3D-Aided Pose Invariant 2D Face Recognition

Xiang Xu
Computational Biomedicine Lab (CBL)
Department of Computer Science, University of Houston, Houston, TX, USA
xxu18@central.uh.edu

***Abstract*— This paper provides a brief overview regarding of the work of the Computational Biomedicine Lab on 3D-aided pose invariant 2D face recognition and my contributions to that system. The goal of my Ph.D. research is to improve 3D-2D face recognition over the state-of-the-art by at least 5%. The current state-of-the-art face recognition systems cannot perform well on the extreme poses. We are using Annotated Face Model (AFM) to address this problem. In the enrollment stage, the system frontalizes the face and extracts the discriminative features by applying the 3D model to estimate a 3D-2D transformation matrix. My thesis will focus on improving the fitting of the AFM to images with expressions along with developing an improved signature for representation.**

I. Introduction

The face recognition system has two steps: enrollment and matching. In the enrollment stage, features are obtained from gallery images to obtain a signature for each image. These signatures are compared to obtain a similarity or distance score during the matching stage. A general face recognition system is depicted in Fig. 1. Recently, face recognition technology has significantly advanced due to the use of deep neural networks. Several systems have achieved human performance or even better. For example, the VGG face descriptor was proposed by Parkhi et al. [4] to extract the feature vectors from the image directly using VGG-Very-Deep-16 CNN architecture. Schroff et al. proposed triplet loss [5] to train the deep neural networks with 200 million labeled faces and obtained performance of 99.63% on Labeled Faces in the Wild (LFW) standard benchmark [1].

However, face recognition is still not a resolved problem in real-world conditions. In unconstrained situations, there is a plethora of images with large variations in head pose, expression, illumination, and occlusions. In addition, training a deep neural network requires abundant labeled image data, which limits the contributions from academic teams that might not have access to datasets with millions of images. To address the pose problem, CBL has proposed a 3D-aided face recognition system using AFM (Fig. 3) [2]. Using a 3D model, it is easy to frontalize the face and extract the features from the frontal face images. The enrollment framework of Kakadiaris et al. [3] is depicted in Fig 2. After face detection, the face alignment module localizes the landmarks on the facial image. With the landmarks and the AFM model, the face reconstruction module reconstructs a personalized 3D model from a 2D face image.

Landmark detection, a.k.a face alignment, plays an important role in the face reconstruction and pose estimation modules in the 3D-aided 2D face recognition system [3]. In prior work, CBL has proposed two landmark localization algorithms [7], [6]. The algorithms use 2D images as input and output the 2D locations of pre-defined landmarks. Both algorithms follow the framework of cascaded regressions. An ensemble of random ferns was used to learn local feature descriptors [6]. In recent work, we explored global and local features obtained from Convolutional Neural Networks (CNN) for learning to estimate head pose and localize landmarks jointly [7].

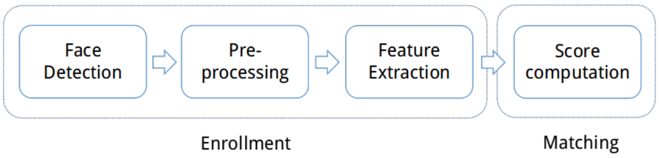

Fig. 1. Depiction of general face recognition system.

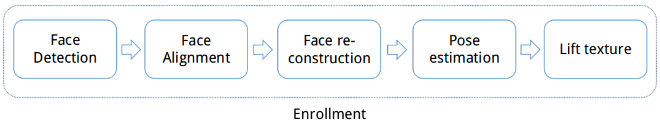

Fig. 2. Depiction of enrollment stage of 3D-aided face recognition system.

II. Previous work

To provide a better fitting result with a 3D model, we explored the high correlation between the 2D head pose and landmark locations. The head pose distributions from a reference database and learned local deep patch features are used to reduce error in the head pose estimation and face alignment tasks. First, we train *GNet* on the detected face region to obtain a rough estimate of the pose and to localize the seven primary landmarks. The most similar shape is selected for initialization from a reference shape pool constructed from the training samples according to the estimated head pose. Starting from the initial pose and shape, *LNet* is used to learn local CNN features and predict the shape and pose residuals. We demonstrated that our algorithm improves both the head pose estimation and face alignment.

III. Acknowledgment

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 2015-ST-061-BSH001. This grant is awarded to the

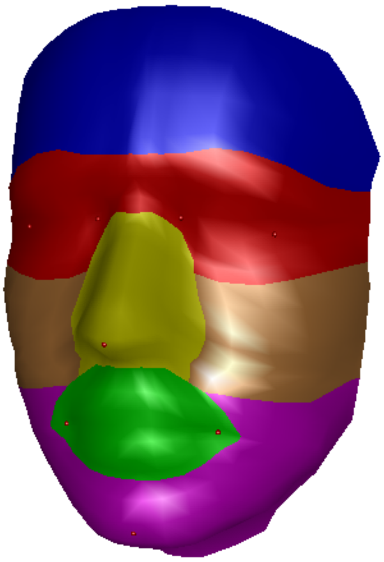

Fig. 3. Depiction of face annotated model.

Borders, Trade, and Immigration (BTI) Institute: A DHS Center of Excellence led by the University of Houston, and includes support for the project "Image and Video Person Identification in an Operational Environment" awarded to the University of Houston. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

References

[1] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller. Labeled faces in the Wild: A database for studying face recognition in unconstrained environments. In *Proc. European Conference on Computer Vision*, Marseille, France, Oct. 17-20 2008.

[2] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):640–649, 2007.

[3] I. A. Kakadiaris, G. Toderici, G. Evangelopoulos, G. Passalis, D. Chu, X. Zhao, S. K. Shah, and T. Theoharis. 3D-2D face recognition with pose-illumination normalization. *Computer Visiona and Image Understanding*, 154:137–151, 2017.

[4] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *Proc. British Machine Vision Conference*, pages 1–12, Swansea, UK, September 7-10 2015.

[5] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823, jun 2015.

[6] X. Xu and I. Kakadiaris. Joint head pose estimation and face alignment framework using global and local cnn features. In *Proc.* 12^{th} *IEEE Conference on Automatic Face and Gesture Recognition*, Washington, DC, May 30-June 3 2017.

[7] X. Xu, S. Shah, and I. A. Kakadiaris. Face alignment via an ensemble of random ferns. In *Proc. IEEE International Conference on Identity, Security and Behavior Analysis*, Sendai, Japan, Feb. 29 - Mar. 2 2016.