International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:10, No:1, 2016

Automatic Checkpoint System Using Face and Card Information

Kriddikorn Kaewwongsri, Nikom Suvonvorn

***Abstract*—**In the deep south of Thailand, checkpoints for people verification are necessary for the security management of risk zones, such as official buildings in the conflict area. In this paper, we propose an automatic checkpoint system that verifies persons using information from ID cards and facial features. The methods for a person's information abstraction and verification are introduced based on useful information such as ID number and name, extracted from official cards, and facial images from videos. The proposed system shows promising results and has a real impact on the local society.

***Keywords*—**Face comparison, card recognition, OCR, checkpoint system, authentication.

I. Introduction

IN the deep south of Thailand, security has been a big issue [1] over the last few decades and concerns many provinces and districts including Yala, Narathiwat, Pattani and some parts of Songkhla (Hatyai included). In these areas and at every entrance to public buildings, such as universities, police stations, hospitals, department stores, and etc., security staff will ask visitors or customers to show their personal ID cards, usually a national identity card or driving license card, to identify themselves. In general, a minimum of two CCTV cameras are used: one is focused on the driver's face, with another for zooming on the card, as shown in Fig. 1. This procedure is recognised by the local authorities as a scanning system, a way to record a person's information before entering into a secured area. The aim of this system is to identify threats as soon as possible in case any security issue is going to happen. However, the authorities are unable to take full advantage of the data using the current method. There are many weaknesses in the actual system. Firstly, it corresponds to the non-adaptive acquisition of materials. The CCTV camera installation is mostly not suitable for this situation. For instance, there is not enough resolution to record the detailed information on the cards. Further, the auto focus of the camera is not fast enough when the cards are suddenly posted in front of it, which causes colour saturation effects. Secondly, the recorded videos are never analysed and indexed, which makes it difficult or impossible to search for eventual threats. In the current situation, security staff must play back all 24 hours of many cameras' videos to search for desired information, which takes a long time and is very inefficient.

Kriddikorn Kaewwongsri is a Master's student and Nikom Suvonvorn a lecturer at the Department of Computer Engineering, Faculty of Engineering, Prince of Songkla University, Hatyai, Songkhla, Thailand 90112 (e-mail: k.kriddikorn@gmail.com, kom@coe.psu.ac.th).

Fig. 1 Checkpoint at an entrance site

In this paper, we propose a more efficient system to replace the practical method of the scanning system, called an automatic checkpoint system. In this system, the appropriate way of CCTV camera installation as a video acquisition sensor is always practical due to the large number of camera networks installed in the area over the past years. The system will recognise the information from the scanned cards, e.g. national ID number and name, from both types of cards, national identity cards and driving licenses. Afterwards, it matches this information with the owner's facial image, which is indexed in a database server. When the person exits from the area, the system will again extract their ID number, use it to retrieve the corresponding facial image, and verify it with the current facial image. The verification result will be shown to the security staff for further procedures. The benefits of this system is that all visitors and customers are totally indexed, which is useful for searching or alerting staff when there are suspect cases. The empowered system will be greatly increased if an indexed ID could be linked to a national information database centre so that a suspected person could be identified and tracked globally.

II. Related Works

Two related research works are concerned: card recognition and information extraction using optical character recognition (OCR) and facial comparison techniques. The OCR converts the characters of scanned images, whether handwritten, typewritten, or printed paper, into text. OCR is applied widely in many fields, especially for ID number recognition. Martin et al. [2] proposed a method to extract ID numbers on images obtained by scanning with a white background. The card is separated from the white background using histogram analysis, which is called peak detection. [3] This method

detects the two top peaks of a histogram and rotates a card using momentum of order. After the card is rotated, information can be extracted using a certain format. Morphological top-hat operator is used to improve text, and the ID number is recognised using an OCR engine. Qu et al. [4] recognised ID numbers of three card types using a template matching method based on similarity voting. The different patterns of numbers' structures are used as features. They used six templates for each number and each template is compared to a query number, and then voted as the possible number. The highest probability would be chosen as the answer, and the ID number is validated by calculating the value of the check digit.

Concerning the related research on face comparison techniques, [5] used Active Shape Model (ASM) to extract the important facial feature set. Then, these features were rectified using facial geometric invariance: the distance between the inner corners of both eyes should be horizontal, and the distance between the chin and nose should be vertical. The age-texture of the image was analysed using Log-Gabor wavelet. The high frequency data in the test image was replaced by high frequency data in the target image, according to suitable sections in the composition map. As a result, the proposed method yielded 100% accuracy for age synthesis on human faces. Juefei-Xu et al. [6] recognised faces using the framework of the periocular region. Walsh-Hadamard transform encoded local binary patterns (WLBP) to be used as features. The WLBP was extracted on the periocular region that maintains the consistency of the same subject across ages. Unsupervised discriminant projection (UDP) was used to build subspace on WLBP featured periocular images. The results showed accuracies of 100% rank-1 identification rate and 98% verification rate at 0.1 % false accept rate on the entire FG-NET database. Mahalingham et al. [7] proposed an age-invariant face recognition algorithm using a graph-based face representation containing the appearance and geometry of facial feature points. They learned an age model for each subject and built graph space for each image. There were two stages for recognition. First, a Maximum a Posteriori solution based on PCA factorisation was used to optimise the search space and choose candidate model sets. Second, a simple deterministic algorithm was used for graph matching between the probe image and the gallery image. The results showed that the proposed method provided good performance in age invariant facial recognition.

III. Automatic Checkpoint System

Our automatic checkpoint system is divided into two parts: Person Information Abstraction (PIA) and Person Information Verification (PIV). The process of PIA consists of extracting the useful information for a person at the entrance and exit of an observing zone. Two cameras are used: one for detecting cards and extracting information, including the ID number and name of a cardholder, and another camera for facial detection. All extracted information is indexed in the database using the ID number. The process of PIV is used for verifying the person who exits from the observing zone. Again, two cameras are used for detecting and extracting the ID number and name from the card and the facial image of the person. The extracted ID number is used to query the corresponding facial image from the database. Then, verification is performed via facial comparison. Fig. 2 shows the overall framework of an automatic checkpoint system.

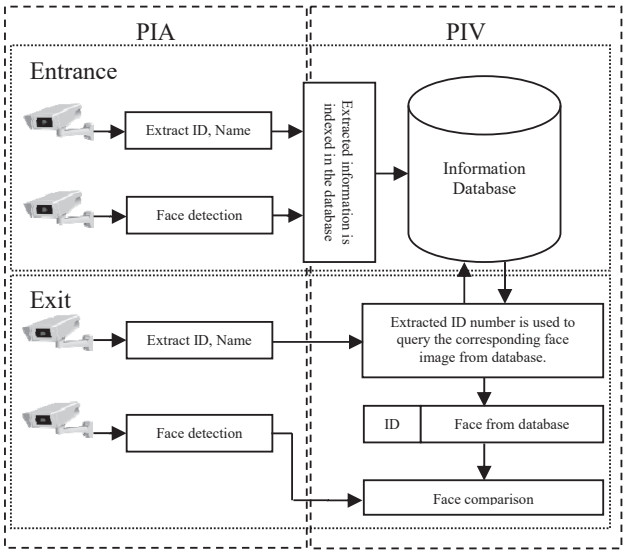


Fig. 2 Framework of automatic checkpoint system

A. Person Information Abstraction (PIA)

Person Information Abstraction is a process for obtaining the information of a person, such as the ID number and name from the card, and the facial image. For the facial image, the Harr-like method [8] is applied simply for detecting the front view of the face, which is then indexed using the ID number extracted from the ID card.

Concerning the extraction of the ID number and name from a card, four necessary steps are required: key frame extraction, card detection, type classification, and OCR. In the first step, the key frame extraction selects the best card image from a sequence of image frames. The assumption is that the high frequency information will increase compared to its background when the card appears in the video. We use edge detection to filter the image pixels with high frequency. Subsequently, the numbers of these pixels are counted. Due to the possible variation caused by light changing, the learning background technique and running average with selectivity is applied in order to define the normal variation values of these pixels, which is set as a background (μ,σ), as shown in (1).

$$\mu_{i+1} = \alpha f_i + (1-\alpha)\mu_i$$
$$\sigma^2_{i+1} = \alpha (f_i - \mu_i)^2 + (1-\alpha)\sigma_i^2 \quad (1)$$

The frame containing the card image is the one with high intensity variation being greater than a threshold, defined to T_i as described in (2).

$$|f_i - \mu_i| > T_i \text{ where } T_i = K\sigma_i \quad (2)$$

With this process, a card may appear in many frames, as shown in Fig. 3 (blue colour). Then, the key frame will be selected from the middle of these blue frames, also shown in Fig. 3 (yellow colour).

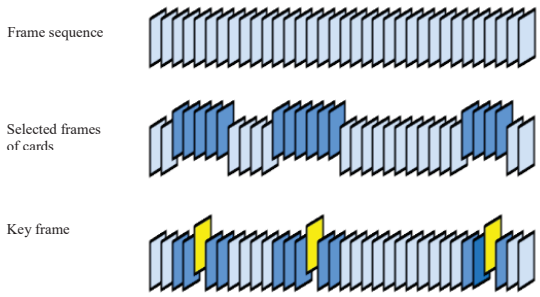


Fig. 3 Key frame selection

In the second step for card detection, contours are extracted from the key frame. The polygon approximation is then applied to the contours in order to fit with the rectangular features. The validated rectangle is then defined as the ROI of card, which will be normalised to a specific size using the affine transform and image warping technique. This step will deal with the variation of sizes and angles of ROIs of cards detected in the key frame, as shown in Fig. 4.

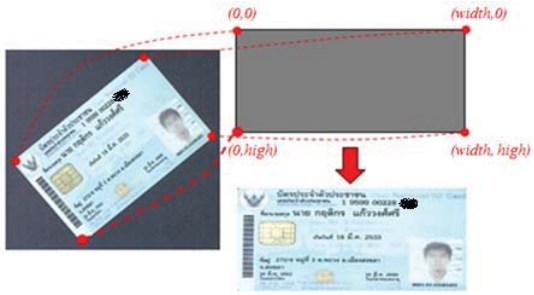


Fig. 4 Normalisation of card's ROI

The third step is for the classification of the card's type. In this stage, two types of cards are concerned: national identification card (ID card) and driving license card. It is necessary to identify the type of card for better character recognition following in the further steps. Equations (3) and (4) are defined for this purpose:

$$\sigma_{b,g,r} = \sqrt{\frac{n\sum I_i^2 - \left(\sum I_i\right)^2}{n(n-1)}} \tag{3}$$

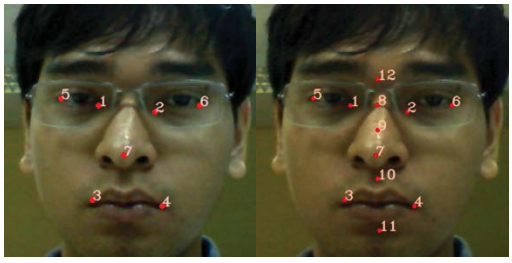
$$\sigma = \sqrt{\sigma_b^2 + \sigma_g^2 + \sigma_r^2} \tag{4}$$

We found that, if the sigma value is lower than the optimal threshold, the card is identified as an ID card. Otherwise, it is a driving license card. When the card type is identified, the pre-defined card patterns of ROIs for the specific card are determined. Segmentation of the image regions for ID number and name could be done using corresponding patterns. Finally, the ROI images of the ID number and name will be binarised by the Otsu method [9] and transformed into text using Tesseract OCR.

B. Person Information Verification (PIV)

The Person Information Verification is a process for verifying a person exiting from an observing zone and comparing the data with his/her own information when previously entering the zone. Initially, the ID number and name from the card and facial image of the person could be extracted using the same method described previously in the PIA. Then, the extracted ID number is used to retrieve the corresponding facial image from the database for facial comparison. The verification is validated only if the comparison score is acceptable.

Our face comparison is based on a comparison of components of the face, such as the eyes, nose, and mouth. Such components must be segmented, which is described below. Firstly, the Flandmark technique [10] is used to detect landmarks on the face, including the left and right eye corners, mouth corners, and nose tip, as shown in Fig. 5 (a) (i.e., 1st-7th points). However, these landmarks are insufficient for approximating the position of the face components. Thus, we define additional reference points for segmenting the face components. The 1st - 7th points are used to estimate the 8th-12th points, as shown in Fig. 5 (b) using (5)-(9). Then, the x and y points obtained from 1st-12th points are used to approximate the boundary of each component, as shown in Fig. 6. Segmentation can be performed using these boundaries. The sizes of components are normalised. Colours are split into the R, G, and B channels, which are normalised by histogram equalisation method and re-combined into a single image, which is finally converted into a gray-scale image.



(a) (b)

Fig. 5 Landmark positions of facial components

$$P8 = \left(\frac{P1_x + P2_x + P7_x}{3}, \frac{P5_y + P1_y + P2_y P6_y}{4}\right) \tag{5}$$

$$P9 = \left(P8_x, \left(\frac{\left(\frac{P5_y + P1_y + P2_y + P6_y}{4}\right) + P7_y}{2}\right)\right) \tag{6}$$

$$P10 = \left(P8_x, \left(\frac{\frac{P3_y + P4_y}{2} + P7_y}{2} \right) \right) \quad (7)$$

$$P11 = \left(\frac{P3_x + P4_x}{2}, \frac{P3_y + P4_y}{2} + 1.5\left(\frac{P3_y + P4_y}{2} - P10_y \right) \right) \quad (8)$$

$$P12 = \left(P8_x, \frac{P5_y + P1_y + P2_y + P6_y}{4} - (P9_y = P8_y) \right) \quad (9)$$

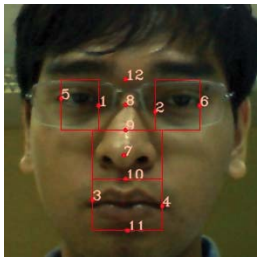


Fig. 6 Facial components

We compare each pair of facial components between two facial images. The square root of the sum of squared error (L2) is calculated to indicate the difference between each facial component pair. Four scores of L2 are obtained from four pairs of facial components. Euclidean distance of the four scores is computed using (11) that indicates the similarity between two facial images. Low Euclidean distances mean that two facial images have high similarity. In contrast, high Euclidean distances mean that two face images have low similarity.

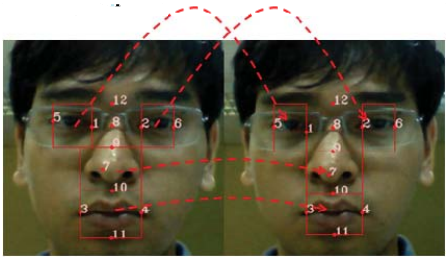


Fig. 7 Calculation between components

$$score_{l\_eye,r\_eye,nose,mouth} = \frac{\sqrt{\sum_i (I1_i - I2_i)^2}}{I1_{wide} \times I1_{hight}} \quad (10)$$

$$score_{total} = \sqrt{l\_eye^2 + r\_eye^2 + nose^2 + mouth^2} \quad (11)$$

IV. Experimental Results

The system is experimented on using an Intel P8700@2.5Ghz with 2GB memory. The codes are implemented using Microsoft visual studio C++ 2012 with OpenCV library. The Flandmark library is used for facial landmark initialisation and Tesseract is used as the OCR engine.

A. PIA Results

We used three driving license cards and four identity cards as a dataset. All cards were placed into three different patterns including straight to camera, rotate left, and rotate right. Each pattern was repeated five times. We obtained 105 image patterns in total for 45 driving card images (three driving cards × three patterns × five times) and 60 identity card images (four identity cards × three patterns × five times). Table I shows the accuracies of information extraction from the cards. We found that the system could select the correct keyframe with 100% accuracy. The accuracies for rotation of the driving and identity cards are 97.77% and 100%, respectively. In some cases, when the card has a transparent plastic cover, the system fails to determine the card's corners correctly. For this reason, error occurs at around 2.23%. We noticed that the system could classify correctly the types of the driving and identity cards at 100% and 96.66%. The accuracies of OCR for the driving and identity cards are 68.18% and 79.31%, respectively. The average time consumed per card was 1686ms.

TABLE I
PIA Accuracies

Card type	Keyframe Selection	Rotated card Correction	Card type Identification	OCR
Driving Card	100%	97.77%	100%	68.18%
Identity Card	100%	100%	96.66%	79.31%

B. PIV Results

In order to experiment with our PIV method, we collected the dataset from four subjects: 20 facial images collected from each subject. Therefore, a total of 80 facial images were used. The dataset was divided into two groups: 40 facial images used as a testing set and the other 40 images used as the request set. To evaluate the capability of facial similarity measurement of PIV method, every facial image for each subject from the request set was compared to all subjects in testing set.

Fig. 8 shows the comparison scores (11) between query facial images and testing images while querying the first subject. We noticed that the scores of the first subject compared to itself (blue line) were the lowest at 1.28 in average. The comparisons with other subjects gave higher values of 2.26, 2.54, and 3.5, correspondingly. Fig. 9 shows the result of the second subject. The self-comparison was around 1.21 and its cross comparisons were 2.26, 2.54, and 3.5, respectively. Again, Figs. 10, 11 show the result of the third and fourth subjects. In these cases, we observed the same trend of self-comparisons giving the lowest values at 1.42 and 1.71. The other comparisons were greater than 2.43. From these experimentations, we determined that the optimal threshold for acceptability of the facial verification using our PIV was 1.99, which was the average of the minimum value of cross-comparison scores (2.26) and the maximum value of the

self-comparison score (1.71). This indicates that if the comparison score is less than 1.99, our system will consider that the person entering and exiting the observed zone using the same ID number is the same person.

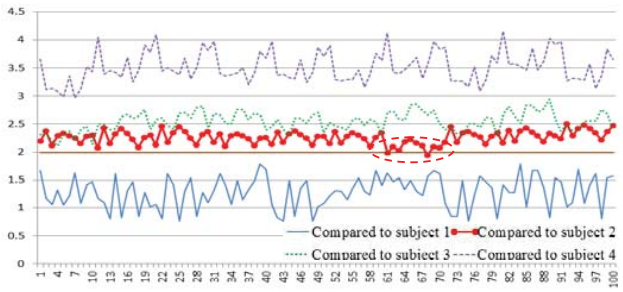


Fig. 8 First subject compared to others

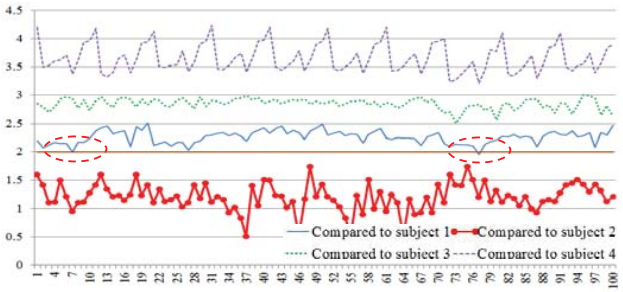


Fig. 9 Second subject compared to others

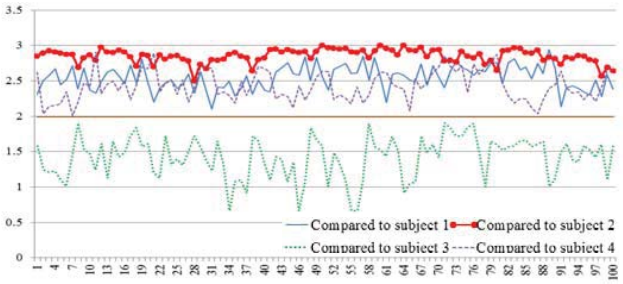


Fig. 10 Third subject compared to others

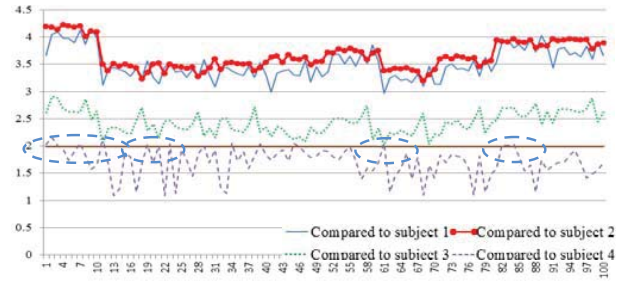


Fig. 11 Fourth subject compared to others

As a result, if the query set images and testset image from the same subject are compared, the comparison score should be lower than other comparisons. In other words, the lower comparison score means that two facial images are similar and can be validated. From the empirical results, the verification of a person will be validated by our PIV method when its score value is less than 1.99, which is the optimal threshold for these datasets.

TABLE II
COMPARISON ACCURACIES

Comparison	1st subject	2nd subject	3rd subject	4th subject
self-compared *(True Positive)	100%	100%	100%	84%
cross- compared *(True Negative)	99.32%	99.32%	100%	100%
correctly	99.66%	99.66%	100%	92%
overall accuracy	97.83%			

Table II shows the accuracies of the verification process while using the optimal threshold of 1.99. We found that the accuracies of the first and second subjects were 99.66% in average. The third and fourth subjects were 100% and 92%, respectively. The red and blue circles in Figs. 8-11 show the false positives and false negative. The overall system can correctly verify persons with up to 97.83% accuracy.

C. System Integration

Fig. 12 shows the integrated automatic checkpoint system. Number one displays the video captured from the first camera, which is focused on the card. Our PIA is executed to extract the key frame. Then, the ID number will appear in the textbox, as shown in area number two. This ID number is then used to retrieve the image of a face from the database server and display it in area number three. Area number four displays the video captured from the second camera, focusing on the face. The PIA for facial detection is applied and shown in area number five. Then, facial comparison is done by our PIV method. The verification result is shown as a text overlay on the facial image in area number three. Note that the green text is for acceptable facial comparison and the red text shows contrast.

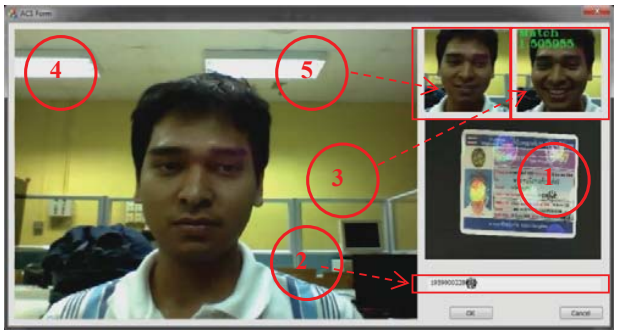


Fig. 12 System integration of Automatic Checkpoint System

V. CONCLUSION

In this paper, an automatic checkpoint system for people is proposed. The PIA method for extracting information from ID cards is introduced, which is robust against the variation of brightness scaling and rotation. The overall results of PIA are rather accurate. However, some enhancements for overshadowing of a watermark on the card are necessary in order to increase the OCR recognition rates. The PIV method

is presented for person verification using facial similarity measurement via its components. The proposed method does not require any training process. The performance of PIV is satisfactory due to the "one-to-one" matching design with additional indexed ID number. In future work, more datasets should be investigated in order to study the robustness and automatic thresholding of the proposed method.

REFERENCES

[1] "Thailand: The Evolving Conflict in the South," Crisis Group Asia Report N°241, 11 December 2012.
[2] F. Martin and V. Maria, "Automatic reading of Spanish identity cards," Document Analysis and Recognition, pp. 470-474, 2001, Seattle, 10 - 13 Sep 2001.
[3] F. Martin, "Analysis Tools for Gray Level Histograms," In Proc. Signal Processing Pattern Recognition and Applications, pp.11-16, 2003.
[4] Lu Y. Qu, Y .Chengand, Y. Xie, "ID Numbers Recognition by Local Similarity Voting," International Journal of Advanced Computer Science and Applications, pp. 3881 – 3888, 4 Oct. 2010.
[5] Ching-Tang Hsieh, Chia-Shing Hu and Chun-Wei Pan, "A Simple Automatic Facial Aging/Rejuvenating Synthesis Method", Systems, Man, and Cybernetics, pp.2982-2988, 9-12 Oct. 2011.
[6] F. Juefei-Xu, K. Luu, M. Savvides, T. D. Bui, and Ch. Y. Suen. "Investigating Age Invariant Face Recognition Based on Periocular Biometrics." Biometrics (IJCB), 2011 International Joint Conference on Biometrics, pp.1 – 7, Washington, 11-13 Oct. 2011.
[7] Gayathri Mahalingam and Chandra Kambhamettu, "Age Invariant Face Recognition Using Graph Matching", Biometrics: Theory Applications and Systems, pp. 1-7, 27-29 Sep 2010.
[8] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features," IEEE CVPR01, pp. 511-518, 2001.
[9] Nobuyuki Otsu. "A threshold selection method from gray-level histograms," Systems, Man, and Cybernetics, pp. 62–66,1979.
[10] M. Uricar, V. Franc and V. Hlavac, "Detector of Facial Landmarks Learned by the Structured Output SVM," In Proc. 7th International