

A Robust Approach of Facial Orientation Recognition from Facial Features

Kishor Datta Gupta

Department of Computer Science,
Lamar University,
4400 S M L King Jr Pkwy, Beaumont, TX 77705, USA
Phone: +1 409-880-7011
kgupta@lamar.edu

Md Manjurul Ahsan

Department of Industrial Engineering,
Lamar University,
4400 S M L King Jr Pkwy, Beaumont, TX 77705, USA
Phone: +1 409-880-7011
mahsan2@lamar.edu

Stefan Andrei

Department of Computer Science,
Lamar University,
4400 S M L King Jr Pkwy, Beaumont, TX 77705, USA
Phone: +1 409-880-7011
stefan.andrei@lamar.edu

Kazi Md. Rokibul Alam

Department of Computer Science and Engineering,
Khulna University of Engineering and Technology,
Khulna 9203, Bangladesh
Phone: +880 41-769468
rokib@cse.kuet.ac.bd

Abstract

Face orientation recognition is an important topic in computer vision and pattern recognition. Due to the non-rigid properties of faces, it is computationally expensive and difficult to achieve good recognition accuracy and robustness in face orientation recognition. In this paper, we propose an image mapping technique for face analysis in smart camera networks with a feature extraction and data from the facial feature. We estimate the face orientation angles in all camera views, based on the matched imaged data. Our objective is to obtain a set of facial structures which can work as landmarks for tracking and recognition of facial expressions.

Keywords: face detection, recognition, computer vision, human-computer interaction, image

1. Introduction

Most face recognition and tracking techniques employed in surveillance and human-computer interaction (HCI) systems rely on the assumption of a frontal view of the human face. In alternative approaches, knowledge of the orientation angle of the face in captured images can improve the performance of techniques based on non-frontal face views. It broadly consists of three parts: first, the face is detected by Haar detection based face detection method; then the face is tracked robustly using four extracted facial features; and finally, the orientation of the face is estimated by using the tracking results obtained independently from the three trackers (Viola and Jones; 2001). First, we use Haar detection method to identify a face area from any picture. The accuracy of this methodology is higher than 90% and very reliable to detect faces, we develop a new algorithm for face orientation recognition. The algorithm based on the combination of four individual tracking-based face orientation estimators that are relied on the seven properties of the face in question respectively: the variation of face regions,

the deformation of face texture patterns, and the trajectory of face motion, Eyes position, mouth position, lip, and nose position (Chihaoui et al.; 2016) (Chutorian et al.; 2007). The combination is achieved by the make data in printed graph type images and compares images by image matching techniques. The algorithm is reliable and able to estimate face orientation efficiently.

2. Problem statement

Despite the success of the existing approaches/systems for extracting face's characteristics using computer vision technologies, current efforts in this area focus on using only a single visual cue such as eyelid movement or line of sight or head orientation to characterize face's state of alertness. The system relying on a single visual cue may encounter difficulty when the required visual features cannot be acquired accurately or reliably (Schiele and Sagerer; 2001). For example, faces with glasses could pose a serious problem to those techniques based on detecting eye characteristics. Glasses can cause glare and may be totally opaque to light, making it impossible for the camera to monitor eye movement. Furthermore, the degree of eye openness may vary from people to people. Another potential problem with the use of a single visual cue is that the obtained visual feature may not always be indicative of one's mental conditions. For example, the irregular head movement or line of sight (like briefly look back or at the mirror) may yield false alarms for such a system. All those visual cues, however imperfect they are individual, if combined systematically, can provide an accurate characterization of a face's level of vigilance. It is our belief that simultaneous extraction and use of multiple visual cues can reduce the uncertainty and resolve the ambiguity present in the information from a single source. The system we propose can simultaneously, non-intrusively, and in real-time monitor several visual behaviors which usually characterize a person's level of alertness while driving (Tian et al.; 1999). These visual cues include eyelid movement, pupil movement, and face orientation. The fatigue parameters computed from this visual cue are subsequently combined probabilistically to form a composite fatigue index that can robustly, accurately, and consistently characterize one's vigilance level (Zhu et al.; 2004).

3. Methodology

The methodology is composed of two main phases such as Face Feature Extraction and create graph image and Matching Graph image with stored images.

3.1. Face Feature Extraction

It consists of four sequential steps such as Face Detection, Feature Extraction, gets features data and Creates image with these data.

3.1.1. Face detection

The core basis for Haar classifier object detection is identifying the Haar-like features. First, Haar classifier cascades are trained for detecting human facial features, such as the mouth, eyes, and nose (Wilson and Fernandez; 2006) (Chihaoui et al.; 2016). To train the classifiers, this gentle AdaBoost algorithm and Haar feature algorithms must be implemented (Zhu et al.; 2004). Intel developed an open source library devoted to easing the implementation of computer vision related programs called Open Computer Vision Library. OpenCV is designed to be used in conjunction with applications that pertain to the field of HCI, robotics, biometrics, image processing, and other areas where visualization is Important and includes an implementation of Haar classifier detection and training (Wanjale et al; 2013). To train the classifiers, two set of images are needed. One set contains an image or a scene that does not contain the object, in this case, a facial feature, which is going to be detected. This set of images is referred to as the negative images. The other set of images, the positive images, contain one or more instances of the object. The location of the objects within the positive images is specified by image name, the upper left pixel, the height, and width of the object. For training, facial features 5,000 negative images with at least one megapixel resolution are used for training (Niese et al.; 2006). These images consisted of everyday objects, like paper

clips, and of natural scenery, such as photographs of forests and mountains. To produce the most robust facial feature detection possible, the original positive set of images needs to be representative of the variance between different people, including, race, gender, and age. A good source for these images are the National Institute of Standards and Technology's (NIST) and the Facial Recognition Technology (FERET) database. This database contains over 10,000 images of over 1,000 people under different lighting conditions, poses, and angles (Wilson and Fernandez; 2006). In training each facial feature, 1,500 images were used. These images were taken at angles ranging from zero to forty-five degrees from a frontal view. This provides the needed variance required to allow detection if the head is turned slightly. Three separate classifiers were trained, one for the eyes, one for the nose, and one for the mouth. Once the classifiers were trained, then it used to detect the facial features within another set of images from the FERET database. The accuracy of the classifier was then computed as shown in Table 1. Except for the mouth classifier, the classifiers have a high rate of detection. However, as implied by, the false positive rate is also quite high.

Table 1. Accuracy of Classifiers

Facial Feature	Positive Hit Rate	Negative Hit Rate
Eyes	93%	23%
Nose	100%	29%
Mouth	67%	28%

3.1.2 Feature extraction

The first step in facial feature detection is detecting the face. This requires analyzing the entire image. The second step is using the isolated face(s) to detect each feature. The result is shown in Figure 1. Since each portion of the image used to detect a feature is much smaller than that of the whole image, detection of all three facial features takes less time on average than detecting the face itself. Using a 1.2GHz AMD processor to analyze a 320 by 240 image, a frame rate of 3 frames per second was achieved. Since a frame rate of 5 frames per second was achieved in facial detection only by using a much faster processor, regionalization provides a tremendous increase in efficiency in facial feature detection.



Figure 1. Face area detection

3.1.3. Get features data

Because of facial symmetry, the horizontal pose can be estimated with positions of both eyes about the face. Our technique is dependent on four features of a face: Right Eye, Left Eye, Mouth, and Nose. It uses respective distance and size of these features for finding the orientation of the faces. One of the most important phase is to get the positions, size, height, width, angle of these features respective to faces.

3.1.4 Create image from data

Using these data, it creates a new picture and uses these data as a starting point for drawing. By using the data, it gets a model and shape of face without color and facial expression (Gourier et al.; 2004), such as Figure 2. It got a model of faces using these features.

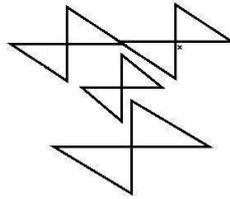


Figure 2. A Sample Image from extracted feature

3.2. Matching Graph image

It calculates two images from all pixels, and calculates the distance between two images' pixels. After summing up the distance using a threshold point it determines whether the images can match or not. If the images are close to the stored image, it passes as recognized. If image A1 has a black point X_a and Y_a and the reference image R1 has a black point is X_r and Y_r , then our calculated distance is: $\sum \sum (X_a - X_r) + (Y_a - Y_r)$.

3.2.1 Orientation Recognition

After matching the images with the reference image, it gets the nearest orientation matches and they could be Font left, Font right, Down-left, Down Right, Up left, Upright, Font Straight, Up Straight, Down Straight. Initially, some constraints must be satisfied to realize a successful correct matching. The facial regions, concerned on eyes and nose points, have the following characteristic: if there is almost one missing point for the region of the same type then the comparison will be performed. There must be the same number of feature points for both eyes and nose separately. If this condition is satisfied then a new comparison will be performed.

4. Experimental Analysis

The methodology of our experimental method is described in Figure 3 below.

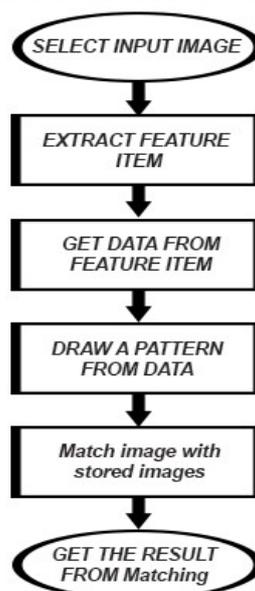


Figure 3. The methodology flowchart

It helps to write our code in C# and to make an application in dot net framework, which collects facial images using a webcam/or other video grabbing tools. Then it implements Haar detection to extract facial features and to draw image pattern for matching both images.

5. Results

We list below our results (Figures 4 and 5) supported by Table 2.

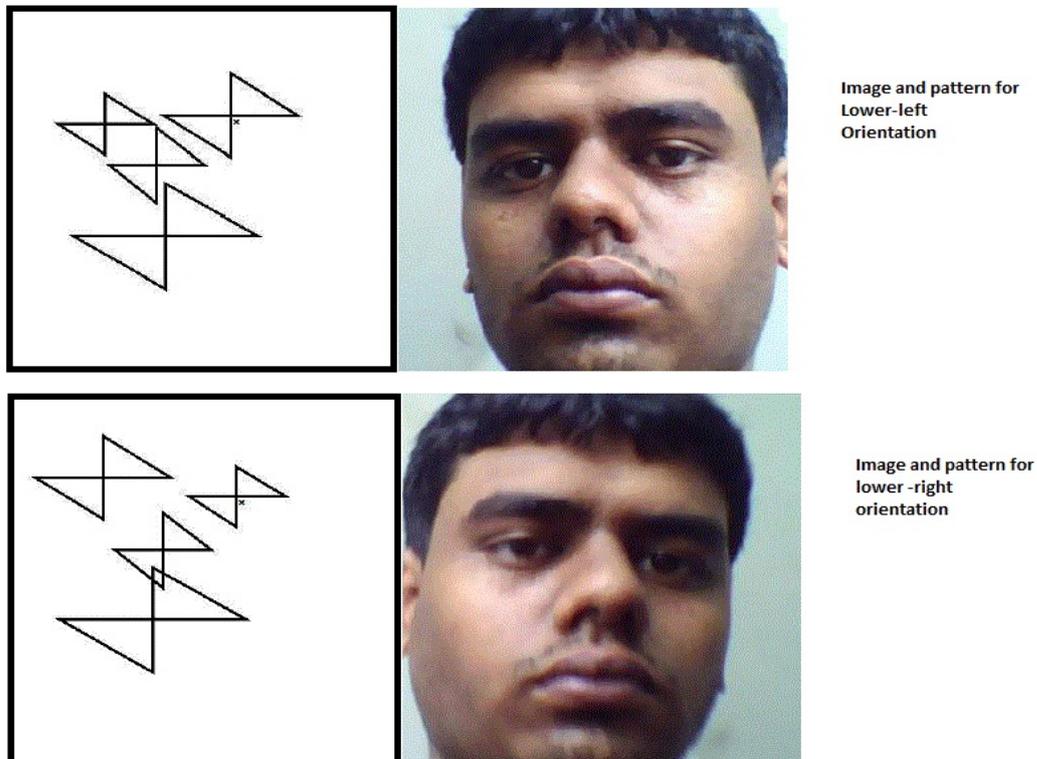
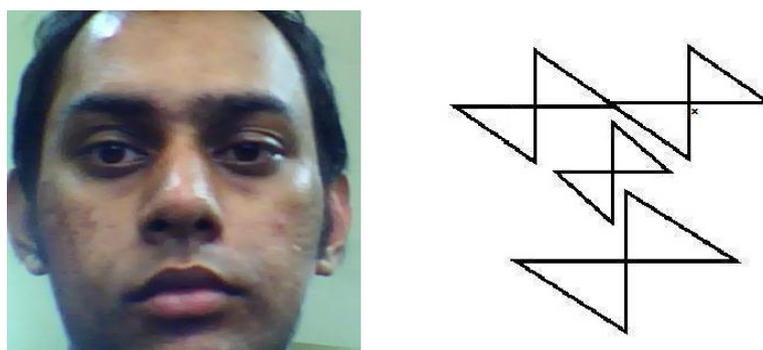


Figure 4. Sample image data 1

And for frontal we get as (Figure 5):



For frontal face

Figure 5. Sample Image data 2

After image matching, we got a positive result at 93% times, for 1000 random sample images tested on the nine criteria of orientation.

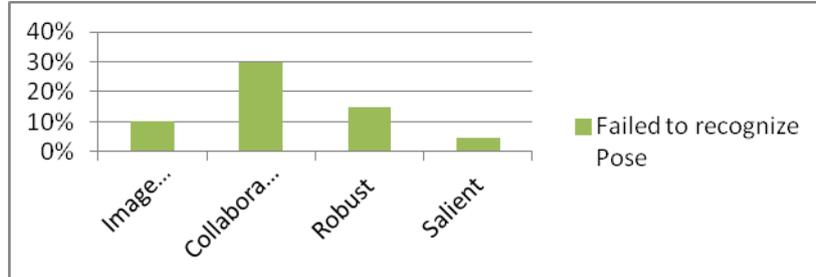
Table 2. Statistics of our matching Result

Face orientation	Total	Missed	Positive	Success Rates
Font Straight	150	4	144	99%
Up Straight	110	20	90	92%
Down Straight	100	23	77	77%
Font left	100	12	88	88%
Font Right	120	18	102	82%
Up left	120	21	89	91%
Up right	115	22	93	95%
Down right	100	38	62	62%
Down left	95	44	51	54%

5.1. Comparisons with other Methods

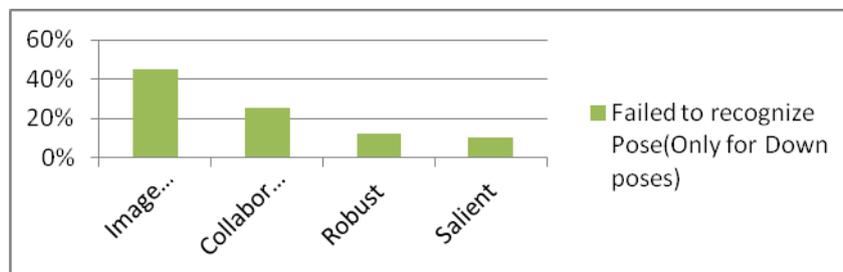
If we make a comparison with three other models, some of these technologies are Multimodal Head estimation, Collaborative Face Orientation Detection. Estimating Face orientation is derived from Robust Detection of Salient Facial Structures. Table 3 shows the percentages of failing to recognize pose.

Table 3.



If we consider only the down face type result (that is, the negative impact on our feature), then the graph of failure rate of our method is higher than other (Table 4).

Table 4.



However, if we consider without the down face type result, then the graph of failure rate of our method is lower than other as shown in Table 5.

Table 5.



As a consequence of our experimental results, our method has a highly improved performance when we do not consider downside pose.

6. Conclusion

Template matching is a fundamental task in the field of image processing because it is applicable to numerous different tasks such as object detection and categorization due to its simplicity of implementation. However, some authors also provide the different algorithm for simultaneous face detection, landmarks localization using deep convolutional neural networks (CNN), which can locate global and local information in faces (Ranjan et al.; 2016). Some authors also suggest the idea of normalizing the salient areas to align the specific areas (Liu et al.; 2017). In this paper, we propose an image matching technique with increased success rate of getting proper pose estimation for human faces. It could be used for artificial intelligence, and game controller development, also can be used in traffic control and robot development. Also, our method is more improved and faster, it needs less stored data and can work faster in a short period of time. Currently, several profit, non-profit organizations are working on face recognition from video evaluation. NIST (National Institute of Standards and Technology) are also working on accuracy and speed of face recognition algorithms applied to the identification of the persons appearing in the video (Grother et al.; 2017).

References

- Chihaoui, M., Elkefi, A., Bellil, W., & Amar, C. B. (2016). A Survey of 2D Face Recognition Techniques, *Computers*, Vol. 5, No. 21, 28 pages.
- Chutorian, E., Doshi, A., & Trivedi, M. (2007). Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation. In *IEEE Intelligent Transportation Systems Conference*, pages 709–714.
- Gourier, N., Hall, D., & Crowley, J. L. (2004). Facial features detection robust to pose, illumination and identity. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, vol. 1, pp. 617-622. IEEE, 2004.
- Grother, P. J., Ngan, M. L., & Quinn, G. W. (2017). *Face In Video Evaluation (FIVE) Face Recognition of Non-Cooperative Subjects*. NIST Interagency/Internal Report (NISTIR) – 8173.
- Liu, Y., Li, Y., Ma, X., & Song, R. (2017). Facial Expression Recognition with Fusion Features Extracted from Salient Facial Areas. *Sensors* 17, no. 4 (2017): 712.
- Niese, R., Al-Hamadi, A., & Michaelis, B. (2006). A stereo and color-based method for face pose estimation and facial feature extraction. In *IEEE International Conference on Pattern Recognition*, pages 299–302.
- Ranjan, R., Patel, V. M., & Chellappa, R. (2016). Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *arXiv preprint arXiv:1603.01249*.

- Schiele, B. & Sagerer, G. (2001). Computer Vision Systems, *Second International Workshop, ICVS 2001 Vancouver, Canada, July 7-8, 2001 Proceedings*. Vol. 2. Springer Science & Business Media.
- Viola, P. & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Conference on Computer Vision and Pattern Recognition*, pages 511–8.
- Wanjale, K. H., Bhoomkar, A., Kulkarni, A., & Gosavi, S. (2013, April). Use of Haar Cascade Classifier for Face Tracking System in Real Time Video. In *International Journal of Engineering Research and Technology*, vol. 2, no. 4. IJERT.
- Wilson, P. I. & John, F. (2006). Facial feature detection using Haar classifiers. *Journal of Computing Sciences in Colleges* 21, no. 4: 127-133.
- Ying-li, T., Kanade, T., & Cohn, J. F. (1999, December). Recognizing Action Unit for Facial Expression Analysis. CMU-RI-TR-99-40 *Technical Report of Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213*. Retrieved from <http://www.cs.cmu.edu/~face/Papers/CMU-RI-TR-99-40.pdf>
- Zhu, Z., Ji, Q., & Lan, P. (2004). Real time non-intrusive monitoring and prediction of driver fatigue. *IEEE Transactions on Vehicular Technologies* 53, no. 4: 1052-1068.