International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:268
Vol.2, No.5, May 2013 DOI:10.15693/ijaist/2013.v2i5.54-58

Study on Video-Based Face Recognition

Sayeesh[1] , Dr. Nagaratna P. Hegde[2]

[1] *Asst. Professor, Dept. of CS & E, Alva's Institute of Engg. & Tech., Shobhavana Campus, Mijar,* Moodbidri – 574 225, Karnataka State.

[2] *Professor, Dept. of CS & E, Vasavi College of Engg., Ibrahimbhag, Hyderabad – 31, Andhra Pradesh.*

***Abstract*- Video-based face recognition has been one of the hot topics in the field of image processing in the last several decades. Face recognition in video has received significant attention. Not only the wide range of commercial and law enforcement applications, but also the availability of feasible technologies after several decades of research contributes to the trend. Compared to traditional face analysis, video based face recognition has advantages of more abundant information to improve accuracy and robustness, but also suffers from large scale variations, low quality of facial images, illumination changes, pose variations and occlusions. Although current face recognition systems have reached a certain level of maturity, their development is still limited by the conditions brought about by many real applications. For example, recognition images of video sequence acquired in an open environment with changes in illumination and/or pose and/or facial occlusion and/or low resolution of acquired image remains a largely unsolved problem. In other words, current algorithms are yet to be developed. This paper provides an up-to-date study on video-based face recognition research.**

***Keywords*- Face recognition, Video-based, Survey**

I. INTRODUCTION

Humans make use of face as an important cue for identifying people. This makes automatic face recognition very crucial from the point of view of a wide range of commercial and law enforcement applications. Face recognition is a task that humans perform routinely and effortlessly in their daily lives. Wide availability of powerful and low cost desktop and embedded computing systems has created an enormous interest in automatic processing of digital images and videos in a number of applications, including biometric authentication, surveillance, human-computer interaction, and multimedia management. Research and development in automatic face recognition follows naturally.

In recent years, face recognition is always an active topic in the field of biometrics. Research in face recognition is motivated not only by the fundamental challenges this recognition problem poses but also by numerous practical applications where human identification is needed. Face recognition, as one of the primary biometric technologies, became more important owing to rapid advances in technologies such as digital cameras, the Internet and mobile devices, and increased demands on security. Face recognition has several advantages over other biometric technologies: It is natural, nonintrusive, and easy to use. Among the six biometric attributes considered by Hietmeyer [1], facial features scored the highest compatibility in a Machine Readable Travel Documents (MRTD) [2] system based on a number of evaluation factors, such as enrollment, renewal, machine requirements, and public perception.

A face recognition system is expected to identify faces present in images and videos automatically. It can operate in either or both of two modes: (1) face verification (or authentication), and (2) face identification (or recognition). Face verification involves a one-to-one match that compares a query face image against a template face image whose identity is being claimed. Face identification involves one-to-many matches that compare a query face image against all the template images in the database to determine the identity of the query face. Another face recognition scenario involves a watch-list check, where a query face is matched to a list of suspects (one-to-few matches).Compared to traditional face recognition in still images, video based face recognition has great advantages listed as follows. Firstly, videos contain more abundant information than a single image. As a result, more robust and stable recognition can be achieved by fusing information of multi frames. Secondly, temporal information becomes available to be exploited in videos to improve the accuracy of face recognition. Finally, multi poses of faces in videos make it possible to explore shape information of face and combined into the framework of face recognition. However, video based face recognition is also a very challenging problem, which suffers from low quality facial images, illumination changes, pose variations, occlusions and so on.

II. FACE DETECTION

Face detection is the first stage of a face recognition system. A lot of research has been done in this area, most of that is efficient and effective for still images only. So could not be applied to video sequences directly. In the video scenes, human faces can have unlimited orientations and positions, so its detection is of a variety of challenges to researchers.

Generally, there are three main processes for face detection based on video. At first, it begins with frame based detection. During this process, lots of traditional methods for still images can be introduced such as statistical modeling method [5], neural network-based method[6], SVM-based method [7], HMM method [8], BOOST method [9] and color-based face detection [10] etc. However, ignoring the temporal information provided by the video sequence is the main drawback of this approach. Secondly, integrating detection and tracking, this says that detecting face in the first frame and then tracking it through the whole sequence. Since detection and tracking are independent and information from one source is just in use at one time, loss of information is unavoidable. Finally, instead of detecting each frame, temporal approach exploits temporal relationships between the frames to detect multiple human faces in a video sequence. In general, such method consists of two phases, namely detection and prediction and then update-tracking. This helps to stabilize detection and to make it less sensitive to thresholds compared to the other two detection categories.

III. FACE RECOGNITION

Face recognition is a visual pattern recognition problem. There, a face as a three-dimensional object subject to varying illumination, pose, expression and so on is to be identified based on its two-dimensional image (three-dimensional images e.g., obtained from laser may also be used).A face recognition system generally consists of four modules: detection, alignment, feature extraction, and matching, where localization and normalization (face detection and alignment) are processing steps before face recognition (facial feature extraction and matching) is performed. Face detection segments the face areas from the background.

In the case of video, the detected faces may need to be tracked using a face tracking component. Face alignment is aimed at achieving more accurate localization and at normalizing faces there by whereas face detection provides coarse estimates of the location and scale of each detected face. Facial components, such as eyes, nose, and mouth and facial outline, are located; based on the location points, the input face image is normalized with respect to geometrical properties, such as size and pose, using geometrical transforms or morphing. The face is usually further normalized with respect to photometrical properties such illumination and gray scale.

After a face is normalized geometrically and photo metrically, feature extraction is performed to provide effective information that is useful for distinguishing between faces of different persons and stable with respect to the geometrical and photometrical variations. For face matching, the extracted feature vector of the input face is matched against those of enrolled faces in the database; it outputs the identity of the face when a match is found with sufficient confidence or indicates an unknown face otherwise.

Face recognition results depend highly on features that are extracted to represent the face pattern and classification methods used to distinguish between faces whereas face localization and normalization are the basis for extracting effective features.

IV. VIDEO-BASED FACE RECOGNITION.

Definition: Video based face recognition in image sequences has gained increased interest based primarily on the idea expressed by psychophysical studies that motion helps humans recognize faces, especially when spatial image quality is low.

Although face recognition has been an active research topic for decades, the traditional recognition algorithms are all based on static images. However, during the last years face recognition in image sequences has gained increased interest based primarily on the idea expressed by psychophysical studies that motion helps humans recognize faces, especially when spatial image quality is low.

Video-based face recognition systems consist of three modules: a detection module, a tracking module and a recognition module. Given a frame of a video sequence, the detection module locates face candidates, while the tracking module finds the exact position of facial features in the current frame based on an estimate of face or feature locations in the previous frame(s). The recognition module identifies or verifies the face, integrating information from previous frames.

In the detection module, motion and/or skin color information may be used for segmenting the face from the background and locate candidate face regions. Face detection techniques similar to those applied for still images are then employed to find the exact location of faces in the current frame, thus initiating face and facial feature tracking. Face tracking techniques include head tracking, where the head is viewed as a rigid object performing translations and rotations, facial feature tracking, where facial features deformations due to facial expressions or speech are viewed as non rigid transformations limited by the head anatomy, and methods tracking head and features. Face and facial feature tracking is sometimes used to reconstruct the 3D shape of the face, which is subsequently used for enhancing face recognition.

The main problem of video-based face recognition is low quality of images in video sequences, while the unquestionable advantage is the abundance of information. This enables the selection of the frames that will be used for recognition and the reuse of recognition information obtained in precedent frames. Also, temporal continuity allows tracking of facial features, which can help in compensating pose or expression variations, while motion, gait and other features may enhance the performance of face recognition, Moreover, the simultaneous comprehensive exploitation of spatiotemporal cues results in increased tracking and identification accuracy. Video based techniques are ideal for surveillance or facility monitoring applications

The whole procedure of video based face recognition is shown in Fig. 1. Related to applications, we can divide video based face recognition methods into two categories: video-image based methods and video-video based methods. The first category can be seen as an extension of still image based face recognition. The second category is more complicated with more abundant solutions proposed. In the following, we will describe both of the categories in detail.

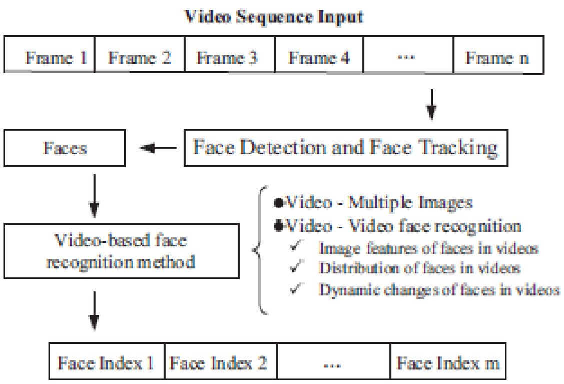

Fig. 1. Process of face recognition in video.

A.Video-image face recognition

Video-image face recognition can be seen as an extension of still image based face recognition. The input of the system is videos while the databases are still face images. Compared to traditional still image based face recognition, how to explore the multi-frame information of the input video is the key to enhance the performance.

One solution is based on frame selection. Faces in videos are tracked by tracking algorithms and those high quality face images of better resolution, pose and clarity are selected for matching based on still image based methods. Eigenfaces [11] and fisherfaces [12] are two of the essential basic techniques in this category. In [13], Satoh proposed a straightforward extension of the traditional eigenface and fisherfaces methods, by introducing a new similarity measure for matching video data. The distance between videos was calculated by considering the smallest distance between frame pairs (one from each video), in the reduced feature space. Raffaella Lanzarotti [14] extracted the 24 facial feature points from the color image, then extracted the feature of the points by Gabor, and the recognition the faces by compare the similarity of points' feature.

Another strategy is based on multi-frame fusion. Rainer Stiefelhagen et al. [15] made use of 3 different metric model to fuse recognition results of different frames. Pascal Frossard et al. [16] adopted the semi-supervised learning and graph theory to convert the problem of face recognition to an optimization task. Video-Based Face Recognition: State of the Art 3 To deal with occlusions, J. Aggarwal et al.[17] adopted patch information from multi-frames to obtain a complete face model for matching. In addition, Jae Young Choi et al.[18] made use of low resolution subspace of high resolution image sets to deal with problem of different resolutions of input videos and databases.

B. Video-video face recognition

Compared to video-image based methods, both the system input and the database in this category are in the form of videos, which is a more difficult problem to solve. Based on the state of the arts, there are mainly three types of solutions of this problem, which are listed as follows:

- Based on feature vector extracted from video input.
- Based on probability density function or manifold to depict the distribution of faces in videos.
- Based on generative models to describe dynamic variance of face in images.

i. Feature vector based methods

The basic idea of this solution is to extract feature vectors from input videos, which are used to match with all the videos in the database.

Horst Eidenberger [19] proposed a Kalman filter-based method to describe invariant face features in videos based on a compact vision model, which achieves high performance in the UMIST database. Park et al.[20] created multiple face templates for each class of face in database according to the video information, dynamic fuse the multiple templates as the feature. Lapedriza et al.[21] builded PCA feature subspaces for the faces of the input and database, which achieves recognition based on the distance between the subspaces measured by geometric angles.

In order to remove the effect of light, gesture and facial expressions, Fukui and Yamaguchi [22] projected the feature space to the constraint subspace. The disadvantages of this method are that they ignore the global probable distribution of each category, and parameters are based experiences or experiments. Ajmal Mian [23] proposed a unsupervised video-based method. Faces from a video sequence are automatically clustered based on the similarity of their local features and the identity is decided on the basis of best temporally cohesive cluster matches.

Since the pose, illumination and expression of face are non-linear in feature space, Wolf and Shashua [24] mapped the feature space into the kernel Hilbert space, combined with Support Vector Machine (SVM) to improve recognition performance. Fan et al.[25] extracted the geodesic distance as the feature to depict the position relations in the manifold space, and use HAC (Hierarchical Agglomerative Clustering) to obtain K samples. Similar to [26] dual subspace probabilistic model is obtained as similarity score to recognize face.

This kind of solution makes use of multi frames of the input video to obtain a discriminant feature representation. However, the spatial information of input videos is neglected, which limits the performance of feature vector based approaches.

ii. Distribution of faces in videos

The main idea of this category is to treat faces in videos as random variables of certain probability. The similarities of faces are measured by similarity of corresponding probability density distributions.

In [27], each faces in the database and input video are modeled with GMM, and Kullback-Leibler divergence is measured as the similarity measurement to achieve recognition. Arandjelovi’c et al.[28] made use of kernel-based methods to map low-dimensional space to high dimensional space, and then use low dimensional space of linear methods (such as PCA) to solve complex nonlinear problems in the high-dimensional space. Zhou et al.[29] mapped the vector space into RKHS (Rep reducing Kernel Hilbert Space) by kernel based methods to calculate the distance between the probability distribution.

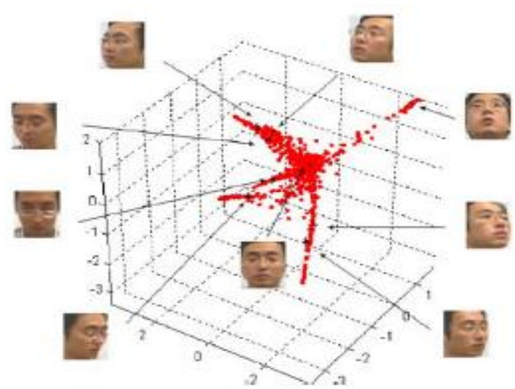

Fig. 2. LLE applied to a sequence of face images corresponding to a single person arbitrarily rotating his head [30].

In [31-33], a multi-view dynamic face model is built to achieve face recognition. Firstly, dynamic face model are constructed including a 3D model, a texture model and an affine change model. Then a Kalman filter is adopted to obtain the shape and texture, which builds a segmented linear manifold for each single person with the face texture reduced by KDA (Kernel Discriminate Analysis). Face recognition is achieved in the following by trajectory matching. However, the 3D model estimation requires a lot of multi-angle images and a larger complexity computational.

Wang et al.[34,30] proposed a incremental online learning model for face recognition in videos. This method use a defined face model to learn new face contour model online, then use a linear model and feature space transformation matrix to generate the face manifold, as shown in Fig. 2. Minyoung Kim et al.[35] integrated face tracking and recognition and employed some priori knowledge. They obtain 70% recognition rate in YouTube and 100% in Honda/UCSD database [36,37].

This kind of solution is much better than the feature vector based solution, which makes use of probability theory to enhance the performance. However, the dynamic change information of faces in videos is neglected, which has potential to improve the video based face recognition.

iii. Dynamic changes of faces in videos

The temporal information in video sequences enables the analysis of facial dynamic changes and its application as a biometric identifier for person recognition [38].

Matta et al. proposed a multi-modal recognition system [39, 40]. They successfully integrated the facial motion information with mouth motion and facial appearance by taking advantage of a unified probabilistic framework.

Some strategies have been developed to integrate tracking and recognition into a single framework. Lee et al. [37] developed the probabilistic appearance manifold approach for tracking and recognition using video sequences. Bayesian inference was employed to include the temporal coherence of human motion in the distance calculation. And they replaced the conditional probability by using the joint conditional probabilities, which were recursively estimated using the transitions between sub-manifolds. In [41], Matta and Dugelay presented a person recognition system that exploited the unconstrained head motion information extracted by tracking a few facial landmarks in the image plane. In [42], Huang and Trivedi developed a face recognition system by employing HMMs for facial dynamic information modeling in videos. Each covariance matrix was gradually adapted from a global diagonal one by using its class-dependent data in training algorithms. Afterwards, Liu and Cheng [43] successfully applied HMMs for temporal video recognition (as illustrated in Fig. 3) by improving the basic implementation of Huang and Trivedi. Each test sequence was used to update the model parameters of the client in question by applying a maximum a posteriori (MAP) adaptation technique.

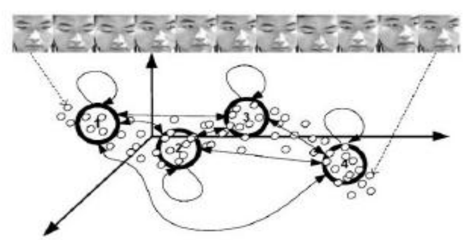

Fig. 3. Example of a hidden Markov model temporally applied to video sequences [43].

V. CONCLUSION

In this paper, we presented some major issues on video based face recognition. Two categories of video based face recognition methods are surveyed and analyzed. We can see that video-image based methods only exploit physiological information of the face while the video-video based methods have more information to be exploited. It is evident that video based face recognition has great potential to make progress and be adopted in real application.

REFERENCES

[1] R. Hietmeyer, “Biometric identification promises fast and secure processing ofairline passengers”. The International Civil Aviation Organization Journal, vol.55, no.9, pp.10–11, 2000.
[2]Machine Readable Travel Documents (MRTD).

http://www.icao.int/mrtd/overview/overview.cfm.

[3] T. Kanade. "Picture Processing by Computer Complex and Recognitionof Human Faces". Ph.D. thesis,Kyoto University, 1973.

[4] W.Y. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, Face Recognition: A Literature Survey, ACM Computing Surveys,Vol:35, 2003.

[5] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 696-710, July 1997.

[6] H.A. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 23-38, Jan. 1998.

[7] B. Heisele, T. Poggio, and M. Pontil, "Face Detection in Still Gray Images," A.I. memo AIM-1687, Artificial Intelligence Laboratory, MIT, 2000.

[8] Ti-QiongXu; Bi-Cheng Li; Bo Wang, Face detection and recognition using neural network and hidden Markov models, Proceedings of the 2003 International Conference on Neural Networks and Signal Processing, pp. 228 - 231,2003.

[9] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," Proc. Conf. Computer Vision and Pattern Recognition, pp. 511-518, 2001.

[10] R.L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face Detection in Color Images," Proc. Int'l Conf. Image Processing, pp. 1046-1049, 2001.

[11]. M.A. Turk and A.P. Pentland. Face recognition using eigenfaces. In Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on, pages 586–591. IEEE, 1991.

[12]. P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces:Recognition using class specific linear projection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 19(7):711–720, 1997.

[13]. S. Satoh. Comparative evaluation of face sequence matching for content-based video access. In Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, pages 163–168. IEEE, 2000.

[14]. S. Arca, P. Campadelli, and R. Lanzarotti. A face recognition system based on automatically determined facial fiducial points. Pattern recognition, 39(3):432–443, 2006.

[15]. J. Stallkamp, H.K. Ekenel, and R. Stiefelhagen. Video-based face recognition on real-world data. In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, pages 1–8. IEEE, 2007.

[16]. E. Kokiopoulou and P. Frossard. Video face recognition with graph-based semisupervised learning. In Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on, pages 1564–1565. IEEE, 2009.

[17]. C. Hu, J. Harguess, and JK Aggarwal. Patch-based face recognition from video.In Image Processing (ICIP), 2009 16th IEEE International Conference on, pages 3321–3324. IEEE, 2009.

[18]. J.Y. Choi, Y.M. Ro, and K.N. Plataniotis. Feature subspace determination in video-based mismatched face recognition. In Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on, pages 1–6. IEEE, 2008.

[19]. H. Eidenberger. Kalman filtering for pose-invariant face recognition. In Image Processing, 2006 IEEE International Conference on, pages 2037–2040. IEEE.

[20]. U. Park, A.K. Jain, and A. Ross. Face recognition in video: Adaptive fusion of multiple matchers. In 2007 IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8. IEEE, 2007.

[21]. A. Lapedriza, D. Masip, and J. Vitria. Face verification using external features.In Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on, pages 132–137. IEEE, 2006.

[22]. K. Fukui and O. Yamaguchi. Face recognition using multi-viewpoint patterns for robot vision. Robotics Research, pages 192–201, 2005.

[23]. A. Mian. Unsupervised learning from local features for video-based face recognition. In Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on, pages 1–6. IEEE, 2008.

[24]. L. Wolf and A. Shashua. Kernel principal angles for classification machines with applications to image sequence interpretation. 2003.

[25]. W. Fan and D.Y. Yeung. Locally linear models on face appearance manifolds with application to dual-subspace based classification. 2006.

[26]. B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recognition. Pattern Recognition, 33(11):1771–1782, 2000.

[27]. O. Arandjelovi´c and R. Cipolla. Face recognition from face motion Manifolds using robust kernel resistor-average distance. In Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on, pages 88–88. IEEE.

[28]. O. Arandjelovi´c, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell. Face recognition with image sets using manifold density divergence. 2005.

[29]. S.K. Zhou and R. Chellappa. From sample similarity to ensemble similarity: Probabilistic distance measures in reproducing kernel hilbert space. IEEE transactions on pattern analysis and machine intelligence, pages 917–929, 2006.

[30]. W. Fan, Y. Wang, and T. Tan. Video-based face recognition using bayesian inference model. In Audio-and Video-Based Biometric Person Authentication, pages 122–130. Springer, 2005.

[31]. Y. Li, S. Gong, and H. Liddell. Modelling faces dynamically across views and over time. In Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, volume 1, pages 554–559. IEEE, 2001.

[32]. Y. Li, S. Gong, and H. Liddell. Video-based online face recognition using identity surfaces. ratfg-rts, page 0040, 2001.

[33]. Y. Li, S. Gong, and H. Liddell. Constructing facial identity surfaces in a nonlinear discriminating space. 2001.

[34]. L. Liu, Y. Wang, and T. Tan. Online appearance model learning for video-based face recognition. In 2007 IEEE Conference on Computer Vision and Pattern Recog nition, pages 1–7. IEEE, 2007.

[35]. M. Kim, S. Kumar, V. Pavlovic, and H. Rowley. Face tracking and Recognition with visual constraints in real-world videos. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pages 1–8. IEEE, 2008.

[36]. K.C. Lee, J. Ho, M.H. Yang, and D. Kriegman. Video-based face recognition using probabilistic appearance manifolds. In Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, volume 1, pages I–313. IEEE.

[37]. K.C. Lee, J. Ho, M.H. Yang, and D. Kriegman. Visual tracking and Recognition using probabilistic appearance manifolds. Computer Vision and Image Under- standing, 99:303–331, 2005.

[38]. F. Matta and J.L. Dugelay. Person recognition using facial video information: A state of the art. Journal of Visual Languages & Computing, 20(3):180–187, 2009.

[39]. U. Saeed, F. Matta, and J.L. Dugelay. Person recognition based on head and mouth dynamics. In Multimedia Signal Processing, 2006 IEEE 8th Workshop on, pages 29–32. IEEE, 2006.

[40]. F. Matta and J.L. Dugelay. Video face recognition: a physiological and behavioural multimodal approach. In Image Processing, 2007. ICIP 2007. IEEE International Conference on, volume 4, pages IV–497. IEEE, 2007.

[41]. F.M.J.L. Dugelay. Person recognition using human head motion information. In Articulated motion and deformable objects: 4th international conference, AMDO 2006, Port d'Andratx, Mallorca, Spain, July 11-14, 2006: proceedings, volume 4069, page 326. Springer-Verlag New York Inc, 2006.

[42]. K.S. Huang and M.M. Trivedi. Streaming face recognition using multicamera video arrays. In Pattern Recognition, 2002. Proceedings. 16th International Conference on, volume 4, pages 213–216. IEEE, 2002.

[43]. X. Liu and T. Cheng. Video-based face recognition using adaptive hidden markov models. In Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, volume 1, pages I–340. IEEE.