Intelligent Transportation and Smart Cities
V. Mezhuyev et al. (Eds.)

doi:10.3233/ATDE240371

Intelligent Recognition and Detection Method for Facial Posture of Motor Vehicle Drivers Based on Computer Vision

Yali CAO[1]
School of Art and Design, Wuhan University of Science and Technology, Wuhan, Hubei, China

Abstract. The driver monitoring system constitutes an essential element in the realm of human-machine interaction within intelligent vehicles, primarily tasked with overseeing and promptly alerting deviations from standard driving behaviors that could potentially lead to traffic accidents. Presently, the evolution of driver monitoring systems in China is at a nascent stage. Challenges persist due to constraints in hardware equipment, resulting in relatively simplistic terminal devices for detecting driver fatigue features. Consequently, this has led to frequent occurrences of false positives and missed detections. This study centers around a non-contact vehicle monitoring device employing machine vision. It delves into the development of a rational, effective, real-time, and accurate mechanism for continuously monitoring driving duration while concurrently recording driving behavior data. Within this framework, the research focuses on the utilization of a modified version of MTCNN for real-time driver face detection. It involves extracting critical driver head posture features, encompassing the pitch angle, yaw angle, and roll angle of the driver's head. By comparing the positions of obtained driver face images, a fatigue driving feature index system is established. This system facilitates the identification, analysis, and discrimination of a driver's fatigue state based on the extracted head feature values. The ultimate aim is to realize a fatigue warning function within the driver monitoring system, thereby enhancing the detection speed of facial fatigue recognition. This endeavor holds paramount significance for road traffic safety, contributing to the continual improvement of driver monitoring systems and consequently mitigating potential risks on the road.

Keywords. Computer vision, motor vehicle driving, intelligent recognition of facial posture, MTCNN algorithm

1. Introduction

Based on data disseminated by the World Health Organization, fatigue driving is responsible for tens of thousands of traffic accidents globally each year, leading to a significant number of casualties. The origins of driver fatigue driving encompass factors such as inadequate sleep and extended periods of driving, culminating in diminished self-control, delayed judgment, and compromised reaction ability due to distraction. This confluence of factors poses challenges in predicting and assessing the occurrence of fatigue driving. Addressing the imperative need to promptly and effectively gauge the

[1] Corresponding Author: Yali CAO, 122163601@qq.com.

fatigue levels of drivers and mitigate traffic accidents resulting from fatigue driving has emerged as a focal point in contemporary intelligent traffic safety systems research.

Within the domain of face recognition, extensive research endeavors have been pursued. Jinbao Li leverages infrared images acquired from depth cameras in conjunction with Local Binary Pattern (LBP) features of faces. This amalgamation facilitates face detection achieved through feature filtering employing a cascaded classifier powered by the Ada Boost algorithm [1]. Shan Zhang, on the other hand, introduces a motion recognition scheme founded on facial analysis, spanning from a comprehensive view to localized scrutiny [2]. Dewei Zheng, in his work, extracts fatigue features by utilizing an optimal set of facial feature points and subsequently employs statistical analysis methods to scrutinize and validate the efficacy of these features [3]. Additionally, some papers employs artificial intelligence (AI)-based face recognition utilizing deep learning techniques [4,5].

In our research, we employ a modified version of the multi-task convolutional neural network (MTCNN), integrating both facial and head features for enhanced fatigue recognition. This approach mitigates the issue of elevated false alarm rates associated with reliance on a single feature, concurrently diminishing interference stemming from inaccuracies in specific feature indicators. Our decision to select the MTCNN algorithm focuses on facial detection of drivers, with subsequent cropping of the driver's facial image. Surveillance cameras capture driver images encompassing facial data, as well as images of the driver's upper body and surroundings. The inclusion of extraneous elements poses challenges to the accuracy of driver facial and fatigue recognition, while also impeding the algorithm's recognition speed. Therefore, employing the MTCNN algorithm for driver facial area detection and subsequent isolation of the facial segment not only streamlines facial and fatigue recognition processes but also enhances overall recognition accuracy and speed.

2. The Principle of Face Detection Algorithm for MTCNN

Multi-task convolutional neural networks can achieve efficient face detection through cascaded network models [6]. The MTCNN algorithm can simultaneously complete tasks of face detection and alignment. Compared to traditional algorithms, it has better performance and faster detection speed. The MTCNN neural network belongs to a cascaded system, which uses three levels of deep convolutional networks to predict facial positions and key point positions in a coarse-grained to fine-grained manner. It consists of P-Net, R-Net, O-Net [7].

The basic principle of MTCNN face detection is to construct a funnel shaped detector through a cascaded three-level convolutional neural network. Firstly, a fully convolutional neural network is used to extract candidate facial images, which is a first level network; The secondary network is used to classify the image blocks provided by the first network, which can filter facial candidate images more strictly; The three-level network is used to filter the image blocks provided by the two-level network and locate the position of facial key points. The implementation process of MTCNN facial detection algorithm is as Figure 1.

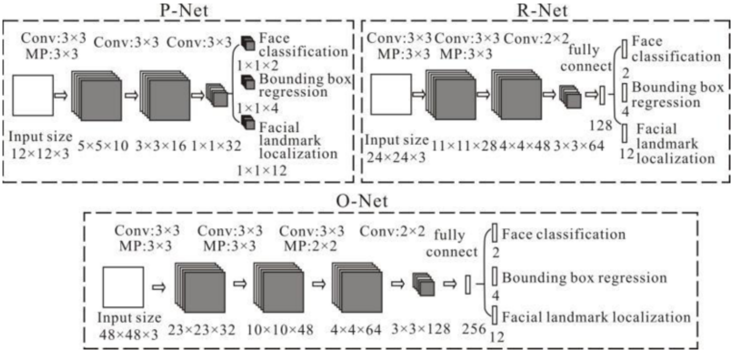

Figure 1. The implementation process of MTCNN facial detection algorithm.

2.1. Image Pyramid

Resize the image to different scales and generate an image pyramid. Then, images of different scales are fed into three sub networks for training, with the aim of detecting faces of different sizes and achieving multi-scale object detection (figure 2). The multi-scale representation of pyramid images is mainly used for image segmentation and fusion. By sampling the original image, images of different sizes are generated to prevent the omission of undetected faces.

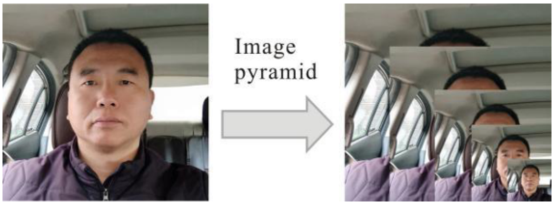

Figure 2. Image pyramid.

2.2. P-Net Network

By constructing a fully convolutional network to predict the position of faces in images of any scale, and obtaining face candidate boxes at coarse granularity through P-Net, various sizes of face bounding boxes will be obtained, and these boxes may not be accurate. At this point, Non-Maximum Suppression (NMS) technology is needed to complete the deduplication of the candidate face boxes. After data deduplication, the coordinate correction offset that can be used to adjust the candidate facial bounding boxes will be provided, as shown in the figure 3.

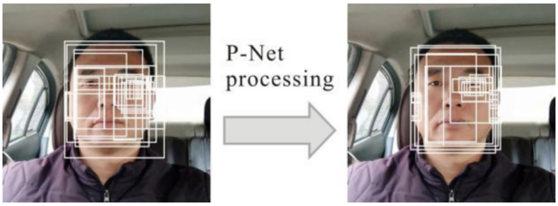

Figure 3. P-Net network.

2.3. R-Net Network

After adding a fully connected layer at the end of the convolutional layer, the resolution of the image has been improved, and the number of layers in the network has also increased. The R-Net neural network will make fine-grained improvements to the output of the P-Net neural network, judging whether the candidate face bounding boxes output by the P-Net are faces, thereby filtering out non-face candidate boxes, providing the offset of the candidate boxes, and then correcting the face boxes through Bounding box regression. NMS technology is used to remove duplicate candidate boxes, optimize the regression and localization of face bounding boxes, as shown in the figure 4.

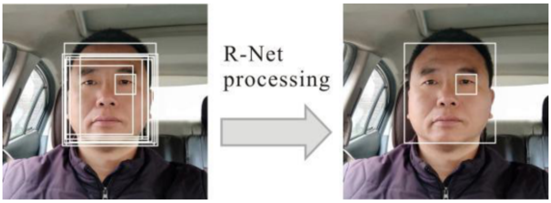

Figure 4. R-Net network.

2.4. O-Net Network

Perform stricter judgments on the facial images output by the R-Net neural network to obtain precise facial positions and key point positions. The O-Net neural network retains more image information while increasing the resolution of the input image, resulting in a more reliable output. Determine whether the candidate's face bounding box is a face, then correct the face box through bounding box regression, and finally complete the data deduplication through NMS. In addition, the O-Net neural network also provides the coordinates of the face's eyes, nose, and the left and right corners of the mouth, as shown in the figure 5.

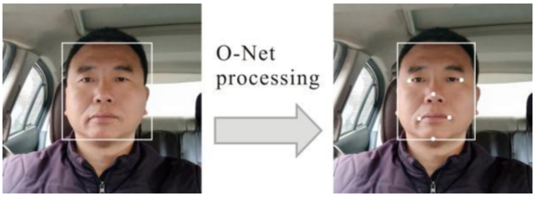

Figure 5. O-Net network.

For the P-Net network, it consists of three output layers, with the first output layer being a human face classifier, which belongs to the classification task; The second output layer is facial frame regression, and the offset result of candidate box correction belongs to the regression task; The third output result is the coordinates of the six key points of the face, which belong to the regression task [8]. The cross entropy cost function can be used to handle classification requirements; L2 loss can be used to handle regression requirements. Each output layer will set weights according to the degree of need, and the losses of each layer will be accumulated according to the weights to calculate the overall loss. The first two layers of the network have a greater weight in the classification task, while the weight of the facial coordinate localization key point calibration in O-Net will be set higher.

3. Principles of Facial Recognition

The driver's image is captured through the camera, and after image preprocessing, the MTCNN algorithm is called to complete the driver's facial detection, crop out the facial image, and correct the facial position.

3.1. Facial Feature Point Alignment

Due to the fact that during the process of collecting driver facial data, the driver may not necessarily be facing the camera. Therefore, when using the MTCNN algorithm to capture facial images, there are various head postures that will affect the computational efficiency of subsequent facial recognition. Therefore, facial alignment operation is an important prerequisite for completing facial recognition.

The driver's facial alignment method calculates parameters such as facial rotation angle by mapping key points of the nose tip, left eye left corner, right eye right corner, left corner of mouth, right corner of mouth and sharp angle of chin, in order to achieve facial alignment [9]. Firstly, obtain a standard face model as the benchmark, and then converge the benchmark face area through shrinkage operation to obtain a face image that can accommodate a larger face area, as shown in the figure 6.

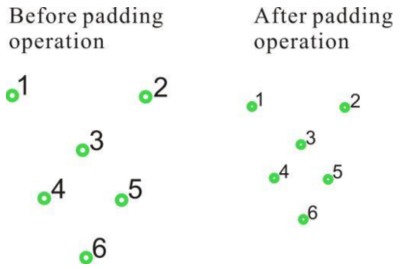

Figure 6. Operation of facial feature point alignment.

Firstly, among all the feature points after the contraction operation, focus on the feature points with numbers ranging from 1 to 6, which correspond to the coordinates of left eye left corner, right eye right corner, nose tip, left corner of mouth, right corner of mouth and chin, as the standard comparison face.

At the same time, real-time coordinates of the driver's facial position are obtained through feature point detection. Then, by analyzing and comparing the key position coordinates of the benchmark face and the real-time face, the transformation matrix is obtained. The calculation process is as follows:

Let $(x_j, y_j)^T$ be the position of the jth keypoint on the face, and $(x_j, y_j)^T$ be the corrected position of the corresponding keypoint on the face. The affine calculation formula (1) is as follows:

$$\begin{bmatrix} x_j' \\ y_j' \end{bmatrix} = A \begin{bmatrix} x_j \\ y_j \end{bmatrix} + B \tag{1}$$

The calculation formulas (2) for parameters A and B in the above equation are as follows:

$$A = \begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix}_{2\times2} \quad B = \begin{bmatrix} b_{00} \\ b_{10} \end{bmatrix}_{2\times1} \tag{2}$$

Finally, the driver's facial image is aligned by calculating the required rotation angle, scaling factor, and other parameters for converting the real-time facial image to the reference face.

3.2. Experimental Objects and Platforms for Driving Behavior Recognition

Related literature indicates a certain correlation between actual vehicle testing and simulated driving [10], therefore, this experiment decided to validate the fatigue recognition algorithm through a car simulation driving platform.

This experiment uses a two seat simulation driving platform, which mainly includes a simulation control system, a visual simulation system, and a vehicle dynamics feedback system. The experimental platform can be divided into two parts: hardware system and software system. The hardware system includes: steering wheel, accelerator pedal, brake pedal, in car rearview mirror, visual display, etc; the software system includes Sim Creator scene modeling system, vehicle script behavior control system, and vehicle dynamics and motion control software. During the experiment, the entire car was manually driven by the driver, allowing the driver to better enter the driving state and obtain more realistic experimental data.

3.2.1. The situation of the experimental driver

Related studies have shown that young drivers are four times more likely to enter a state of fatigue than other age groups [11]. Therefore, the participants in this experiment are mainly young male drivers.

3.2.2. Simulation of experimental scenarios

The road scene environment in this experiment was simulated and generated by the Center Channel computer, and the road scene environment designed by the visual simulation system was projected onto a circular screen through a projector. Related studies have found that drivers are more likely to enter a fatigue state on straight driving roads. In order to help drivers enter a state of fatigue driving as soon as possible, this experimental scenario is set as a straight highway, with a length of 90km and a lane width of 3.5m. The speed is maintained at 90km/h, and the lane is designed as a two-way four lane highway.

3.2.3. The process of experimental operation

The main experimenter explained the purpose and detailed process of this experiment to the participants. Introduce the usage of simulated driving instruments to the experimenters, inform them of the precautions and operations to be completed during the driving process. Let the participants enter the driving simulator to adjust the seat height and backrest angle suitable for the driver, so that the driver is in a suitable driving state and has a good field of vision. The main test personnel ran a simulated driving scenario and driving simulator, allowing the participants to test drive for 8 minutes and familiarize themselves with the steering wheel operation. The entire experimental time is about 60 minutes, and participants are required to maintain a speed of 90km/h as much as possible while complying with traffic rules.

4. Facial Feature Recognition and Data Analysis

Numerous works have been conducted in the domain of head pose estimation [12,13]. In our study, real-time recording of the driver's facial video was completed through a camera installed in the cab facing the driver's face, and data analysis was conducted on facial features.

During the simulated driving training process of the experimenters, it was found that when the driver's fatigue level is deep, there will be changes in head posture, such as frequent nodding movements, to identify the driver's fatigue state. And through facial pose estimation algorithms, changes in the Euler angle of the driver's head can be obtained [14], thereby completing recognition of actions such as nodding and tilting.

The head pose estimation algorithm mainly obtains the angle information of facial orientation. Specifically, it is based on machine vision methods to infer head posture information from facial images, namely calculating three Euler angles: pitch, yaw, and roll [15]. The pitch angle is a nodding motion rotating on the X-axis, the roll angle is a shaking motion rotating on the Y-axis, and the yaw angle is a shaking motion rotating on the Z-axis, as shown in the figure 7.

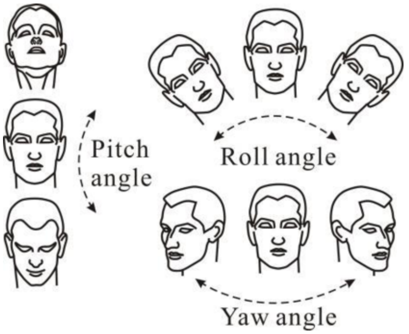

Figure 7. Three types of angle information for head posture.

Among them, the pitch angle can recognize the driver's nodding action, while the yaw angle can recognize the driver's actions that are not visible to the front for a long time, providing a basis for identifying subsequent unsafe driving behaviors.

Considering the hardware conditions of embedded devices and the real-time requirements of algorithm processing, based on facial feature point detection, 3D pose changes are inferred based on 2D coordinates to obtain the pitch angle of the driver's head. Obtain the coordinates of 2D facial key points and select a total of 6 key points, including the nose tip, left eye left corner, right eye right corner, left corner of mouth, right corner of mouth and sharp angle of chin, as the 2D facial model.

Face model matching is performed in a three-dimensional plane, using the 6 key points of a standard 3D face model that correspond to the 2D face key points. The specific 2D-3D facial model coordinates are shown in the Table 1.

Table 1. 2D-3D facial model coordinates

Facial position	3D coordinates
Left eye left corner	(-165.0, 170.0, -135.0)
Right eye right corner	(165.0, 170.0, -135.0)
Nose tip	(0.0, 0.0, 0.0)
Left corner of mouth	(-150.0, -150.0, -125.0)
Right corner of mouth	(150.0, -150.0, -125.0)
Sharp angle of chin	(0.0, -330.0, -65.0)

Finally, solve the mapping relationship between the key points of 2D and 3D facial models, and convert the world coordinate system into a camera coordinate system. Assuming that the position of 3D points in world coordinates (U, V, W) is known, and the rotation matrix R and translation t (external parameters of the camera) are obtained, the position of the point in the camera coordinate system (X, Y, Z) can be obtained by using formula (3).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = R \begin{bmatrix} U \\ V \\ W \end{bmatrix} + t = [R|t] \begin{bmatrix} U \\ V \\ W \\ 1 \end{bmatrix} \tag{3}$$

Changing camera coordinates to image coordinates is changing from 3D coordinates to 2D coordinates. fx and fy represent the focal length of the camera in the x and y directions, respectively, while cx and cy represent the optical center of the camera in the x and y directions. S represents the scaling factor. X and Y correspond to pixel coordinate systems. Calculate the position (x, y) of a point in the image coordinate system using formula (4).

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = s \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{4}$$

Therefore, the relationship between pixel coordinate system and world coordinate system is shown in formula (5):

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = s \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} [R|t] \begin{bmatrix} U \\ V \\ W \\ 1 \end{bmatrix} \tag{5}$$

According to the derivation results of the above formula, the computer predicts the driver's head posture. Real-time image acquisition of the user's head is carried out through the camera, and the two-dimensional position information of multiple parts of the user's head is recognized.

Then, by comparing the two-dimensional position information of the parts with the preset standard position information, the current state of the user's head can be determined, and the recognition of head posture can be achieved, and the detection diagram is shown in the figure 8.

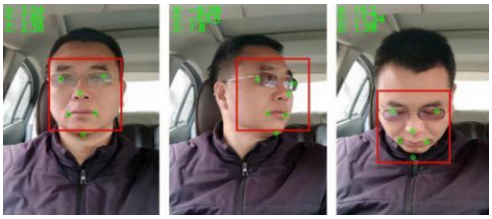

Figure 8. Detection diagram of head posture.

On the basis of detecting the driver's head posture angle, the pitch angle in the head posture angle is selected as the basis for judging whether the driver nods. To verify the effectiveness of pitch angle recognition for nodding actions, data from the same experimenter in normal driving and nodding states were taken, and the results are shown in the figure 9.

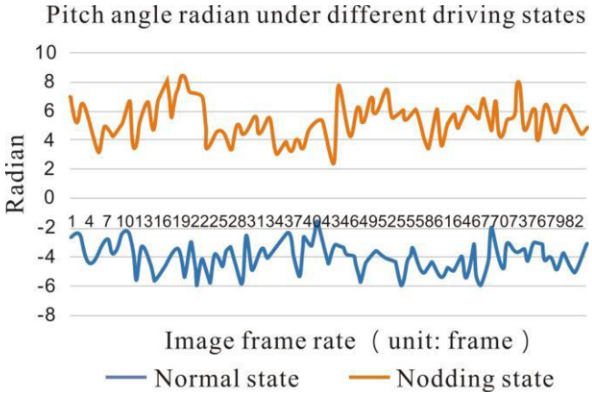

Figure 9. Pitch radian frame rate under different driving postures.

There is a clear boundary between the pitch angle differences of drivers in different states. Further analysis of variance on the experimental data revealed that $p<0.001$, indicating a significant difference in pitch angle between normal and nodding states. This parameter serves as the basis for identifying nodding actions.

When the monitoring system detects that the pitch angle of the driver's head posture exceeds the set range in more than 50 out of 80 consecutive driver images, it determines that the driver has nodded and detects drowsiness in real-time. Experimental results have shown that this method can effectively identify the fatigue status of drivers.

5. Conclusions

In our study, we utilized a modified version of MTCNN for real-time facial detection. Through this approach, we extract crucial head posture features, encompassing the pitch angle, yaw angle, and roll angle of the driver's head. By meticulously comparing the facial positions within the acquired driver's facial images, we discern and discriminate the driver's fatigue state based on the extracted feature values. This comprehensive methodology not only elevates the precision of fatigue recognition but also significantly enhances the detection speed of facial fatigue recognition. As a result, our study culminates in the development of a fatigue recognition model aligning with pertinent traffic regulations, thereby contributing substantively to the field of intelligent traffic safety systems.

Acknowledgment

This paper is sponsored by the fund for cooperative and collaborative education projects between industry and universities of the Higher Education Department of the Ministry of Education: "Research on the Practical Path of Industrial Design Teaching Based on the Management Model of Innovative Talents Cultivation in Information Technology" (Project No. 230729214007292)

References

[1] Li J B, Research on Fatigue Driving and Dangerous Driving Behavior Detection Algorithms Based on Deep Video, Qingdao University, 2020

[2] Zhang S, Research on Computer Vision Based Driving Behavior Recognition Algorithm, Tianjin University, 2019.11.

[3] Zheng D W, Facial Fatigue Expression Recognition Based on Machine Learning, Beijing University of Posts and Telecommunications, 2019.

[4] J S, M M, O P and S M, "Artificial Intelligence based Face Recognition using Deep Learning," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2022, pp. 1017-1024.

[5] V. C. R, V. Asha, B. Saju, S. N, T. R. Mrudhula Reddy and S. K. M, "Face Recognition and Identification Using Deep Learning," 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2023, pp. 1-5.

[6] Yin X, Liu X. Multi task convolutional neural network for pose invariant face recognition IEEE Transactions on Image Processing, 2017, 27 (2), pp. 964-975.

[7] Zhang K, Zhang Z, Li Z, et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks IEEE Signal Processing Letters, 2016, 23 (10), pp. 1499-1503.

[8] Shukri D Y S M, Asmuni H, Othman R M, et al. An improved multiscale regression algorithm for motion blurred iris images to minimize the intra individual variables Pattern Recognition Letters, 2013, 34 (9), pp. 1071-1077.

[9] Cai C. Research on facial recognition method based on facial reference point alignment. Huazhong University of Science and Technology, 2013.

[10] Godley S T, Triggs T J, Fields B N. Driving simulator validation for speed research Accidental Analysis and Prevention 2002, 34 (5), pp. 589-600.

[11] Filip P, Taillard J, Klein E, et al. Effect of fatigue on performance measured by a driving simulator in automotive drivers Journal of Psychological Research, 2003, 55 (3), pp. 197-200.

[12] Liu Y, Gong Y, Lu Z and Zhang X, "Accurate Head Pose Estimation Based on Multi-Stage Regression" 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 2022, pp. 1326-1330.

[13] Menan V, Gawesha A, Samarasinghe P and Kasthurirathna D, "DS-HPE: Deep Set for Head Pose Estimation" 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2023, pp. 1179-1184.

[14] Luo W J. Research on Non Safe Driving Behavior Detection Algorithm Based on Machine Vision. Hunan University, 2018.

[15] Kazemi V, Sullivan J. One Millisecond Face Alignment with an Ensemble of Regression Trees. IEEE Conference on Computer Vision&Pattern Recognition, 2014.