

# Multimodal Expression-Invariant Face Recognition Using Dual-Tree Complex Wavelet Transform

Fazael Ayatollahi<sup>1</sup>, Abolghasem A. Raie<sup>2</sup>

<sup>1,2</sup>Electrical Engineering Department  
Amirkabir University of Technology  
Tehran, Iran

<sup>1</sup>fazael\_ayat@yahoo.com

<sup>2</sup>raie@aut.ac.ir

Farshid Hajati<sup>3,4</sup>

<sup>3</sup>Electrical Engineering Department, Tafresh University,  
Tafresh, Iran

<sup>4</sup>School of Engineering, Griffith University, QLD 4111,  
Australia

hajati@tafreshu.ac.ir

**Abstract**—A new multimodal face recognition method which extracts features of rigid and semi-rigid regions of the face using Dual-Tree Complex Wavelet Transform (DT-CWT) is proposed. DT-CWT decomposes range and intensity images into eight sub-images consisting of six band-pass sub-images to represent face details and two low-pass sub-images to represent face approximates. In this work, support vector machine (SVM) has been used as the classifier. The proposed method has been evaluated using the face BU-3DFE dataset containing a wide range of expression changes. Findings include the overall identification rate of 98.1% and the overall verification rate of 99.3% at 0.1% false acceptance rate.

**Keywords**—face recognition; multimodal; complex wavelet; expression invariant; SVM

## I. INTRODUCTION

In the last two decades, face recognition has been known as one of the attractive topics in the computer vision research. Earlier face recognition methods used 2D intensity images for recognition. 2D face recognition methods achieve acceptable recognition rate under controlled conditions, but their performance drops dramatically with changing conditions such as environmental illumination, pose, occlusion, and expression. However, 3D scans are less influenced by environmental illumination and pose variations compared to 2D images [1].

One of the important challenges in 3D face recognition is the expression variation [2]. Expression-invariant 3D face recognition methods are divided into three main categories: global-based, region-based, and isometric. The global-based methods try to extract invariant features from the whole face surface. Xu et al. [3] proposed an automatic method to extract Gabor feature from both intensity and range images. They aligned all images using Iterative Closest Point (ICP) algorithm and a neutral model, and convolved intensity and range images with Gabor filters in five scales and eight orientations. After combination, LDA and AdaBoost learning techniques were used to extract the most efficient features and classification. Wang et al. [4] developed an automatic 3D face recognition system that extract three kinds of features including Haar wavelet, Gabor and Local Binary Pattern (LBP) from a Signed Shape Difference Map (SSDM) generated from two aligned 3D scans. They used a boosting algorithm to determine the most discriminative features and to build a strong classifier. Tang et

al. [5] proposed a method which extracted Local Binary Pattern (LBP) from depth and surface normal information after spherical feature-based division of 3D scans. Three different matching strategies and the nearest neighbor algorithm were applied for classification.

In the region-based methods, regions being less sensitive to expression are extracted and compared. Mian et al. [6] proposed an automatic multimodal face recognition method by combining a 3D Spherical Face Representation (SFR) and Scale Invariant feature Transform (SIFT) as a rejection classifier. Then, regions of 3D scans consisting of nose forehead and eyes are compared and classified by Iterative Closest Point (ICP) algorithm. Lei et al. [7] presented a region-based 3D face recognition method that determined rigid and semi-rigid region of 3D scans of the face and extracted four kinds of low geometry features to classify by Support Vector Machine (SVM).

Isometric methods model the expression variations by using isometric transformation [2]. These methods use geodesic distance being less sensitive to isometric variation. Berretti et al. [8] proposed a 3D face recognition method in which face scans were divided by a set of isometric strips and represented using 3D Weighted Walkthroughs (3DWW) as the descriptor. They used a graph-based method to match and compare query and gallery face scans. Li et al. [9] presented an expression-invariant 3D face recognition method using the geodesic level curves determined from the nose tip in the range images. They resampled intensity images using geodesic level curves to extract the equal-sized feature vectors. Hajati et al. [10] combined geodesic distances in the global and the local regions to transform face texture maps to new ones. Pseudo-Zernike Moments (PZM) were used to measure similarities between transformed images.

We present a novel multimodal face recognition method which is invariant under expression changes using both 2D and 3D face scans. As can be seen in Fig.1, we preprocess the image and find rigid and semi-rigid regions of 2D and 3D face scans because these regions are less influenced by expression. Then robust feature is extracted from these regions by DT-CWT. DT-CWT is a shift-invariant and good directional selectivity transform. The scans are decomposed at one scale by this wavelet transform and make six band-pass sub-images

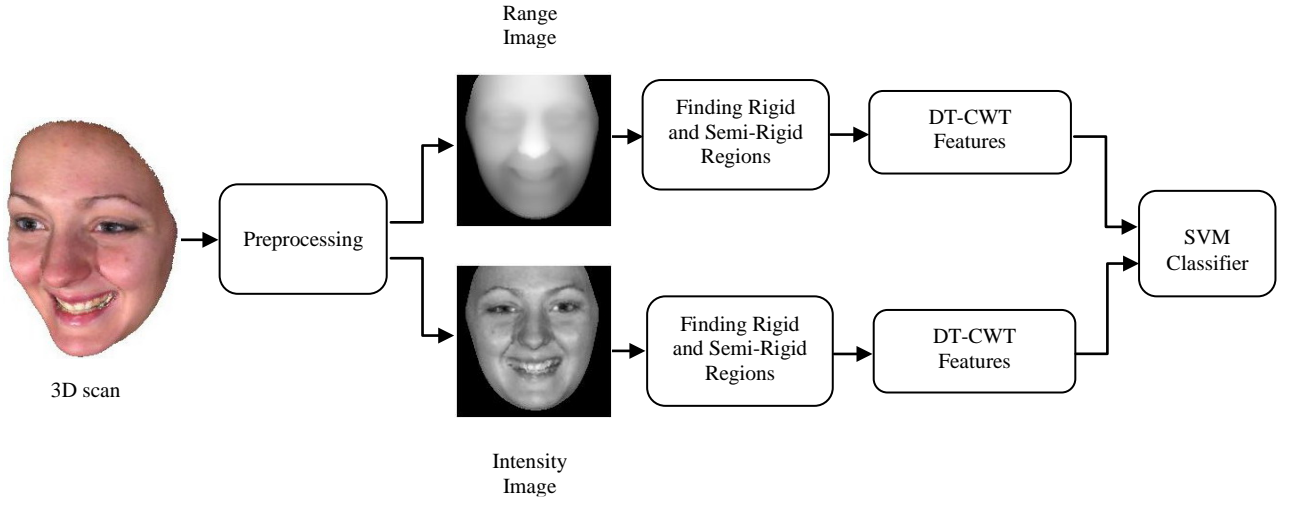


Fig. 1. Overview of the proposed method

to determine details and two low-pass sub-images to determine approximates. Another property of this transform is that it can recognize edges in six angles;  $-15^\circ$ ,  $-45^\circ$ ,  $-75^\circ$ ,  $+15^\circ$ ,  $+45^\circ$ ,  $+75^\circ$ . By combining 2D and 3D scans feature, final feature vectors are generated and classified by Support Vector Machine (SVM).

This paper is organized as follows. Section II reviews Dual-Tree Complex Wavelet Transform (DT-CWT). Section III introduces the proposed region-based multimodal face recognition method by combining DT-CWT and SVM. In Section IV, experimental results are discussed, and Section V concludes the paper.

## II. DUAL-TREE COMPLEX WAVELET TRANSFORM

DT-CWT was introduced to overcome the main drawbacks of Discrete Wavelet Transform (DWT). Fig. 2 shows one-dimensional DT-CWT filter bank which consists of two separate branches; one for real and another for imaginary part of this complex transform. Since each branch is a separate DWT filter bank, designing and implementation of DT-CWT is very easy with no extra mathematical cost. The outputs of real and imaginary parts of the filter banks are approximately Hilbert transform of each other.

Two-dimensional DT-CWT is an extension of one-dimensional DT-CWT. As 2D-DWT is generated by the product of the low-pass function ( $\phi$ ) and high-pass function ( $\psi$ ) and generates the three wavelet functions LH, HL, and HH the Two-dimensional DT-CWT is generated in the same manner. However, Two-dimensional DT-CWT is a complex transform; consequently, low-pass function ( $\phi$ ) and high-pass function ( $\psi$ ) are complex. These functions are computed using Eqs. (1) and (2).

$$\phi(x) = \phi_h(x) + j\phi_g(x) \quad (1)$$

$$\psi(x) = \psi_h(x) + j\psi_g(x) \quad (2)$$

The above equations generate six below wavelet functions [11-12].

$$\psi_{1,1}(x, y) = \phi_h(x) \psi_h(y) \quad (3)$$

$$\psi_{1,2}(x, y) = \psi_h(x) \phi_h(y) \quad (4)$$

$$\psi_{1,3}(x, y) = \psi_h(x) \psi_h(y) \quad (5)$$

$$\psi_{2,1}(x, y) = \phi_g(x) \psi_g(y) \quad (6)$$

$$\psi_{2,2}(x, y) = \psi_g(x) \phi_g(y) \quad (7)$$

$$\psi_{2,3}(x, y) = \psi_g(x) \psi_g(y) \quad (8)$$

DT-CWT is approximately analytic and without symmetry. As the result, it is an oriented wavelet transform and is able to detect edges and features in 6 angles which consist of  $-15^\circ$ ,  $-45^\circ$ ,  $-75^\circ$ ,  $+15^\circ$ ,  $+45^\circ$ ,  $+75^\circ$  as shown in Fig. 3.

Some properties of DT-CWT are approximately shift-invariance, good directional selectivity, low data redundancy, and optimize linear computation that can be used to approximate geometric structure of the face.

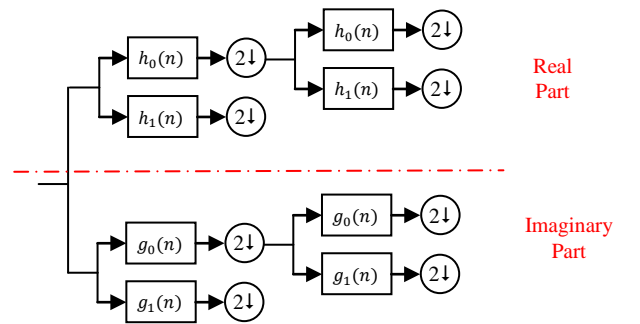


Fig. 2. One-dimensional DT-CWT filter bank

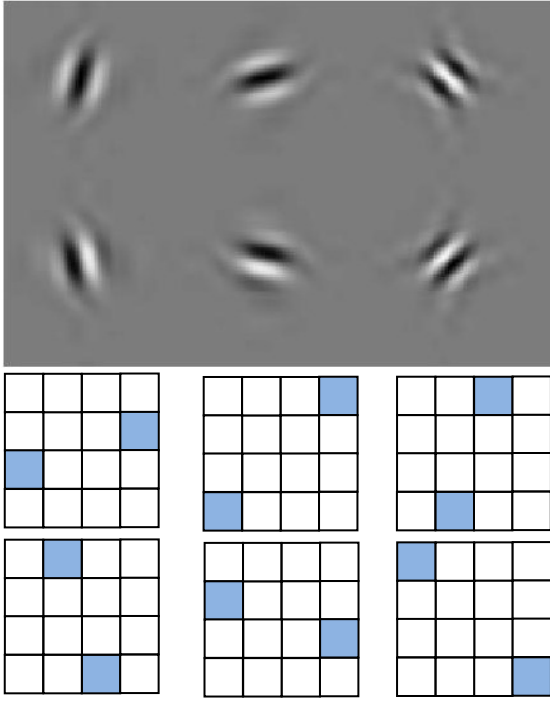


Fig. 3. 2D DT-CWT wavelets in space domain and their (idealized) support of spectrum in 2D frequency plane in 6 different angles

### III. THE PROPOSED METHOD

#### A. Preprocessing

In this work, we use BU-3DFE dataset [13] which contains 3D scans as point clouds. In point clouds, the points are sampled non-uniformly and it is necessary to resample them uniformly. First, the nose tip in 3D scans should be found. The nose tip has the maximum value in the  $z$  axis in the frontal face. Then, the points that have integer distance from the nose tip are interpolated in vertical and horizontal axes and related depth value is determined. Finally, images are cropped from the nose tip in a determined radius. These points form a 2D image drawn so that  $x$  and  $y$  are the values of the horizontal and the vertical axes of the image and the intensity shows the corresponding depth values. Such an image is called a range image and the nose tip is the center point in this representation.

The above steps are repeated for the intensity images in dataset with the difference that intensity information is used instead of depth information. As a result, there are two images. One of them consists of the intensity information and the other has the depth with corresponding points.

Throughout the face, the mouth region shown the most variations compared to other regions; the nose region has the least and the eye-forehead have the less variations by the expression because of their rigidity. Therefore, we use a binary mask which is shown in Fig. 4 to crop rigid region (nose region) and semi-rigid region (eyes-forehead). To generate this mask, some landmarks such as mouth and eyes corners are used [7]. This mask is applied to both intensity and range

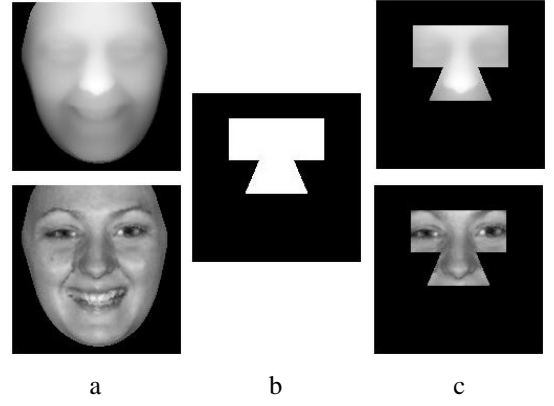


Fig. 4. (a) range and intensity images; (b) binary mask; (c) masking result

images and then a robust feature that will be described in the next section is extracted.

#### B. Feature extraction

To perform DT-CWT to masked intensity and range images in arbitrary scale, their sizes should be increased to the multiple of  $2^J$  where  $J$  is an arbitrary scale. When images are decomposed by DT-CWT, it is necessary to increase the size by adding the constant value pixel to the borders.

The images should be decomposed to the  $J$  level. Magnitudes of Coefficients are used as feature because this feature is approximately shift-invariant. As an instance, the semi-rigid region feature changes a little under different expressions (like disgust and surprise). Therefore, an approximately shift-invariant transform generates the same features. Now, the  $J$ th level sub-bands are preserved and the others are discarded; thus, each image contains six band-pass sub-images which show the details, and two low-pass sub-images which show approximates. Therefore, we have sixteen sub-images for each scans; eight sub-images for the intensity image and eight sub-images for the range image. These sub-images are shown in Fig. 5. The advantage of the discarding of other levels is the dimension reduction of the final vector. Consequently, techniques like PCA and LDA are not necessary.

Sixteen sub-images of each scan are reshaped as vectors aligned respectively (the range vectors and then the intensity vectors) to make final feature vector. This large vector is normalized to zero mean and unit variance.

To classify vectors Support Vector Machine (SVM) is used. Since SVM is a binary classifier, One Versus the Rest (OVR) technique is used to classify multi classes.

### IV. EXPERIMENTAL RESULTS

#### A. Dataset

The BU-3DFE dataset has been used in this study. It is one of the important datasets for face recognition under expression variations [2]. This dataset includes 3D scans with relating texture of 56 women and 44 men (totally 100 people) in various expressions. Each person has totally 25 scans

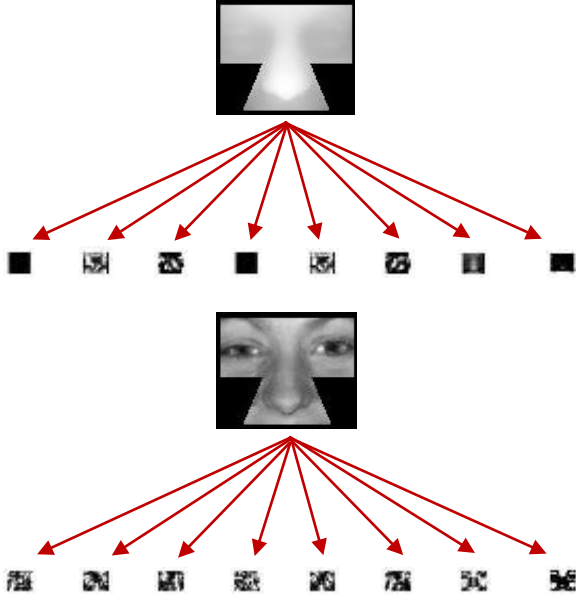


Fig. 5. Range and intensity masked images and their eight sub-images, from left hand six first sub-iamges show details and the last two show the approximation

containing one neutral and six kinds of expressions. Fig. 6 presents these expressions which include anger, happiness, sadness, surprise, disgust, and fear in four different intensities.

To test the algorithm on BU-3DFE dataset we divided scans of each person into two sets of “low intensity” and “high intensity”. Low intensity set contains 1200 scans with low intensity expressions and high intensity set contains other 1200 scans with high intensity expressions. Note that neutral scans are not considered. We choose “low intensity versus high intensity” and “high intensity versus low intensity” strategies for classification. In “low intensity versus high intensity” strategy the train set is the low intensity set and the test set is the high intensity set and in “high intensity versus low intensity” vice versa [7].

### B. Comparison of the different features

To test the efficiency of the extracted features, we compare this feature and two other similar ones: Gabor and DWT. To apply these three kinds of features in the same manners a



Fig. 6. An example of 25 images of a person in BU-3DFE dataset [10]

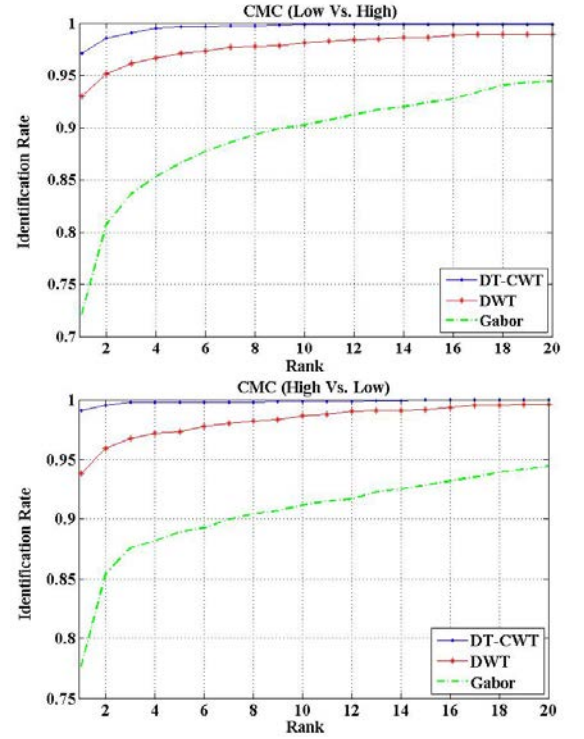


Fig. 7. Identification rate to compare DT-CWT, DWT and Gabor features

certain scale is chosen and other scale coefficients are discarded. Note that Gabor features are applied in eight different directions consequently, for approximately time unification PCA algorithm is used to reduce the feature vector dimension to DT-CWT feature vector size.

The identification rates are shown in Table I. As expected, DT-CWT feature has higher performance than the two others because of their mentioned properties.

### C. Recognition results

Cumulative Match Curve (CMC) of our proposed approach and the two other methods mentioned in section B are presented in Fig. 7. The rank-1 identification rate for low intensity vs. high intensity and high intensity vs. low intensity are 97.1% and 99.1%, respectively. In addition, the verification rate at 0.1% False Acceptance Rate (FAR) for low intensity vs. high intensity is achieved 99% and for high intensity vs. low intensity is 99.7%.

The rank-1 identification rate for each expression of anger, happiness, sadness, surprise, disgust, and fear is provided in Table II. This parameter, in every expression with both strategies, is higher than 95%. In addition, the open mouth problem in some expressions like surprise that decreases

TABLE I. THE RANK-1 IDENTIFICATION RATE COMPARISON OF DIFFERENT FEATURES

features	R-1 Identification rate (%)	
	Low int. vs. high int.	High int. vs. low int.
DT-CWT	97.1	99.1
DWT	93	93.8
Gabor	72.2	77.7

TABLE II. OUR APPROACH PERFORMANCE UNDER DIFFERENT EXPRESSION

Expression	R-1 Identification rate (%)	
	<i>Low int. vs. high int.</i>	<i>High int. vs. low int.</i>
Angry	96	98.5
Disgust	95	99
Fear	97.5	99.5
Happy	99	99
Sad	98.5	99.5
Surprise	96.5	99

TABLE III. COMPARISON OF RANK-1 IDENTIFICATION RESULT

Method	No. of probes	R-1 Identification rate (%)
<b>Our Approach</b>	1200	<b>98.1</b>
Smeets et al. [14]	900	89.9
Mpiperis et al.[15]	1250	86
Kaushik et al. [16]	695	98.9
Yinjie Lei et al. [7]	1200	98.2

recognition rate in many other methods has been solved. Comparison results between our approach and some other new methods which use the same dataset are presented in Table III. Though we have used all six expressions and more probe images, we achieved a better performance than Smeets et al. [14]. In addition, with approximately the same number of probe images, we achieved a higher rate compared to the Mpiperis et al. [15]. Kaushik et al. [16] reported above 98% recognition rate which seems is better than our result but they have used a limited number of probes. Furthermore, our result is similar to Lie et al. [7].

## V. CONCLUSION

In this paper, we proposed a novel region-based multimodal face recognition method which was invariant to the expression variations. In this method, we used both 2D and 3D information obtained from BU-3DFE dataset. We combined a binary mask to remove non-rigid region of the face and Dual-Tree Complex Wavelet Transform to extract the feature. DT-CWT is a shift invariant and good directional selectivity transform applied in a way that generates an appropriate low dimension feature vector which can be directly classified by SVM without any dimension reduction techniques. For these properties, DT-CWT extracts superior feature compared to Gabor and DWT. Also, our approach achieved higher performance compared to the other methods by 98.1% and 99.3% of the rank-1 overall identification rate and overall verification rate at 0.1% FAR, respectively.

## REFERENCES

- [1] A. Abate, M. Nappi, D. Riccio, G. Sabatino, "2D and 3D face recognition: A survey," Pattern Recognition Letters, vol. 28, no. 14, pp. 1885-1906, 2007.
- [2] D. Smeets, P.Claes, J. Hermans, D. Vandermeulen, P. Suetens, , "A Comparative Study of 3-D Face Recognition Under Expression Variations," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on , vol.42, no.5, pp.710-727, Sept. 2012.
- [3] C. Xu, S. Li, T. Tan, L. Quan, "Automatic 3D face recognition from depth and intensity Gabor features," Pattern Recognition, vol. 42, no. 9, pp. 1895-1905, Sept. 2009.
- [4] Y. Wang, J. Liu, X. Tang, "Robust 3D Face Recognition by Local Shape Difference Boosting," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.32, no.10, pp.1858-1870, Oct. 2010.
- [5] H. Tang, B. Yin, Yanfeng Sun, Y. Hu, "3D face recognition using local binary patterns," Signal Processing, vol.93, no. 8, Aug. 2013.
- [6] A.S. Mian, M. Bennamoun, R. Owens, "An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.29, no.11, pp.1927-1943, Nov. 2007.
- [7] Y. Lei, M. Bennamoun, A.A. El-Sallam, "An efficient 3D face recognition approach based on the fusion of novel local low-level features," Pattern Recognition, vol. 46, no. 1, Jan. 2013.
- [8] S. Berretti, A. Del Bimbo, P. Pala, "3D Face Recognition Using Isogeodesic Stripes," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.32, no.12, pp.2162-2177, Dec. 2010
- [9] L. Li, C. Xu, W. Tang, C. Zhong, "3D face recognition by constructing deformation invariant image," Pattern Recognition Letters, vol. 29, no. 10, July 2008.
- [10] F. Hajati, A.A. Raie, Y. Gao, "2.5D face recognition using Patch Geodesic Moments," Pattern Recognition, vol. 45, no. 3, pp. 969-982, March 2012.
- [11] I.W. Selesnick, R.G. Baraniuk, N.G. Kingsbury, "The Dual-Tree Complex Wavelet Transform," IEEE SIGNAL PROCESSING MAGAZINE, Nov. 2005.
- [12] Chao-Chun Liu, Dao-Qing Dai, "Face Recognition Using Dual-Tree Complex Wavelet Features," Image Processing, IEEE Transactions on, vol. 18, no. 11, pp. 2593-2599, Nov. 2009
- [13] L. Yin, X. Wei, Y. Sun, J. Wang, M.J. Rosato, "A 3D facial expression database for facial behavior research," in: Proceedings of the Seventh International Conference on Automatic Face and Gesture Recognition (FGR),pp. 211-216, 2006.
- [14] D. Smeets, T. Fabry, J. Hermans, D. Vandermeulen, P. Suetens, "Fusion of an Isometric Deformation Modeling Approach Using Spectral Decomposition and a Region-Based Approach Using ICP for Expression-Invariant 3D Face Recognition," Pattern Recognition (ICPR), 2010 20th International Conference on, pp. 1172-1175, Aug. 2010.
- [15] I. Mpiperis, S. Malassiotis, M.G. Strintzis, "Expression-Compensated 3D Face Recognition with Geodesically Aligned Bilinear Models," Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on, pp. 1-6, 2008.
- [16] V.D. Kaushik, A. Budhwar, A. Dubey, R. Agrawal,; S. Gupta, V.K. Pathak, P. Gupta, "An Efficient 3D Face Recognition Algorithm," New Technologies, Mobility and Security (NTMS), 2009 3rd International Conference on, pp. 1,5, 20-23 Dec. 2009.