller images (modules), a covariance matrix

is formed from the mean-subtracted face modules, and the Eigenvectors and Eigenvalues are extracted. A weight set is then generated for each person in the database by projecting each person's face image onto M Eigenvectors corresponding to the largest Eigenvalues. M is a number much less than the number of pixels in the face image. This feature extraction process essentially compresses the face image to a small feature vector that defines a weighting for the Eigenvectors (Eigenfaces) that were calculated in the training process.

The real-time processing for the face matching portion only contains a small number of operations. First the face is broken into modules and projected onto the Eigenvectors calculated by the training process to define weight vectors in the feature space. The weight vectors are then compared with the weights of every person in the database using the simple Euclidean distance measure,

$$D_p = \sqrt{\sum_{i=1}^M (x_i - x_p)^2} \quad (3)$$

where D_p is the distance between the input face and the face of person p in the database. In normal face recognition algorithms, nearest neighbor methods are used to determine the most similar person. However, since the people being tracked are not likely to correspond to the people trained by the system, the most similar candidate is likely to fluctuate frequently. Rather than use the most similar person to define the tracker, we maintain a log of the similarity scores for all people in the database, which forms a subsequent feature vector.

From experimental analysis, it was noted that the feature vectors can be accurately classified by analyzing the correlation between the vectors. This accounts for any shifts in the mean between frames making it more appealing than the Euclidean distance measure. The correlation coefficient between the input and tracker similarity vectors is given by,

$$r_{xy} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{(n-1) \sigma_x \sigma_y} \quad (4)$$

where x and y are the input and tracker feature vectors respectively, \bar{x} and \bar{y} are the sample means, σ_x and σ_y are the sample standard deviations, and n is the number of people in the face database. Using the correlation measure to properly differentiate people, the feature vectors of the tracking units are regularly updated to provide the most current representation of the face. As the face exhibits changes in face pose angle and expression, the corresponding similarity scores change as well. Updating the features allows the system to provide better results during these variant conditions, which correctly assumes that pose, lighting, and face expression changes do not occur instantaneously.

3. Face Recognition

The Kalman face tracker provides a fast and reliable means to continuously locate the faces in the input video frames. The tracking system is then used to support a higher level face recognition system containing a database of a large number of individuals. By tracking the faces in the scene, the face recognition system can perform multiple tests on an unknown face to provide more reliable results. In addition, the tracking system allows the recognition process to be performed as a separate activity, permitting it to be computationally expensive while still maintaining the real-time ability of the overall system. In this paper, the feature extraction for face recognition is performed using an Adaptively Weighted Modular Principal Component Analysis (AWMPCA) approach. To provide higher recognition accuracy, AWMPCA enhances the MPCA technique by selectively focusing only on regions of the face containing significant details. The system dynamically determines the areas of special interest by analyzing the variance of each module. Modules with higher variance scores are given more consideration in the classification process by assigning proportional weight factors. Therefore, face regions such as the eyes and nose contain a greater impact on the overall face recognition result. The training portion of AWMPCA is performed in the same manner as MPCA, in which the training images are divided into modules and PCA is performed on each module. However, the classification process differs from the MPCA technique due to the variance and weighting calculations.

4. Experimental Results

The face tracking and recognition system was implemented in C++ using the OpenCV libraries. A system was created to capture real-time video frames from an IP surveillance camera and perform the Kalman tracking, Modular PCA based face feature extraction and low level matching techniques. Multiple tests were then performed on the overall system to evaluate its capabilities. The testing procedures focus on the main contributing algorithms presented in this paper, including the Kalman tracker and low-level matching functionality of the system. The results of the

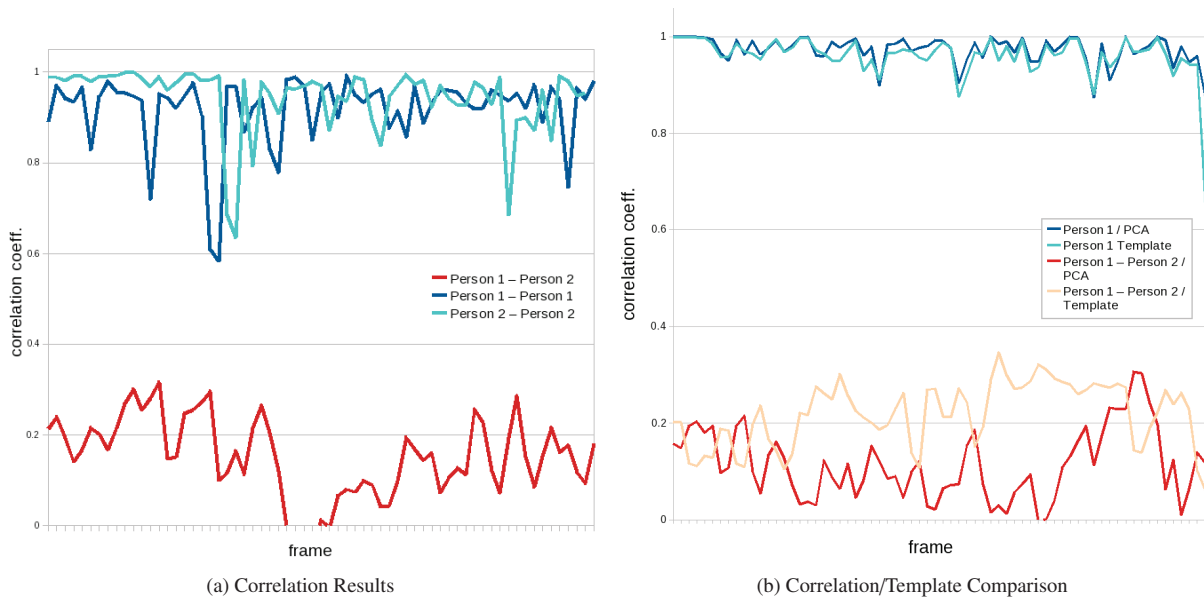


Figure 2: Feature vector results.

high level face recognition system are primarily dependent on the chosen recognition method. The tracking system is intended to serve as an added preprocessing technique designed to increase the exposure of a person's face to the recognition system, which inherently increases the overall accuracy of the given system. Therefore the system is tested to verify that the feature vector correlation matching method has the ability to distinguish between multiple faces. This correlation method is then compared with the traditional face template matching method used in [2]. After determining the multi-face capabilities of the system, we then analyze the Kalman tracker based on its performance compared to the standard face detection methods.

The proposed system was first tested for its low-level face matching capability. The test consisted of providing an input video containing two people, extracting the features of each face in every frame, and calculating correlation scores. The correlation scores were divided into three categories, the inter-frame correlation between person 1 and person 1 (r_{11}), person 2 and person 2 (r_{22}), and person 1 and person 2 (r_{12}). For example, r_{11} would be calculated by performing the correlation computation on person 1's feature vector from frame 1 with person 1's feature vector from frame 2. Figure 2a shows the test results of the correlation measurements taken during a video sequence containing two people with varying pose angle and expression. The bottom red line is a record of the correlation coefficient r_{12} . The results show that there is little correlation between the feature vectors of the two people and the maximum correlation coefficient is around 0.3. On the other hand, the two blue lines in the upper portion of the graph illustrate the correlation coefficients between the same person over two frames, r_{11} and r_{22} . The coefficient magnitudes are much higher with a minimum value of roughly 0.6. Clearly, correlation between the two feature vectors is adequate for matching small sets of faces. It is also noted that the correlation scores for r_{11} and r_{22} significantly decrease on certain frames. This corresponds to quick pose angle changes by each face in the frame. In this particular example, the face changes its pose angle which results in a significant decrease in correlation. Since the correlation score remains above the selected threshold, the new feature vector is accepted by the tracker. This results in a much higher correlation in the following frame, allowing the system to be adaptive to pose and expression variations. Next, the system's feature based matching system is compared to the standard template matching method used in [2]. Again, a video containing two people is processed by the system. The correlation scores are generated using the feature vectors and the raw face image data (face template). The results are shown in figure 2b, where the dark blue and light blue lines represent the correlation between one person for feature based and template based methods respectively. The red and beige lines represent the correlation between two people using feature based and template based methods respectively. The results show that for identifying similar faces, both the template method and the feature based methods perform roughly the same way. However, when correlating non identical people, the feature based method provides improved results by recording lower scores. Finally, the Kalman tracking abilities of the system were employed.

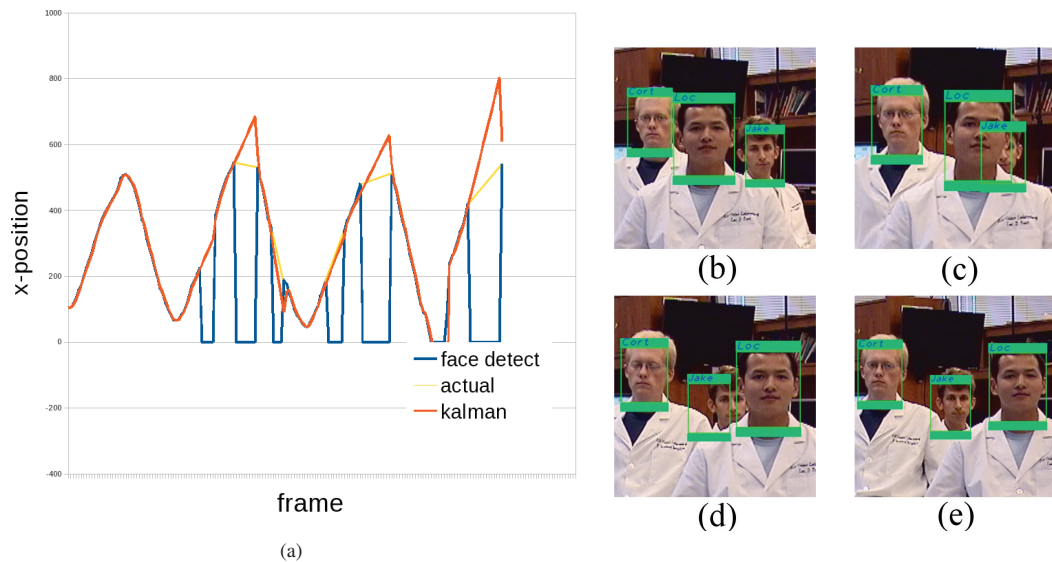


Figure 3: a) Comparison of Kalman face tracking and standard face detection. b) Initial frame with no occlusions. c) Occlusion with Kalman prediction. d) Partial occlusion with Kalman prediction. e) Kalman correction.

A video input of one person moving periodically from the left portion of the screen to the right while experiencing face occlusions and pose changes was applied to the proposed system. The system predicted the x-coordinate of the face first by performing face detection, followed by an interpretation by the Kalman tracker. The results are shown in figure 3a. As expected, the face detector fails during portions of occlusion and poor face pose angle, while the tracker attempts to predict the location in a linear manner. Figure 3(b-e) shows the final output of the Kalman face tracking and recognition system. As the figure shows, the system is able to track multiple people even with the loss of input data, which in this case is due to occlusions as shown in Fig. 3c.

5. Conclusions

In this paper we have shown a face tracking system capable of predicting the movements of multiple people based on a low-level matching system. The tracker distinguishes between multiple faces by generating feature vectors obtained from the principal components of a generic face database. This method was shown to provide basic classification through a correlation metric which is needed to perform multiple face tracking. The tracking system can also predict the location of faces in the event of occlusions and poor pose angle. This system provides a sound basis for a high-level face recognition system, in which the tracker maintains a record of close matches with a larger database.

References

- [1] P. Sankaran and K. V. Asari, "Nonlinear manifold representation of a face image for classification," *International Conference on Artificial Neural Networks in Engineering - ANNIE 2008*, vol. 18, Nov. 2008, pp. 705-710.
- [2] Z. Shaik and K. V. Asari, "A robust method for multiple face tracking using Kalman filter," *Applied Imagery Pattern Recognition Workshop*, 2007, pp. 125-130.
- [3] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, May 2004, pp. 137-154.
- [4] Kalman, "A new approach to linear filtering and prediction problems," *Transaction of the ASME - Journal of Basic Engineering*, 1960, pp. 33-45.
- [5] G. Welch and G. Bishop, "An introduction to the Kalman filter," University of North Carolina at Chapel Hill, 2006, <http://www.cs.unc.edu/~welch/kalman/kalmanIntro.html>.
- [6] B. Moghaddam, A. Pentland, T. Starner, "View-based and modular Eigen spaces for face recognition," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, June 2004.
- [7] R. Gottumukkal, K. V. Asari, "An improved face recognition technique based on modular PCA approach," *Pattern Recognition Letters*, vol. 25, 2004, pp. 429-436.
- [8] P. Phillips, P. Flynn, T. Scruggs, "Overview of the Face Recognition Grand Challenge," *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.