



International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

<http://gssrr.org/index.php?journal=JournalOfBasicAndApplied>



Facial Landmark Detection and Estimation under Various Expressions and Occlusions

Abdulganiyu Abdu Yusuf ^{a*}, Zahraddeen Sufyanu ^b, Fatma Susilawati
Mohamad ^c, Kutiba Nanaa ^d

^{a,b,c,d} Faculty of Informatics and Computing, 21300 Gong Badak campus, Universiti Sultan Zainal Abidin
(UniSZA), Terengganu, Malaysia.

^aEmail: abdulg720@gmail.com

^bEmail: sufyanzzz@yahoo.com

^cEmail: fatma@unisza.edu.my

Abstract

Landmark localization is one of the fundamental approaches to facial expressions recognition, occlusions detection and face alignments. It plays a vital role in many applications in image processing and computer vision. The acquisition conditions such as expression, occlusion and background complexity affect the landmark localization performance, which subsequently lead to wrong classification. In this paper, the writers bestowed the challenges of various landmark detection techniques, number of landmark points and dataset types been employed from the existing literatures. Meanwhile, advanced technique for facial landmark detection under various expressions and occlusions was presented. This was carried out using Point Distribution Model (PDM) to estimate the occluded part of the facial regions and detect the face. The proposed method was evaluated using University Milano Bicocca Database (UMB). This approach gave more promising result when compared to several previous works. However, the technique detected images despite varieties of occlusions and expressions. It can further be applied on images with different poses and illumination variations.

Keywords: PDM; Facial landmark; Occlusion; Expression; UMB

* Corresponding Author:

E-mail address: SI0441@putra.unisza.edu.my, abdulg720@gmail.com.

1. Introduction

Face recognition is a process of recognizing an individual from their facial attributes. It belongs to the class of biometric system [1]. Occlusions degrade the performance of face recognition evidently [2,3]. Therefore, when occluding object is known, a specific strategy can be developed to estimate and compensate the occluded regions, for reliable face recognition [3]. Facial expressions are sometimes used for human communication, as it provides natural and immediate indication about a person's intentions and emotions. Face landmarking is defined as detection of certain characteristic points on the face. The points used to represent the vital information needed to classify an individual; this is achieved by building a model [4]. The knowledge of landmarking can be applied in login authentication and security units such as police department, criminal investigation, and immigration department among others.

Regardless of good landmark detector, there are cases where landmark values cannot be computed, due to missing data (the occlusion). At the same time, landmark approach should try to detect as many points as possible, this leads to complex and more general systems.

Human faces vary from one another; it is therefore difficult to differentiate faces under high occlusion. More so, to detect and remove occlusions on the face images quickly and automatically becomes a largely unsolved problem. This makes face detection and recognition among the toughest problems in the fields of computer vision and biometrics. Moreover, aligning faces robustly and precisely is one of the most important steps to solve the challenges in facial landmark detection [5]. In order to overcome these challenges, significant contributions have been made to aid in the process of identification, in different scenarios [3,6]. However, there is still need for significant attentions in terms of missing data, control points (labeling), misalignment, restoration, expression and illumination variations.

Information about shape variations are usually collected to build a model. The model represents a predefined number of landmark points, which depends on the complexity of the object's shape and desired level of detailed descriptions of where it is needed. In this study, we present a technique for facial landmark detection and estimation under various expressions and occlusions. The approach aimed at detecting the position of face and facial features (eye, mouth, and nose) despite occluding object by hair, scarf, hand, and so on. It will also be tested to fit on faces with different expressions such as smile, angry, and open mouth. We obtained faster detection and fitting time compared to other reported techniques in the literature.

The rest of the paper is organized as follows: Section 1 introduces the concepts of facial landmarks with regard to expressions and occlusions. Section 2 illustrates related works of current applications. In section 3, the materials and methods used in the study are developed. The results and discussion of the proposed system are stated in Section 4. Finally, conclusion and future work are drawn in Section 5.

2. Related Work

Facial landmark localization is a prerequisite for face recognition either in two dimensions (2D) or three dimensions (3D). Various algorithms and constraints have been proposed to detect and handle the face

occlusion.

2.1 Reported challenges on 2D and 3D methods

A method for removing glasses from a frontal face image was proposed by the authors in [7]. The occluded region by glasses was first detected and then generated a natural looking facial image without glasses using Principal Component Analysis (PCA) reconstruction. Besides that, a more general solution is needed for more challenging situations especially where the occlusions are unpredicted. A possible solution for this problem is local approaches, for instance, the proposed work in [8,9] divided the face into local regions which were analysed independently. Moreover, the authors in [10] proposed a part-based local representation method using Locally Salient Independent Component Analysis (LSA-ICA).

Contrary to this approach, another method was investigated in [11] which eliminated parts that correspond to occlusions, and that may hinder the recognition accuracy. The authors in [12,3] presented a strategy that approaches the occlusion problem by performing a restoration of the faces: the occluded regions were detected, and the non-occluded regions were used to recover the missing information. However, accumulation of error at each stage deteriorated the final performance of the system. Similarly, the research in [13] divided the faces into rectangular regions, and on the basis of their gray level histograms, the probability of the occlusion was estimated.

In more advanced study, the authors in [14] proposed a part based recognition method, which used Average Regional Models (ARM) and matched the various parts of the face independently. Various fusion techniques were used to integrate the similarity between the face parts and the corresponding ARMs. However, significant improvement was reported with respect to Iterative Closest Point (ICP) matching in the case of occlusions (by hands, hair or eyeglasses) and facial expressions. In another development, the authors in [15] presented an approach that provided general labeling over a wide region of face, which was robust to occlusion and pose variations.

The Landmarking scheme used in [15,16] involved 14 and 22 points respectively, which were manually selected across the faces. However, the systems were costly when the number of points increases. The studies of [16] and [17] on landmark detection and pose estimation worked robustly, even if half of the face is missing. The selected landmark points for both studies were 5 - 8 and were not fully automated. There is need for increased landmark points for more robustness. In another efforts by [18,19], a fully automated and multimodal (2D and 3D) algorithm for facial image synthesis was proposed. The labeled landmark points were 57 and 79 points. But there was no consideration of other timing issues such as rotating and aligning 3D models, which instantly compared to the modeling time. And the recognition rate was affected in the case of some missing landmarks. Additionally, the accuracy of landmark localization increased proportionally to number of landmarks considered [20]. When the number landmark points increases from 3 - 68, 50% improvement was observed.

In comparison between 2D and 3D methods, one advantage of landmarking in 3D is that, it enables alternate processing techniques for landmarks since there are multiple ways of representing 3D face data. For example,

point clouds, multiple profiles, curvature, shape index and depth maps [21] have been used for face recognition, but not fully exploited for landmarking. These advantages cannot be found in 2D landmarking which is largely affected by severe lighting and pose variation.

Conversely, the drawback of 3D face raw data is that, it demands more pre-processing steps compared to 2D. For instance, the face surface must be smoothed, spikes and discontinuities must be removed, and gaps must be filled in. Moreover, the execution time is dependent on the scene complexity. Therefore, 3D landmarking are considered costly and time consuming. This is because, significant effort is required to produce an acceptable result [22]. Again, the time required to generate new views in 2D using interpolation is independent of the scene complexity.

2.2 Other factors and challenges

Researchers reported different techniques via various databases being used on landmark points, conducted using 2D and 3D methods and achieved considerable results often. Face landmarking algorithm that works well under and across all intrinsic variations of faces, and delivers the target points in a time efficient manner has not yet been feasible [22]. Table 1 depicts comparison of some recent landmarking techniques. Other factors that hinder the performance of facial landmark detection are:

- **The accurate number of landmarks:** The accuracy and number of landmark points vary based on the intended application. For instance, higher level tasks, such as facial expression understanding or facial animation, requires greater number of landmarks within the range of 20–30 and 60–80, as well as higher spatial accuracy [23,24]. Few number of landmark points were selected in [25,26] as shown in table 1. The results obtained were not applicable under significant pose variations.
- **The acquisition conditions:** Acquisition conditions such as illumination, resolution and background complexity affect the landmark localization performance. This was attested by the fact that landmark localizers trained in one database have usually inferior performance when tested on another database [27], [28].
- **Face variability:** Landmark appearances differ due to intrinsic factors such as face variability between individuals, and due to extrinsic factors such as partial occlusion, pose and camera resolution. Moreover, facial landmarks can only be partially observed due to occlusions of hair, hand movements or self-occlusion due to extensive head rotations. The other two major variations that compromise the success of landmark detection are facial expressions and illuminations [22].

Table 1: Comparison of some recent landmarking techniques

Mode	Techniques	Landmark points	Databases	Remark
2D	SIFT and ASM [29]	58	BioID and FRGC V2	• No localization results
	ASM, SIFT, and Probabilistic PCA [30]	45	IMM and CMU-PIE	• No localization results
	PDM of landmarks and local descriptors [31]	12	CMU Multi-PIE	• No localization results
	Conditional Regression Forest [32]	10	LFW	• 2D normalized errors per landmark. Close to human accuracy and real time. • Different expression and lighting variation are considered.
3D	Generic Algorithm [26]	3	FRGC V1	• Not applicable under significant pose variations
	PDM, Shape index and Curvedness index [25]	5	BU3DFE	• No localization results. • Not applicable under significant pose variations
	Shape index, Spin images and FLM [33]	8	FRGC V2 and UND Ear	• Applicable under significant yaw and expression variations.

3. Materials and Methods

The experiment was conducted using UMB database [34]. This is a collection of multimodal (3D + 2D images) facial acquisitions. The database is composed of 1473 images of 143 subjects. The subjects include 98 males and 45 females. Most of the subjects have been with eyeglasses, holding phones, hat, partially occluded by hair and other miscellaneous objects. The total number of occluded faces are 578. And the size of each image is 640x480 pixels.

A total number of 100 neutral faces were selected for training the model. Another 100 faces were used as test set. The set includes multiple occluding objects (by hand, hair and scarf) and different facial expressions such as angry, smile, and open mouth. The images were cropped to retain only the region of interest. For the purpose of experiment, Matlab software was used on personal computer with the following specifications; Intel Celeron of 1.0 GHz processor and 2.00 GB DDR3 RAM. Figure 1 shows some sample of the neutral images and test images from the UMB database.

The images were converted to gray-scale, to reduce the amount of information in the image and then cropped to size 220×240 to remove the background and retain only the region of interest. The images were normalized by scaling, rotation and translation so that the faces were centered into a definite coordinates to enable the model easily fit the appropriate regions. We used the principle of PDM to capture the statistical face variations. The

number of landmarks should be adequate to show details of the overall shape. For accuracy, 66 control points were used to represent the faces as a sequence of connected landmarks similar to [35]. Finally, model fitting was achieved through transformations between the model points and candidate vertices. Candidate vertices include candidate inner eye, nose tip and mouth tips vertices. To extract the candidate interest points, the face model was categorized into five regions; region 1 represents Right Eye and Right Eye Brows (REREB), region 2 represents Left Eye and Left Eye Brows (LELEB). The Nose Tip (NT) is represented by region 3 and region 4 represents Mouth Tips (MT) while region 5 represents Edges of the Face (FE). Figure 2 shows the block diagram of the developed system, and figure 3 depicts the modeled regions.



Fig 1: Sample of neutral images obtained from UMB database (first row) and corresponding occluded faces by different types of objects (second row).

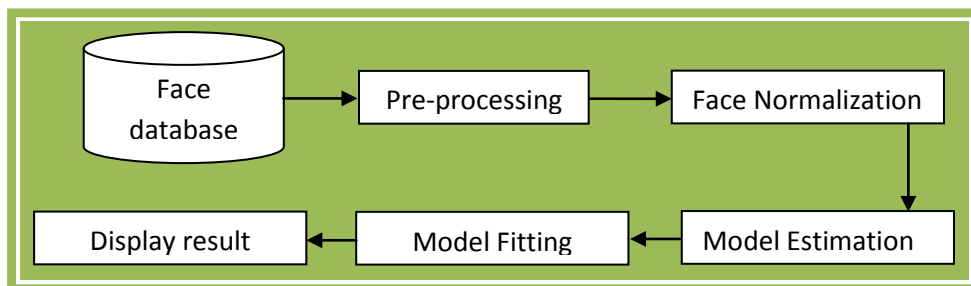


Fig 2: Block diagram of the proposed system

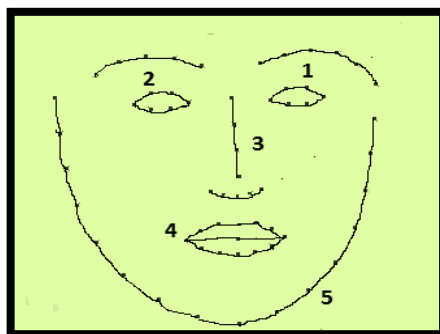


Fig 3: Model divided into five regions

4. Results and Discussion

The detected occluded faces were categorized into a set of weakly and strongly occluded faces based on the amount of details covered. If at least three out of the five regions (3/5) are covered then the face is classified as strong occlusion, otherwise as a weak occlusion.

Figure 4 shows some of the results of face detection using this model with different occlusion and expression. Images a, b, c are faces with expression and d, e, f, are faces with occlusions. In figure 4 'a', the occluded regions are; LELEB, MT and FE. Three types of expressions were taken into consideration in the experiment, these are: angry, smile and open mouth. The results performed well for the expression of the type angry and smile. In the same figure, 'a' and 'b', the three out of the five regions (of the facial features) were partially covered. Hence, they were classified as strong occlusion. On the other hand, image 'c' of the same figure was classified as weak occlusion since two out of the five regions (2/5) of the face were covered.

In table 2, a strategy was developed to estimate the location of the occluded regions based on the positions and number of landmark points represented on the face regions. Table 3 summarizes the results of the face and facial features detection implemented in this paper. The average detection and fitting time per image (for occluded faces) were 5.43 sec and 1.18 sec respectively. And for the expressions, the average detection and fitting time per image (for expressions) were 4.56 sec and 1.08 sec respectively.

Finally, most of the strong occluded faces covered the region FE. Whereas weak occluded faces mostly covered regions REREB and LELEB. We obtained faster detection and fitting time in faces with expressions than in faces with occlusions. The major drawback of this work is that, the fitting accuracy decreases in the case of strongly occluded faces. Although, the eyes in the neutral images were closed, yet, the model detected the eye regions. This study focused on expressions and occlusions using Point Distribution model, and more promising result was reported compared to many studies in the literature. Table 3 demonstrates the detection results of this technique and the previous studies.

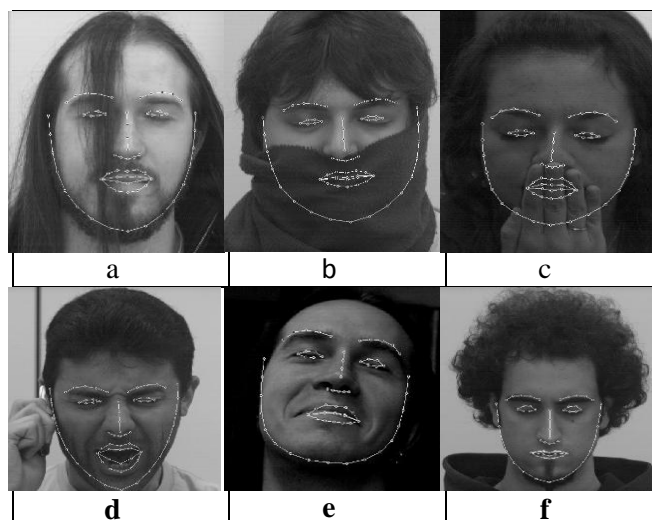


Fig 4: Model fitted on different face with occlusion (a-c) and different expression (d-f)

Table 2: Regions and their corresponding positions of the landmark point

Region name	Position of the landmark points	No. of landmark points
REREB	1 – 11	11
LELEB	12 – 22	11
NT	23 – 31	9
MT	32 – 48	16
FE	59 – 66	17

Table 3: Detailed results obtained from the experiment

Acquisition type	Fitting time (sec)	Detection time (sec)	
Occlusion type {	Hand	1.18	5.48
	Hair	1.11	4.92
	Scarf	1.24	5.88
Expression type {	Angry	1.04	4.40
	Smile	1.08	4.48
	Open Mouth	1.13	4.81

Table 4: Comparison of the proposed and previous methods

S/N	Techniques	Detection time (sec)
1	Gabor Wavelets [36]	16.00
2	ICP [37]	7.52
3	PDM & ICP [38]	3.00
4	PCA [39]	6.18
5	Our approach (with expression)	4.56
6	Our approach (with occlusion)	5.43

5. Conclusion and Future Work

This paper discussed the various challenges in face landmarking, and proposed an approach that detected facial landmarks under various occlusions and expressions. The proposed technique was evaluated using UMB database, which is one of the most challenging facial databases with different types of occlusions and expressions. A PDM was employed with a total number of 66 points to fully describe the shape variations in accordance with five divided regions. The application of this model on very many occlusions and expressions is considered unique in this study, contrary to many existing works in the literature where single variation (like

expression, occlusion or pose) was reported. Despite the strong occlusions and difficult expressions, yet, we obtained detection within shortest possible time. The model worked robustly, even if more than half of the face is entirely covered. The model can further be applied to work on different poses and illuminations. Finally, we hope to improve the technique (i) by testing it on different databases (ii) by improving the accuracy of fitting the model in terms of pose changes and (iii) by increasing detection and fitting time.

Acknowledgements

We appreciate University Sultan Zainal Abidin (UniSZA), Kuala Terengganu, Terengganu, Malaysia for the academic support of this work. Our profound gratitude goes to Kano State Government, Nigeria for the financial backing.

References

- [1] D. N. Parmar and B. B. Mehta, "Face Recognition Methods & Applications," *International Journal of Computer Technology & Applications*, vol. 4, no. 1, pp. 84–86, 2013.
- [2] A. Colombo, C. Cusano, and R. Schittine, "Detection and Restoration of Occlusions for 3D Recognition," In *IEEE International Conference of Multimedia and Expo*, 2006, pp. 1541-1544.
- [3] A. Colombo, C. Cusano, and R. Schettini, "Three Dimensional Occlusion Detection and Restoration of Partially Occluded Faces," *Journal of Mathematical Imaging and Vision*, 40(1):105–119, 2011.
- [4] I. Dryden and K. Mardia, *Statistical Shape Analysis*. Wiley, 1998.
- [5] D. Ramanan, "Face Detection, Pose Estimation, and Landmark Localization in the Wild," *2012 IEEE Conf. Computer Vision & Pattern Recognition*, June 2012, pp. 2879–2886.
- [6] M. V. Gupta and D. Sharma, "A Study of Various Face Detection Methods," *International Journal of Computer and Communication Engineering*, vol. 3, no. 5, pp. 3–6, 2014.
- [7] J. S. Park, Y. H. Oh, S. C. Ahn, and S. W. Lee, "Glasses Removal from Facial Image using Recursive Error Compensation," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 27(5), 2005, pp. 805-811.
- [8] A. M. Martinez, "Recognition of Partially Occluded and/or imprecisely Localized Faces using a Probabilistic Approach," In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, 2000, pp. 712–717.
- [9] A. M. Martinez, "Recognizing Imprecisely Localized, Partially Occluded and Expression Variant Faces from a Single Sample per Class," *IEEE Trans. Pattern Anal. Mach. Intell.* 24(6), 2002, pp. 748–763.
- [10] J. Kim, J. Choi, J. Yi, and M. Turk, "Effective Representation Using ICA for Face Recognition Robust to Local Distortion and Partial Occlusion. In *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(12): 2005, 1977–1981.
- [11] F. Tarrés and A. Rama, "A Novel Method for Face Recognition under Partial Occlusion or Facial Expression Variations," In *Proc. 47th Int'l Symp. ELMAR*, 2005, pp. 163–166.
- [12] A. Colombo, C. Cusano, and R. Schettini, "Gappy PCA Classification for Occlusion Tolerant 3D Face Detection," *Journal of Mathematical Imaging and Vision*, 35(3):193–207, 2009.

- [13] W. Zhang, S. Shan, X. Chen, and W. Gao, "Local Gabor Binary Patterns Based on Kullbackleibler Divergence for Partially Occluded Face Recognition," *IEEE Signal Processing Letters*, 14(11):875–878, 2007.
- [14] N. Alyuz, B. Gökberk, and L. Akarun, "A 3D Face Recognition System for Expression and Occlusion Invariance," In *Proceedings of 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2008, pp. 1–7.
- [15] C. Creusot, N. Pears, and J. Austin, "3D Face Landmark Labelling," *Proc. ACM Work. 3D object Retr. - 3DOR '10*, 2010, p. 27.
- [16] P. Perakis, T. Theoharis, G. Passalis, and I. Kakadiaris, "Automatic 3D Facial Region Retrieval from Multi-pose Facial Datasets," In *Proc. Eurographics Workshop on 3D Object Retrieval*, Munich, Germany, Mar. 30 - Apr. 3 2009, pp. 37–44.
- [17] G. Passalis, P. Perakis, T. Theoharis, and I. Kakadiaris, "Using Facial Symmetry to handle Pose Variations in Real-world 3D Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, Oct. 2011, pp. 1938–1951.
- [18] A. Ansari, M. H. Mahoor, and M. Abdel-mottaleb, "Normalized 3D to 2D Model-Based Facial Image Synthesis for 2D Model-Based Face Recognition," *IEEE GCC Conference and Exhibition*, Dubai, United Arab Emirates, February 2011, pp. 178–181.
- [19] J. Heo and M. Savvides, "3-D Generic Elastic Models for Fast and Texture Preserving 2-D Novel Pose Synthesis," *IEEE Trans. Information Forensics Security*, vol. 7, no. 2, pp. 563–576, Apr. 2012.
- [20] S. Milborrow, and F. Nicolls, "Locating Facial Features with an Extended Active Shape Model," In *Computer Vision–ECCV (Marseille, France)*, pp. 504-513, 2008.
- [21] B. Gokberk, M. O. Irfanoglu, and L. Akarun, "3D Shape-based Face Representation and Feature Extraction for Face Recognition", *Image Vis. Comput.* 24(8), 857–869, 2006.
- [22] O. Çeliktutan, S. Ulukaya, and B. Sankur, "A Comparative Study of Face Landmarking Techniques," *EURASIP Journal on Image and Video Processing*, pp. 1–27, 2013.
- [23] K. E Ko, and K.B. Sim, "Active Appearance Model for Facial Emotion Recognition," In *Proc. of Int. Symposium on Industrial Electronics Development (Seoul, Korea)*, 2009, pp. 1019–1022.
- [24] A. Asthana, A. Khwaja, and R. Goecke, "Automatic Frontal Face Annotation and AAM Building for Arbitrary Expressions from a Single Frontal Image Only," In *16th IEEE International Conference International Conference on Image Processing (ICIP)*, (Cairo, Egypt), 2009, pp. 2445–2448.
- [25] P. Nair and A. Cavallaro, "3D Face Detection, Landmark Localization, and Registration using a Point Distribution Model," *IEEE Transactions on Multimedia*, vol. 11, no. 4, pp. 611–623, June 2009.
- [26] T. Yu and Y. Moon, "A Novel Genetic Algorithm for 3D Facial Landmark Localization," In *Proc. 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems*, Arlington, VA, Sep. 20 - Oct. 1 2008.
- [27] P. J Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," In *Proceeding of International Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 947–954.
- [28] H.C. Akakin, and B. Sankur, "Robust 2D/3D Face Landmarking" In *Proceeding of Conference on 3DTV (Kos, Greece)*, 2007, pp. 1–4.

- [29] D. Zhou, D. Petrovska-Delacretaz, and B. Dorizzi, "Automatic Landmark Location with a Combined Active Shape Model," In Proc. IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, Sep. 2009, pp. 49–55.
- [30] D. Zhou, D. Petrovska-Delacretaz, and B. Dorizzi, "3D Active Shape Model for Automatic Facial Landmark Location Trained with Automatically Generated Landmark Points," In Proc. 20th International Conference on Pattern Recognition (ICPR), Telecom SudParis, Paris, France, Aug. 23-26 2010, pp.3801–3805.
- [31] B. Efraty, M. Papadakis, A. Proffitt, S. Shah, and I. Kakadiaris, "Facial Component Landmark Detection," In Proc. IEEE International Conference on Automatic Face and Gesture Recognition, Mar. 21-25 2011, pp. 278–285.
- [32] M. Dantone, J. Gall, G. Fanelli, and L. van Gool, "Real-time Facial Feature Detection using Conditional Regression Forests," In Proc. 25th IEEE Conference on Computer Vision and Pattern Recognition, June, 2011.
- [33] P. Perakis, T. Theoharis, G. Passalis, and I. Kakadiaris, "3D Facial Landmark Detection Under Large Yaw and Expression Variations," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 7, pp. 1552–1564, July 2013.
- [34] UMB database, <http://www.ivl.disco.unimib.it/umbdb/description.html>
- [35] A. Asthana, S. Zafeiriou, S. Cheng and M. Pantic. Robust Discriminative Response Map Fitting with Constrained Local Models. In CVPR 2013.
- [36] J. D'Hose, J. Colineau, C. Bichon, and B. Dorizzi, "Precise Localization of Landmarks on 3D Faces using Gabor Wavelets," In BTAS 2007, pp. 1-6, Sept. 2007.
- [37] T. C. T Faltemier, K. W. K Bowyer, and P. J. P Flynn, "A Region Ensemble for 3D Face Recognition , "IEEE Trans. Information Forensics and Security, vol. 3, no. 1, pp. 62-73, March 2008.
- [38] F. T. Haar and R. C. R. Velkamp, "Expression Modelling for Expression-invariant Face Recognition," Computers and Graphics, vol. 34, no. 3, pp. 231-241, 2010.
- [39] H. Drira, B. B. Amor, M. Daoudi, and R. Slama, " 3D Face Recognition under Expressions, Occlusions and Pose Variation," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, no. 9, September 2013.