International Academy of Engineering and Medical Research, **2017**
Volume-2, ISSUE-4
Published Online April 2017 in IAEMR (http://www.iaemr.com)

A Review on Multimodal Face Pair Matching System

[1]Ms. Divya K. Sawant
PG Student
Department of Computer Engineering
K. K. Wagh Institute of Engineering, Education and Research, Nashik, Maharashtra, India
divyasawant93@gmail.com

[2]Prof. Nitin M. Shahane
Associate Professor
Department of Computer Engineering
K. K. Wagh Institute of Engineering, Education and Research, Nashik, Maharashtra, India
nmshahane@kkwagh.edu.in

Abstract — A face recognition system can be thought as an identification or verification system. Face pair matching is a challenging task which aims to determine whether two face images represent the same person. Many times the face images being compared are having complicated facial variations and limited expressive information which becomes a difficult problem. To address these issues, the proposed approach concentrate on exploiting an additional set of face images called as cohort set. The proposed system is able to perform multimodal face pair matching using cohort information. The inputs to this system are pair of face images to be matched and the set of cohort face images. All cohort images are ranked separately based on pair of face images to generate two lists. Then cohort information are extracted from two sorted cohort list and combined with direct matching score of the two input face images to form modality. The final decision of matching is made by fusing all available face modalities.

Keywords — Cohort set, Face recognition, Multimodal fusion.

I. INTRODUCTION

The recognition of human faces has been a long standing problem in computer vision and pattern recognition. A face recognition system can be thought as an identification system, a verification expert, or a pair matching system. There is a pre-enrolled face database in identification and verification system, but there is no pre-enrolled template database in face pair matching system. The role of identification system is to identify the given probe face image comes from which subject in the pre-enrolled database, while the role of verification system is to verify which person from among a set of people the picture represents, if any.

Face pair matching system differs from these two systems. In this system, the input is pair of face images and the goal is to decide whether the face image represent same person or not. If they are from the same person, then it is called a genuine pair otherwise it is called an impostor pair. [1]

Applications of face recognition are in law enforcement and video surveillance [2], HCI [3], demography estimation [4], smart cards, information security purpose. But among all these applications, automatically recognizing humans by analysing their faces has been one of the most difficult problems in computer vision and pattern recognition. This task is extremely hard because of limited expressive information i.e. only photometric information is available in two face images. Facial descriptors have been devised in handling different sources of facial variations which includes patch-based LBP codes, learning-based descriptor, and discriminant face descriptor. Cosine similarity metric learning is a similarity measure for better matching.

Single image per person require less effort for collecting, low cost for storing and processing, so one cannot have enough information to predict the result in test samples. This issue results in exploiting an additional set of samples called as cohort set. Utilizing information from cohort set increases performance of face pair matching task. [5]

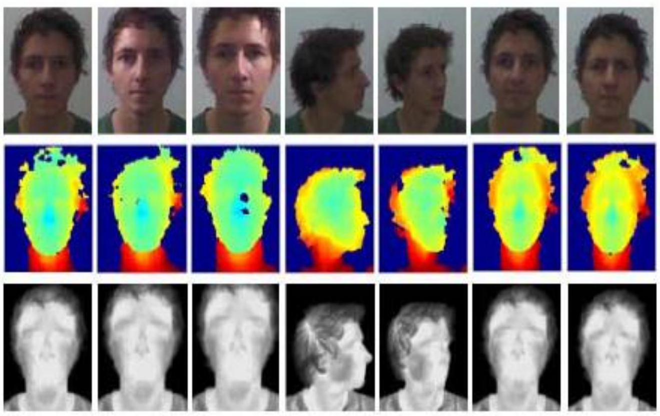

Fig. 1. Several face images of RGB, Depth, Thermal modalities resp.

Complicated facial expressions are handled by face modalities like depth and thermal. For face pair matching task, fusing of different modalities results in reduction of diverse corrupting factors. Therefore the main focus is on multimodal cohort based face pair matching system.

In the below sections we are going to discuss about related work done for the proposed research area. We refer some existing research paper for completing this task. It is given as follow:

II. RELATED WORK

Schroff et al. [6] proposed a data driven approach for face similarity. This approach is based on using Doppelganger list comparison. In this method an image of a face is described by an ordered list of identities from library. For given input pair of probes, the look-alike ranked list for each input probe from the Library is computed. Then the similarity between the two input probes is determined by the similarity between these two ranked lists. Ranking is based on identity. These ranked Doppelganger lists are similar for similar people. But the major disadvantage of this method is that direct comparison between faces is required only in similar imaging conditions. This method is robust to pose variation means performance remains high across full range of poses. Pixel wise measures are problematic to use across poses.

Yin et al. [7] proposed an Associate-Predict Model for face recognition. This model handles the issue of recognizing similarity measure between two human faces under significantly different condition like pose, illumination and expression. This model is built on an extra generic identity data set, in which each identity contains multiple images with large intra-personal variation. When two faces under significantly different conditions, suppose non-frontal and frontal faces, this model first “associate” one input face with alike identities from the generic identity date set. Then using the associated faces, one can “predict” the appearance of one input face under the setting of another input face. And can also “predict” whether two input faces are from the same person or not. By using an extra data set i.e. “memory” and the “associate-predict” model, the intrapersonal variation can be effectively handled. The generalization ability of this model is improved by switching mechanism in which it can adaptively switch between the direct matching and associate-predict based matching. This model may be less discriminative due to the limited size of identity data set.

John Wright et al. [8] proposed Robust Face Recognition via Sparse Representation. Sparse signal representation offers the key to addressing the problem of automatically recognizing human faces from frontal views with varying expression, illumination, occlusion and disguise. Based on a sparse representation, a general classification algorithm for image-based object recognition is developed. Feature extraction and robustness to occlusion issues in face recognition are handled by this new framework. For feature extraction, if sparsity in the recognition problem is properly harnessed, the choice of features becomes less important than the number of features used. Determining whether the number of features is sufficiently large and whether the sparse representation is correctly computed is a critical part.

This framework can handle errors due to occlusion and corruption uniformly by utilizing the fact that these errors are often sparse with respect to the standard pixel basis. It achieves striking recognition performance for severely occluded or corrupted images by a simple algorithm with no special engineering. The full potential of sparsity in robust object detection and recognition together is yet to be uncovered.

Weihong Deng et al. [9] proposed the Extended Sparse Representation-based Classifier(ESRC). This method applies intraclass variant dictionary to represent the possible variation between the training and testing images. ESRC has better generalisation ability than SRC for undersampled face recognition under variable expression, illuminations, disguises and ages. ESRC for undersampled face recognition is effective even if there is only one training image per subject. SRC always requires a rich set of training images of each subject that can span the facial variation of that subject under testing condition. However many important applications on law enforcement and homeland security can only offer a few or even single facial images per subject. This is often called under sampled problem of face recognition which is a challenge. Use of intraclass variant dictionary improves the accuracy of face recognition which is based on sparse representation. Local features like gabor wavelet and LBP instead of pixel feature can largely improve sparse representation leads to best performance. For local features dimension reduction such as random projections would lose useful information.

Meng Yang et al. [10] proposed Sparse Variation Dictionary Learning (SVDL) method. Face recognition with the single training sample per person (STSPP) is a very challenging problem due to lack of information to predict the variations in the query sample. SVDL method learns a sparse variation dictionary from a generic training set to improve face recognition performance with a STSPP. The SVDL is adaptive to the gallery set by simultaneously learning a projection from the gallery set to the generic set. Hence for relation between generic and gallery set can be exploited. Learned sparse variation dictionary can more effectively help the single training sample to represent query image.

Liao et al. [11] proposed partial face recognition for alignment-free sparse representation approach. Partial faces mostly arise from images captured by surveillance cameras or handheld device like mobile phones. This approach does not require face alignment by eye coordinates or any other fiducial points. Alignment-free face representation method is based on

International Academy of Engineering and Medical Research, 2017
Volume-2, ISSUE-4
Published Online April 2017 in IAEMR (http://www.iaemr.com)

multi-keypoint descriptors (MKD). Superior methods recognise holistic as well as partial faces without requiring alignment. Commercial Off-the shelf (COTS) face recognition system is not able to handle the general partial face recognition problem since they need to align faces by facial landmarks that may occluded. Face VACS needs localisation of two eyes and PittPatt determines predefined landmarks.

Tistarelli et al. [12] developed an approach called picture specific cohort score normalisation. This approach extracts discriminative cohort score coefficients from a pool of sorted cohort samples using polynomial regression. It enhances the robustness of face matching under challenging conditions. Face representation of specific subject allows to maximize the discrimination capability for each individual. Computation model is developed by this approach which gives peculiar information for each subject. By making the analysis for each user any identity claim is adapted to the user or most simply to the model associated with the user. This is achieved by using a polynomial regression-based cohort normalisation to face pair matching scores. After that normalisation is performed for variations in the score distribution appears in the two faces in a given pair of images. This subject specific cohort normalisation has been developed for single image holistic face matching, but it cannot do video-to-video face recognition. Component based face recognition is also yet to be uncovered.

Wolf et al. [13] proposed learned discriminative model by combining multiple descriptors and learned background statistics. The local patch similarities statistics are captured by combining multiple face image descriptors. Then from unlabelled background samples, image similarities are better evaluated. There are number of effective similarity measures. Classification performance is improved by employing a unique pair matching pipeline from labelled background samples. This system performs well in a multi label face recognition problem. Scalability is limited for this method because of inherent efficiency constraints.

Li et al. [14] developed a probabilistic elastic matching method. In this method, firstly in training set all spatial appearance features are extracted from face images. Means we build a Gaussian Mixture Model (GMM). This is also called as Universal Background Model (UBM) or UBM-GMM. For given face pair which are represented by its spatial appearance features for each Gaussian component in the UBM-GMM. Then a pair of features that gives highest probability is to be found. This pair of features which gives highest probability is called a corresponding feature pair. Then calculate difference vector for all these corresponding feature pairs and concatenate all vectors to form long vector. After that this vector is fed into SVM classifier for prediction.

SIMONYAN et al. [15] developed an approach which is based on fisher vector representation of a particular face image. This approach includes SIFT features and Fisher vector encoding. A large number of sophisticated face landmark detectors are avoided by using dense features. This approach also presents a large margin dimensionality reduction framework. It is well for high-dimensional Fisher vector representations. An effective and efficient face descriptor computation Pipeline is obtained which can be easily applied to large-scale face image repositories. This system is based on a single feature type. The main problem with this system is that it is not able to handle multi-feature image representation.

III. SYSTEM ARCHITECTURE

Fig. 2 describes the overall system structure. To perform better matching the background information is exploited from an extra cohort set. Cohort set contains synchronized images of different modalities. This system is divided into face pair matching pipeline, cohort identity list comparison (CILC), cohort sample list comparison (CSLC) and classification.

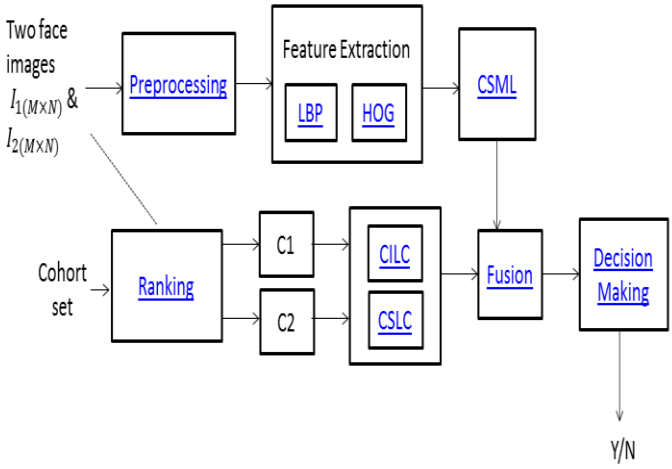

Fig. 2. System Architecture

Processing of proposed work is takes places as following way:

1. For each particular modality, calculate similarity score by matching two input face images.
2. Calculate two sorted cohort list by ranking all cohort images according to their similarity with respect to two input face images.
3. Cohort coefficients are extracted from two sorted cohort list.
4. Fusion of similarity score and extracted cohort coefficients to form final contribution.
5. In same way obtain contributions of other modalities.
6. Final decision making is made by fusing all contributions.

International Academy of Engineering and Medical Research, **2017**
Volume-2, ISSUE-4
Published Online April 2017 in IAEMR (http://www.iaemr.com)

IV.CONCLUSION

In this review paper, several existing techniques have studied and analysed in section II. Traditional methods of face recognition using cohort information and different modalities work effectively and efficiently for face identification. This method handles complicated facial variations by fusing different modalities. For the lack of representative information due to the few available face images, this method exploits a cohort set for additional information to better drive the matching. Therefore performance is achieved by fusing different modalities.

V. REFERENCE

[1] G. B. Huang, M. Mattar, T. Berg, and E. L. Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” Univ. Massachusetts, Amherst, MA, USA, Tech. Rep. 07-49, 2008, vol. 1, no. 2.

[2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” ACM Comput. Surv., vol. 35, no. 4, pp. 399–458, 2003.

[3] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, “A survey of affect recognition methods: Audio, visual, and spontaneous expressions,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 1, pp. 39–58, Jan. 2009.

[4] H. Han, C. Otto, X. Liu, and A. K. Jain, “Demographic estimation from face images: Human vs. machine performance,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 6, pp. 1148–1161, Jun. 2015.

[5] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, “Face recognition from a single image per person: A survey,” Pattern Recognit., vol. 39, no. 9, pp. 1725–1745, 2006.

[6] F. Schroff, T. Treibitz, D. Kriegman, and S. Belongie, “Pose, illumination and expression invariant pairwise face-similarity measure via Doppelgänger list comparison,” in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011, pp. 2494–2501.

[7] Q. Yin, X. Tang, and J. Sun, “An associate-predict model for face recognition,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2011, pp. 497–504.

[8] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 2, pp. 210–227, Feb. 2009.

[9] W. Deng, J. Hu, and J. Guo, “Extended SRC: Undersampled face recognition via intraclass variant dictionary,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 9, pp. 1864–1870, Sep. 2012.

[10] M. Yang, L. Van Gool, and L. Zhang, “Sparse variation dictionary learning for face recognition with a single training sample per person,” in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2013, pp. 689–696.

[11] S. Liao, A. K. Jain, and S. Z. Li, “Partial face recognition: Alignment free approach,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 5, pp. 1193–1205, May 2013.

[12] M. Tistarelli, Y. Sun, and N. Poh, “On the use of discriminative cohort score normalization for unconstrained face recognition,” IEEE Trans. Inf. Forensics Security, vol. 9, no. 12, pp. 2063–2075, Dec. 2014.

[13] L. Wolf, T. Hassner, and Y. Taigman, “Effective unconstrained face recognition by combining multiple descriptors and learned background statistics,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 10, pp. 1978–1990, Oct. 2011.

[14] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang, “Probabilistic elastic matching for pose variant face verification,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2013, pp. 3499–3506.

[15] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Fisher vector faces in the wild,” in Proc. Brit. Mach. Vis. Conf., 2013, pp. 1–12.

Authors’ Profiles

Ms. Divya K. Sawant received the B. E. degree in Computer Engineering from Savitribai Phule Pune University, Maharashtra, India. Currently pursuing M.E. degree in Computer Engineering from Savitribai Phule Pune University. Her current research interest includes digital image processing and information retrieval.

Prof. Nitin M. Shahane Associate Professor in Department of Computer Engineering, K. K. Wagh Institute of Engineering Education and Research, Nashik. His current research interests include digital signal processing, digital image processing, pattern recognition, machine learning, data mining and mathematical modelling.