

Face Detection and Pose Estimation Based on Evaluating Facial Feature Selection

^{1,3}Hiyam Hatem, ¹Zou Beiji, ¹Raed Majeed, ²Mohammed Lutf and ¹Jumana Waleed

¹*School of Information Science and Engineering, Central South University, Changsha 410083, China*

²*Department of Electronics and information Engineering, Huazhong University of Science and Technology, Wuhan, China*

³*Department Of Computer Science, Collage of Sciences, Baghdad University, Iraq
hiamhatim2005@yahoo.com, bjzou@vip.163.com, aed.m.muttasher@gmail.com,
jumana_waleed@yahoo.com, mohammed.lutf@gmail.com¹*

Abstract

The detection of faces is one of the most requesting fields of research in image processing and Visual estimation of head pose is desirable for computer vision applications such as face recognition, human computer interaction, and affective computing. In this paper, we propose completed method for face pose estimation, face and face parts detection, feature extraction, tracking. This paper proposes using an improved AdaBoost algorithm, which is much better than normal AdaBoost. We use the de-facto Viola-Jones method for face and face part detection. From the robustness property of Haar-like feature, we first construct the strong classifier more effective to detect rotated face, and then we propose a novel method that can reduce the training time. We adopt affine motion model estimation as a tracking method. The combination enables efficient detection around the search area limited by tracking.

Experimental results demonstrated its effectiveness and robustness against different types of detection and pose estimation in the input face images, including faces that appear in a wide range of image positions and scales, and also complex backgrounds, occlusions, illumination variations and multi-pose head images.

Keywords: *face detection; head pose estimation; Haar –like features*

1. Introduction

Face detection is one of the primary yet sophisticated topics in computer vision and visual pattern recognition communities and it is the first crucial step for facial analysis algorithms and is one of the most important problems in computer vision like head tracking, face recognition, face verification, and facial expression recognition. With the advent of the Internet and low-price digital cameras, as well as a powerful image editing software (such as Adobe Photoshop and Illustrator), ordinary users have more access to the tools of digital doctoring than ever before [1]. It has become a frequent need of our life as it finds its uses in areas like surveillance system, digital monitoring, PC, camera, social networking, cell phones and the like.

Automatic human detection and tracking is an important and challenging field of research and has many application areas. Tracking is the most challenging step of tracking system, which localizes and associates the feature across a series of frames. The increasing use of computer vision in surveillance, replacing human beings, has initiated the research in the field

of face detection. Early research is biased to human recognition rather than tracking. Monitoring the movements of human being raised the need for tracking. Monitoring movements are of high interest in determining the activities of a person and knowing the attention of person [2].

Establishing a system similar to human beings for face recognition is an unmanageable job. Although there has been a significant evolution in this aspect, nevertheless, there is no perfect resolution for all shells, also despite the successes in the last two decades; the state-of-the-art face detectors still have problems in dealing with images in the wild due to large appearance variations.

The performance of various faces based applications, from traditional face recognition and verification to the modern face clustering, tagging and retrieval, relies on accurate and efficient face detection [3].

The goal of face detection is to determine if there are any faces in the image or not and, if present, return the location and the bounding box of each face in the image. Human faces are difficult to model as it is necessary to account for all possible appearance variations caused by changes in scale, location, orientation, facial expression, lighting conditions and partial occlusions, etc. [4].

In spite of all these difficulties, tremendous progress has made in the latest several decades and many systems have shown impressive performance. One of the most remarkable breakthrough made in recent years is the first real-time face detector designed by Viola and Jones [5,6]. The amazing real-time speed and high detection accuracy of Viola and Jones' face detector can attributed to three factors: the integral image representation, the cascade framework, and the use of Adaboost to train cascade nodes.

This paper presents a proposal of an integrated method for face detection, tracking, and head pose estimation, we also focus on both face detection and facial fiducially landmark localization at the same time. We use the de-facto Viola-Jones method for face and face part detection. In addition, the increase of the diversity of input features for Adaboost classifier verified to improve the classification performance of our face detector. The Adaboost algorithm proposed to efficiently select discriminating features (weak classifiers) from a large pool of available features and reinforce them into the final ensemble classifier. We adopt affine motion model estimation, which can accommodate the shape change, as a tracking method.

The rest of this paper organized as follows. Section 2 addresses related works on face detection and gives a brief review of several key issues for face detection. In particular, we focus on feature extraction, feature selection for detecting the face. Section 3 describes the proposed methodology, including features selection and head pose estimation. Section 4 we describe the experiment result .Section 5 we conclude our work with further discussion.

2. Related Work

There has been a multitude of methods for face detection, but one of the most popular methods is the seminal work of Viola and Jones [5-6]. It proposed by Viola and Jones from the University of Cambridge in 200, it proposed a method to combine integral image based Haar -like feature, adaboost based classifier and cascade based fast inference. The amazing real-time speed and high detection accuracy of Viola and Jones' face detector, it became very popular, and many methods offered further enhance it. The result of detection gives the face location parameters and it could be required in various forms, for instance a rectangle covering the central part of the face, eye centers or landmarks in clouding eyes, nose and mouth corners, eyebrows, nostrils, etc.

The most similar work on deformable face representation is [7], which exploited local parts around landmarks for joint face detection, landmark localization and pose estimation with promising performance.

However, there are still problems in [7]. It did not consider the local variation of part and ignored the global structural information. The defined structural model in [7] is not robust to occlusion since the location of parent node may affect the location of its child node. In particular, the detection model in [7] can see as a special case of our face model by removing the hierarchical structure and the part subtype option. Recently, [8-9] proposed methods to speed up part based face detection by cascade classifiers.

The Adaboost algorithm selects Haar-like features and combines them into an ensemble classifier in a cascade mode. The integral image and cascade framework make the detector run fast, whereas Adaboost is the key to obtain a high detection rate for a cascade node. Currently, many face detection systems follow Viola and Jones' cascade-based framework, which computes a great number of weak classifiers formed by Haar-like features at all possible positions and scales in a sliding window and then boosts these weak classifiers into a strong classifier to predict whether, or not a face is present in the window [4].

In view of this problem, Mita *et al.* [10] Suggested constructing a weak classifier using joint Haar-like features that capture co-occurrence of multiple Haar-like features. Nevertheless, the performance improvement in their system is limited, since only simple Haar-like features are considered. Several works [10,11] proposed to explore more homogeneous feature types in order to improve detector's performance. However, expansion of feature numbers and types automatically increases the size of feature set and storage memory. Since feature space enlarges dramatically, obviously, the exhaustive search mechanism used in the standard Adaboost algorithm cannot effectively manage the search process. This in turn makes the training time longer, which is by far one of the main reasons that stop many methods for exploring other feature types.

Compared with these prior works, there are three main contributions of our paper: proposed face detector outperforms the most of the successful face detection algorithms, it shows that including parts improves the detection performance when face images are large and the details of the eyes and mouth does not clearly visible, in addition it accurate estimation of the head pose.

3. Face Detection and Tracking

Face is an important part of human anatomy that varies from individual to individual, but can be located with the use of certain context information. Facial feature detection methods classified into two: local and global methods. In local method, each face component like eyes, mouth and nose detected separately while in global method all facial parts detected jointly and a model based on the relative position of these features is constructed. The goal of face detection method is to determine the presence of face in the frame and return its location. Numerous techniques developed to detect faces in images.

This section explains face and face part detection. Then affine motion model estimation for face tracking explained. Conventional methods used for head pose estimation explained last. The method proposed Viola-Jones detector [5] chosen as a detection algorithm in our framework because of its high detection rate, and its ability to run in real time. This detector is comprised of three main concepts: Employing rectangular Haar-like features and a learning method based on Adaboost, and the attention cascade structure.

3.1. Robustness of Haar-like Features

A Haar-like feature can be defined as the difference of the sum of pixels in two or more adjacent rectangular regions. By changing the position, size, shape and arrangement of these rectangular regions, Haar-like features can capture the intensity gradient at different locations, spatial frequencies and directions [4].

The possible wavelets in the candidate area are vastly numerous. It is inefficient to produce a classifier using all possible wavelets as features. Selecting effective features for detection is necessary. Haar-like features, computed according to the following equation:

$$ii(x, y) - \sum_{x' \leq x, y' \leq y} i(x', y') \quad (1)$$

Where $ii(x, y)$ is the integral image at pixel location (x, y) , and $i(x', y')$ is the original image. Calculation of the sum of a rectangular area in the original image is extremely efficient, requiring only four additions for any arbitrary rectangular size. It can be calculated using the following equation:

$$\sum_{(x,y) \in ABCD} ii(x, y) - ii(A) + ii(D) - ii(B) - ii(C) \quad (2)$$

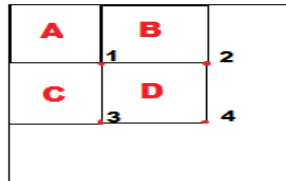


Figure 1. Integral Image at Location 1 is the Sum of Pixels in Region A; at Location 2 Sum of Pixels in Region A+B, at Location 3 C+A and at Location 4 A+B+C+D

Haar-like features can capture the intensity gradient at different locations, spatial frequencies and directions. In [5], Viola and Jones proposed a basic set of four types of Haar-like features for detecting frontal faces. Using all possible sizes, they generated around 180,000 features for a 24×24 pixel image.

In our approach, we adopt a total of nine generalized Haar-like features, including a group of extended Haar-like features proposed in [11] and four basic Haar-like features, to increase the detector's performance. Figure 2 compares the four basic Haar-like features applied by Viola and Jones as well as the generalized Haar-like features used in our approach.

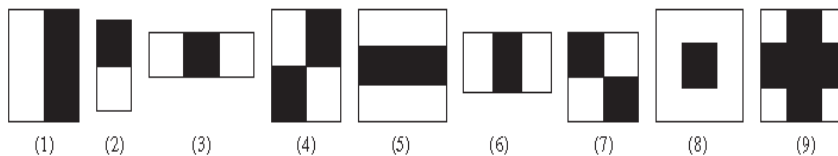


Figure 2. Generalized Haar-like Features used in our Face Detector; (1–4) Original Haar-like Features used by Viola and Jones, (5–9) extended Haar-like Features Proposed by Pham and Cham [11]

3.2. Adaboost for Feature Selection

AdaBoost is a learning classification function which when given a set of features and a training set of positive and negative images, to learn a classification function any number of machine learning approaches could be used. AdaBoost used to train the classifiers as well as to select a small set of features. When used in its real form, AdaBoost learning function boosts the performance of a simple learning algorithm (sometimes called weak). AdaBoost has capability to achieve large margins rapidly that is one of the key features of this algorithm [12].

The method adopts Adaboost as a learning algorithm that constructs a strong classifier composed of weak classifiers. They connected in a cascade. The cascade classifier rejects objects that are apparently false in the early stages, thereby drastically reducing the computation time. Thus, Haar-like feature based Adaboost can effectively learn to construct the best informative weak classifiers from a great number of diverse faces and the convergence of Adaboost also preserved by the robustness. For instance, positions and sizes of eyes of each diverse upright face are different, but a third type Haar-like feature can cover the region under small variances of faces. It used for the construction of strong classifiers as a linear combination of weak classifiers, as shown in the following equation:

$$H(x) = \text{sign} \left(\sum_{t=1}^T a_t h_t(x) \right) \quad (3)$$

Where $h_t(x)$ is a weak classifier, C_{it} is the weight and $H(x)$ is the final strong classifier. In the training process, every weak classifier configured to detect those features that misclassified in previous classifiers [13].

3.3. Attention Cascade Structure

Cascade classifier method is a face detection algorithm based on Haar-like features. A boosted cascade classifier is very popular for rapid classification. However, the raw output of the boosted cascade classifier equilibrated. Although a boosted cascade, classifier is a fast version of Adaboost, the classifier used in limited applications because no one has presented a calibration method for it.

Cascade classifier method is a face detection algorithm based on Haar-like features. The white region's pixel value sum subtracts the black region's pixel value sum can get the characteristic value of feature rectangle, which is used as the face detection's basis [14].

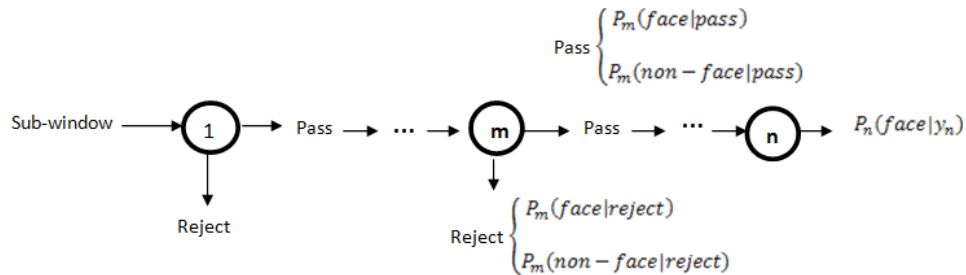


Figure 3. The Process of a Boosted Cascade Classifier: Rejected Sub Windows at Each Stage Subsequently Categorized into Four Groups

As shown in Figure 3 shows the classification process of a boosted cascade classifier. At each stage, all sub windows categorized to only four types; rejected faces, rejected non- faces, passed faces, and passed non-faces. It is a critical component in the Viola-Jones detector, where the main idea behind it is in building smaller boosted classifiers [9]. Each node is a collection of weak classifiers. Nodes are forming a degenerate decision tree, called a cascade.

The number of classifiers in a node usually increases with level, where later stages have more classifiers, because each node tries to pass all the positive sub-windows to further stages, while still rejecting some of the negative ones. Input sub-window passes a series of nodes, where each node makes a binary decision whether to reject it, or pass it on to the next stage. That way, only a small amount of Sub windows will pass through to the latter stages, with most of the negatives rejected early on, thus vastly improving efficiency. The goal of the proposed calibration method is to obtain a posterior probability using the boosted cascade classifier.

3.4. Joint Training of Cascaded Pose Regression

We use the Cascaded Pose Regression (CPR) [15] framework in this work given its efficiency and accurate performance for estimating face landmark locations. We follow the main steps of CPR evaluation procedure. A CPR consists of a cascade of T regressors' $R^{1:T}$. An estimation of a shape starts from an initial guess S^0 , and progressively refines the estimation by an update in each iteration, until the final stage of regression is applied.

Algorithm 1 poses regression

Input: image I, initial pose S^0 , regress $R^{1:T}$
Output: Estimated pose S^T
1: for t=1 to T do
2: $f^t = h^t(I, S^{t-1})$
3: $\Delta S = R^t(f^t)$
4: $S^t = S^{t-1} + \Delta S$
5: end for

As demonstrated in Algorithm 1, where f^t is shape index features, ΔS is apply regress R^t , we use two stages of regression, *i.e.* at each iteration, multiple regresses are utilized and they share the same pose for feature calculation that is from the previous iteration. We also use the random fern as the primitive regresses and follow their training scheme that directly minimizes the alignment error.

We use the interpolated shape-indexed features proposed in [16]. The latter uses a reference location between the locations of two landmarks this is more robust against large pose variations and shape deformations.

4- Experimental Results

In this section, several experiments carried out to evaluate the performance of our face detector. We first demonstrate the superiority of face detection using Haar-like features in terms of the classification efficiency and detection performance. We then show that the Adaboost based feature selection significantly reduces the training time and the number of features required in our system. Finally, we extend our detector to profile faces and discuss

how the detector's accuracy affected by poses variations. We verify the proposed work over two published and widely recognized datasets: ESOGU Face Detection Database [17], Face Detection Data Set and Benchmark [18]. On all the datasets, the proposed method significantly outperforms the state-of-the-art performances, especially on FDDDB. The detection performance of a faster detector with 500 exemplars reported as well to show the proposed approach could have decent performance with a more practical setting.

The Results for various Databases as follow:

-Result on ESOGU Face Detection Database

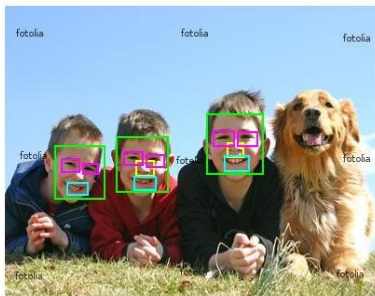
Where the ESOGU face detects from the Eskisehir OsmanGazi University, The original database contained 285 higher-resolution color face images, including faces that appear in a wide range of image positions and scales, also complex backgrounds, occlusions, illumination variations and multi-pose head images. We extended the database by adding 382 new images. Thus, the database now includes 667 images that contain 2042 annotated frontal faces. For many faces, the sizes of the images are too small to search for the parts, thus we up sample the face region in such cases. However, returned parts mostly fail for up sampled low resolution faces. To this end, we visually counted the number of faces returned by the people's tagging tool of this system. We considered all returned face images, as true positives although some faces do not satisfy PASCAL VOC overlapping criterion. We also ignored all false positives coming from the background [17].

-Face Detection Data Set and Benchmark

FDDDB contains 5,171 annotated faces in 2,845 images. Images in this dataset extracted from news articles those present large face appearance variations. The evaluation procedure is standardized and researchers expected to use the same evaluation program to report the results. Due to the annotation mismatch, quite a few true positive detections evaluated as false alarms because their overlapping ratio with the ground truth is just right below the threshold [19].

Experimental conditions are Intel Core i3-2350M CPU, 2.30 GHz, 2G RAM, Windows 7, and MATLAB2012a version. The results are encouraging, and both face and landmark detectors can integrate into face recognition systems without hesitation (most of the face recognition methods

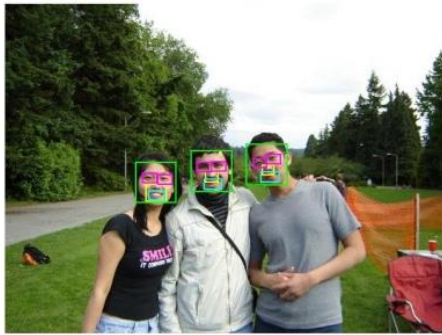
Proposed in the literature relies on manual face alignment and there are only a few studies that implement a fully automatic face recognition system). In addition, results on current face recognition databases are mostly saturated, and it is necessary to introduce more challenging face recognition databases.



(a)



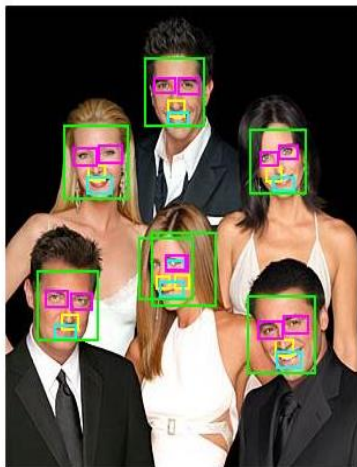
(b)



(c)



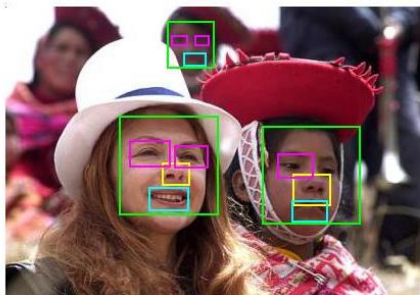
(d)



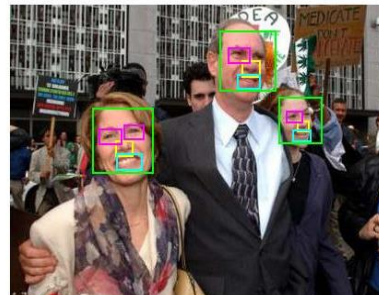
(e)



(f)



(g)



(h)

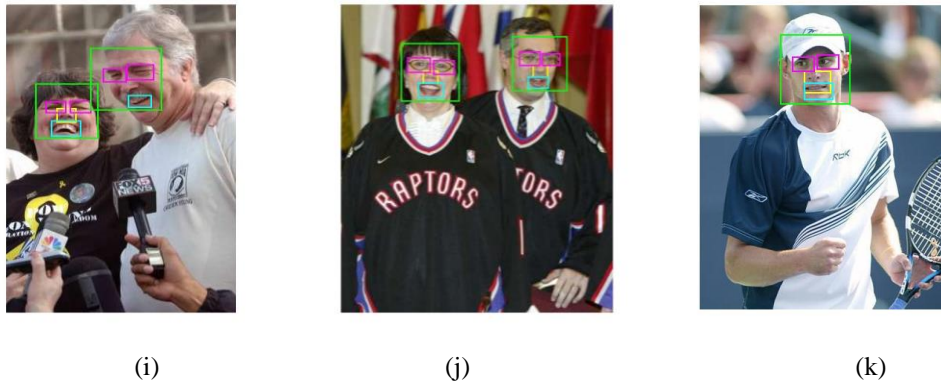


Figure 4. Some examples of the output of our cascade detector of images from the ESOGU face dataset (top three rows) and the Faces in the Wild dataset (the remaining rows). Green rectangles show true human annotations correct detections, and pink rectangles show the eye, yellow rectangle shows the nose and blue rectangle show the mouth.

As shown in Figure 4 Most of the faces and face parts are correctly detected, but there are a few missed detections and false positives. Part detectors typically locate the eyes, nose and mouth correctly if the face image region is large, Part detectors also correct to locate eyes if the person wears sunglasses (*e.g.* Figure 3 (j)) or if the face region not clear (*e.g.* Figure 3 (g)). False detection of face parts often occurs because the standard position and size are not considered. However, there are a few missed of face part detections because of the faces are small, not clear or the face direction in some image are not frontal (*e.g.* Figure 3 (g top), (I right)).

The face parts are detected efficiently only within the face region. The face motion model applies to both the face and face parts. This method results in the robust detection against partial occlusion. The combination of detection and tracking enables efficient detection only around the search area limited by the tracking. It also reduces false detection because of processing continuity between two frames. Because of the combination, a profile face can estimated accurately. In addition, the method re-initializes the position and size of the face and face parts in every frame. The initialization corrects tracking jitter immediately. Furthermore, undetected face revised in terms of position and size using detected face parts.

5. Conclusion

This paper proposed an integrated method for face detection, tracking, and head-pose estimation. The combination of face detection using propose Haar-like feature method to detect people's faces, noses, eyes, mouth ,and tracking by the affine motion model estimation enables detection of a profile face. We proposed head-pose estimation using cascade object detector, we present an efficient boosted exemplar based face detector utilizing exemplar-based weak detector.

This approach makes the exemplar-based face detection practical by largely reducing the number of required exemplars and training discriminative exemplar detectors. Furthermore, by making the definition of exemplar more general to incorporate both face and non-face images, the efforts of collecting exemplars are relieved and negative exemplars can built purposely to suppress false alarms.

Finally, this paper presents a set of detailed experiments on difficult face detection and tracking data set that has been widely studied. This data set includes faces under a wide range of conditions including illumination, scale, and pose and camera variation. Nevertheless the system which work under this algorithm are subjected to the same set of conditions and but the algorithm is flexible enough to adjust according to the changing conditions. Detailed comparisons with the other methods remain as a subject for future work.

Acknowledgements

This research is supported by the National Science Foundation of China under Grant No. 60970098, Modelling for Human Visual System and its application; The National Science Foundation of China under Grant No. 61173122, A Fast Three-dimensional Reconstruction Based on Image by Simulating Mechanisms of Human Visual System.

References

- [1] Y. Ke, W. Min, F. Qin and J. Shang, "Image Forgery Detection Based on Semantics", International Journal of Hybrid Information Technology, vol. 7, no. 1, (2014), pp. 109-124.
- [2] S. V. Tathe and S. P. Narote, "Real-Time Human Detection and Tracking", Annual IEEE India Conference (INDICON), (2013).
- [3] J. Yan, X. Zhang, Z. Lei and S. Z. Li, "Face detection by structural models", Image and Vision Computing, vol. 32, no. 10, (2014), pp. 790-799.
- [4] H. Pan, Y. Zhu and L. Xia, "Efficient and accurate face detection using heterogeneous feature descriptors and feature selection", Computer Vision and Image Understanding, vol. 117, no. 1, (2013) January, pp. 12-28.
- [5] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", in: Proceedings of CVPR 2001 IEEE International Conference on Computer Vision and Pattern, (2001), December 8-14, Kauai, HI, USA.
- [6] P. Viola and M. Jones, "Robust real-time face detection", International journal of computer vision, vol. 57, no. 2, (2004), pp. 137-154.
- [7] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2012), pp. 16-21 Jun, USA.
- [8] H. Cevikalp, B. Triggs and V. Franc, "Face and landmark detection by using cascade of classifiers", Automatic Face & Gesture Recognition, 2013 IEEE International Conference on, (2013), April 22-26, Shanghai.
- [9] J. Yan, X. Zhang, Z. Lei and S. Z. Li, "Real-time high performance deformable model for face detection in the wild", International Conference on Biometric, IEEE, (2013), June 4-7, Madrid.
- [10] T. Mita, T. Kaneko and O. Hori, "Joint Haar-like features for face detection", in: Proceedings of ICCV 2005 10th IEEE International Conference on Computer Vision, (2005), October 17-21, Beijing.
- [11] M. T. Pham and T. J. Cham, "Fast training and selection and Haar features using statistics in boosting-based face detection", in: Proceedings of ICCV 2007 11th IEEE International Conference on Computer Vision, (2007), October 14-21, Rio de Janeiro.
- [12] J. Chatrath, P. Gupta, P. Ahuja, A. Goel and S. M. Arora, "Real Time Human Face Detection and Tracking", International Conference on Signal Processing and Integrated Networks (SPIN), (2014), 20-21 February, Noida.
- [13] H. Leventi, C. Livada and I. Gali, "Towards Fixed Facial Features Face Recognition", 21st International Conference on Systems, Signals and Image Processing, (2014), May 12-15, Dubrovnik.
- [14] M. Ren, S. Zhang, Y. Lei and M. Zhang, "CUDA-based Real-time Face Recognition System", Digital Information and Communication Technology and its Applications (DICTAP), 2014 Fourth International Conference on, (2014), May 6-8, Bangkok.
- [15] P. Dollár, P. Welinder and P. Perona, "Cascaded pose regression", In Proc. IEEE Conf. Computer Vision and Pattern Recognition, (2010), June 13-18, San Francisco, CA.
- [16] X. P. Burgos-Artizzu, P. Perona, and P. Dollár, "Robust face landmark estimation under occlusion", In Proc. IEEE Conf. Computer Vision, (2013), December 1-8, Sydney, VIC.
- [17] <http://mlcvdb.ogu.edu.tr/facedetection.html>
- [18] V. Jain and E. Learned-Miller, "Fddb: A benchmark for face detection in unconstrained settings", Technical report, (2010).

- [19] L. Haoxiang, L. Zhe, J. Brandtz, S. Xiaohui and H. Gang, "Efficient Boosted Exemplar-based Face Detection", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2014), June 23-28, Columbus, OH.

Authors



Hiyam Hatem, she received the BS degree from the department of computer science, Baghdad university, Iraq, 2003. The master degree from Huazhong University of science and technology Wuhan, China, in 2010. Currently, she is a PhD student in School of information science and Engineering at Central south university, Changsha, China. Her research interests include face processing and recognition, object detection, pattern recognition, computer vision, and biometrics.



Zou Beiji, he holds a Ph.D. in Computer science and technology from Hunan University, China. He is currently a Professor in School of Information Science and Engineering of Central South University, China. He worked at Zhejiang University, China as a visiting scholar from Jul.1999 to Jun. 2000 and at Tsinghua University, China as a post-doctor from Jan. 2002 to Nov. 2003 respectively, engaged in research about Human Facial Expressional Recognition and Animation. He worked in the research field of Multimedia Technology at Griffith University in Australia from Dec. of 2003 to Dec. of 2004. Interest area of research: Digital Image Processing, Computer Graphics, Multimedia technology, Software Engineering.



Raed Majeed Muttasher, he received his B.Sc. in Computer Science from Baghdad University, colleges of science, computer department in 2004. He received the Master Degree in applied computer technology from Wuhan University, school of computer in 2011, P.R. China. He is currently working toward his Ph.D. degree at Central South University, School Of Information Science And Engineering, P. R. China. His research interests include 3D Object Recognition, 3D Modeling, Pattern Recognition, and Image Processing.



Mohammed Lutf, he is a Ph.D. student in the Department of Electronics and Information Engineering at Huazhong University of Science and Technology, Wuhan, China. He is an active researcher in the area of Arabic character and handwriting recognition. He received his B.S. Degree in Telecommunication Engineering from Dalian Maritime University, Dalian, China, in 2005 and an M.S. Degree in Communication and Information Systems from Huazhong University of Science and Technology, Wuhan, China, in 2010.



Jumana Waleed, is a Ph.D. student in the School of information science and Engineering at Central South University, Changsha, China. Her research activity focuses on image processing, and information security working on digital watermarking. She received the B.S. degree in computers sciences from the Al-Yarmouk University College, Iraq, in 2004, and the M.S. degree in Computer Science/Data Security from the University of Technology, Baghdad, Iraq, in 2009.