Analyzing the Effect of Eye Center Localization on Accurate Landmark Localization in a Facial Image

Manir Ahmed
Department of Electronics and Communication Engineering,
National Institute of Technology Silchar,
Assam, India
manirahmed@ieee.org

Ram Kumar Karsh
Department of Electronics and Communication Engineering,
National Institute of Technology Silchar,
Assam, India
tnramkarsh@gmail.com

Rabul Hussain Laskar
Department of Electronics and Communication Engineering,
National Institute of Technology Silchar,
Assam, India
rhlaskar@ece.nits.ac.in

***Abstract*— Localization of facial landmarks on a human face is an important step for many face-related computer vision applications. The most of the earlier techniques (AAMs, CLMs) has achieved good performance in landmark localization but they always limited by the initialization of landmarks. In this paper, the initialization problem is solved by taking the eye center as references to the mean face shape. In the proposed method, the eye centers are estimated using multi-scale iris shape feature first and then the constrained local model is applied for landmark localization where initialization is done using mean face shape taking eye centers as references. The performance of eye center estimation and landmark localization method are evaluated on AR and Multi-PIE databases. For eye center estimation three normalized eye localization error is considered whereas for landmark localization RMSE and detection rate are considered. For landmark localization, a total of 130 and 68 landmarks are considered for AR and Multi-PIE database respectively. The experimental results suggest that the proposed method has achieved improved performance as compared to some of the other methods.**

Keywords— facial landmark localization, eye localization, face alignment, constrained local model, support vector machine

I. INTRODUCTION

The landmark localization is an important step for many face-related applications in computer vision community. The landmark points on a facial image convey important information of cognitive stage of a person. Thus, this can be used in facial expression recognition [1], face alignment and recognition [2], face hallucination [3] etc. The landmarks are some important points on a face such as corner of mouth, corner and center of eye, nose tip etc. The automatic landmark localization is considered as a difficult problem due to the varied facial appearance and presence of external noise. The appearance of a face varied due to the structural changes between persons, different expressions, head poses, colors etc., however the external noises includes hair, spectacles, illumination variations, glisten of eyeglasses etc. Several approaches are proposed by the researchers to solve the above difficulties in the last couple of decades. The Active Appearance Model (AAM) [4] and its variants [5] achieve great success in the field of landmark localization. These methods are holistic in nature which required complete face appearance. The main limitation of these approaches is the initialization of the landmarks. The failure cases of these methods are also seen in the large deformation facial images and in varying lighting conditions. On the other hand, the part-based models shows higher accuracy as compare to the holistic approach. Some examples of part-based models are component based active shape model [6], tree-structured model [7] etc. The component-based active shape model (CompASM) [6] consider landmark as independent parts of the facial shape model. This method can localize the landmarks under different facial expression. Similar to the AAMs, CompASM is suffered from the initialization problems.

The tree structure model [7] uses mixture-of-tree structure models to perform three tasks simultaneously i.e. face detection, landmark localization and pose estimation. In [8], an extended tree-structured model is proposed to detect more landmarks compare to the original tree-structured model [xx] in frontal faces. The extended tree-structured model shows high accuracy for landmarks localization. These models do not require the initialization of landmarks but show high processing time as well as less accuracy.

In recent days, the constrained local model (CLM) and its variants [9-12] show better performance as compared to the AAMs. This model builds patch experts for each landmark individually which makes it robust to large appearance deformation and occlusions. Though the CLMs show better performance, but these are also suffering from initialization problems.

In this paper, we propose a better way to represent the CLM model named as eye center guided constraints local model (ECG-CLM). Here, the major limitation of original CLM approaches i.e. the initialization of landmarks is handled taking eye centers as references to the mean face shape. This paper shows the landmark localization performance for different normalized eye center localization errors to justify the impact of precise eye center localization on other facial landmark localization.

The structure of the paper is as follows. In Section II, the detail of the proposed method is described. The results and discussion is shown in section III. Section IV shows the conclusion and futures works.

II. EYE CENTER GUIDED CONSTRAINED LOCAL MODEL

In this section, the details of the proposed eye center guided constrained local model is described. The block

diagram of the proposed ECG-CLM model has been shown in Fig. 1. First of all, the face region is detected from the input image using the face detector. Then the eye center is estimated using the multi-scale iris shape feature reported in [13]. The initialization problem of the constrained local model is solved by taking the eye centers as references to the mean face shape. The details of each block are described in the following section.

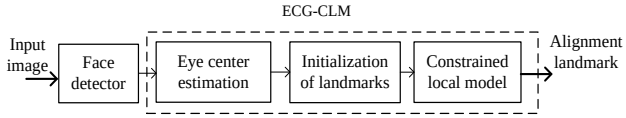
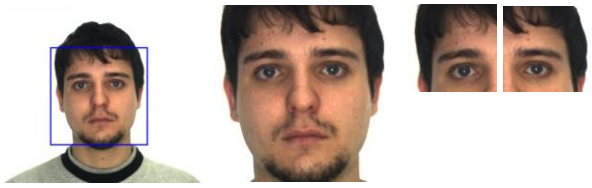


Fig. 1. Block diagram of the proposed landmark localization method

A. *Face detector*

The Viola-Jones face detector [14] is used for detecting the face region from the input image. Instead of using one face detector as earlier works [9] used, this paper uses 3 face detectors to extract the face region. One is a frontal face detector and the other two are profile face detectors. The frontal face detector is used first to detect the face region and if it fails then the other two face detectors are used. This is done to improve the face detector performance in this paper. For eye center estimation, we only use the upper half portion of the face region as shown in Fig. 2. For the left and right eye detection, the upper half portion again divided vertically into two portions as shown in Fig 2(c).



(a) (b) (c)

Fig. 2. (a) Input image, (b) Face detector output, and (c) Upper half portion divided vertically into two regions

B. *Eye center estimation*

The eye centers are estimated using the multi-scale iris shape feature on the face region detected by the face detector. The multi-scale iris shape feature first reported in [13] where it is used to detect eye candidates under head pose variations. The multi-scale iris shape features are the various scale version of the original iris shape feature [15]. The combined response of different scale versions of the iris shape feature is considered as the response of a multi-scale iris shape feature. In this paper, 3 different scale version of iris shape feature is used as multi-scale iris shape feature as shown in Fig. 3.

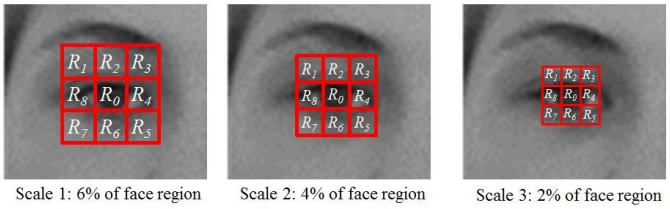


Fig. 3. Size of iris shape feature for scale 6%, 4% and 2% of face region centered at a pixel of the eye region.

The iris shape feature is constructed with 9 square cells where the middle cell is called as iris cell and other cells are called as surrounding cells. Each surrounding cells' size is same as the iris cell. The iris cell size depends on the size of the face region detected by the face detector. In this paper, three different iris cell sizes (6%, 4%, and 2% of face region size) are considered. The multi-scale iris shape features are named depending on the iris cell size as shown in Fig. 3. The scale of the iris shape feature is depended on the face boundary detected by the face detector. According to [13], the iris size is less than 0.07 times of the face boundary size.

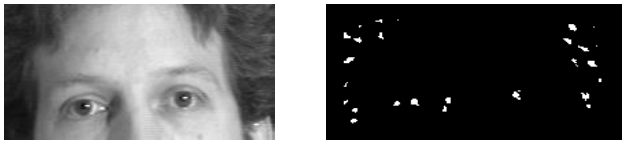
The multi-scale iris shape feature is applied to detect the eye candidates from the face region. The multi-scale iris shape feature is applied to the pixel of an image should satisfy the two conditions: (i) the mean intensity of surrounding cells of the multi-scale iris shape feature should greater than the mean intensity of iris cell, and (ii) the mean intensity of the iris cell should less than the face region detected by the face detector. Upon satisfying the above two criteria for a pixel, the following calculations are done for three scales i.e. 2%, 4% and 6%.

$$\bar{R}_0 = \frac{\sum_{x,y \in R_i} I(x,y)}{n} \quad (1)$$

$$Score_{scale} = \left(\bar{R}_0 - \bar{R}_{Fcae}\right) \times \sum_{i=1}^{8} \left(\bar{R}_i - \bar{R}_0\right), \quad scale = 1,2,3 \quad (2)$$

$$S = \frac{1}{3} \sum_{scale=1}^{3} Score_{scale} \quad (3)$$

where, n is the number of pixels found in the R_i region, $\bar{R}_0$ is the mean intensity of the R_0 cell, $\bar{R}_{Face}$ is the mean intensity of the face region detected by the face detector and $\boldsymbol{S}$ is the response score of the multi-scale iris shape feature. An example of a multi-scale iris shape feature response score applied to an image as shown in Fig. 4. It can be observed that the multi-scale iris shape feature is generated some group of pixels. Each group is considered to be the eye candidates and a maximum score of the group is considered to be the center of the eye candidate. Since the detected eye candidates have some false eye candidate too, so a verification process is applied to detect the true eye candidates.



(a) (b)

Fig. 4. (a) Input image, (b) Output response of Multi-scale iris shape feature

In the verification stage, the eye images are cropped with a size of 30×40 pixels taking eye center as an anchor point. Then each eye images are verified with a SVM based classifier trained with Histogram of oriented gradients (HOG), Local Binary Pattern (LBP) and Cell Mean Intensity (CMI) Features. The highest classifier scores found on two half of upper part of face region considered as required eye pair.

C. Constrained local model

A constrained local model is a part-based method to localize the facial landmarks. This model uses some patch/region experts to register the non-rigid face region. Combine responses of these patch experts are jointly optimized to estimate the global non-rigid variation of the face region. The CLM procedure mainly consists of two parts: (i) CLM model building, and (ii) CLM search process.

1. CLM model building

The CLM model building is a training phase. In this phase, two models are prepared mainly: (i) shape model and (ii) patch model. The shape model is prepared to learn the shape variations of the facial images. The point distribution model (PDM) [4] with principal component analysis (PCA) is used to learn the shape model. The non-rigid shape variation can be expressed as a linear combination of mean shape and Eigen vectors. It can drive as equation (4).

$$x = \bar{x} + VP \tag{4}$$

where, $\bar{x}$ is the mean shape, V be the Eigen vectors and P is the non-rigid wrap. The Procrustes analysis is done in the training images before applying the PCA techniques. This is required to remove the scale, rotation and translation variation of training images. The Procrustes analysis also registers the shape of the training images as the inter-ocular distance is 75 pixels. The shape model only learns the shape variations of the training images.

The patch models are prepared to learn the appearance variation across each landmark points in the training images. For this purpose, a linear support vector machine (SVM) [16] based classifier is used. The positive and negative patch samples are cropped from the training images after the shapes are registered. The positive patches are cropped from the annotated landmark points of the training images and the negative samples are cropped by shifted original landmark points. The number of the patch model is the same as the number of the landmark used to represent the shape of a training image.

2. CLM search process

The CLM search process is the testing phase where the built CLM shape model and patch models are used to locate the facial landmarks in the test image. The CLM search process can be divided into the following steps.

1. Initialization of the facial landmark points.
2. Generate local region from the landmarks.
3. Patch experts are applied to generate the response images.
4. Fit the response images using optimization.
5. Update the landmark positions.
6. The steps 2-5 are repeated until all the landmarks reach its best positions.

Among all the steps, we mainly focused on Step 1 i.e. initialization of the facial landmark points. Earlier works initialize the landmarks points using mean face shape and the boundary box returned by the face detector [17]. Such type of initialization is heavily depended on precise localization of the face boundary box. If there is slight variation in scale or the orientation displacement in comparison to ground truth, there have high chances of falling in local minima. Instead of depending on the boundary box of the face detector, this paper takes the position of eye centers as references to the mean face shape which shows proper initialization.

In this paper, the inter-ocular distance of the mean face shape keeps as a constant distance i.e. 75 pixels. Thus, the eye center of the test image has to be aligned with the eye position of mean face shape. Here, the eye center of the test images is detected by multi-scale iris shape feature as described in section II (B) and then align them with mean face shape using Procrustes technique. Such type of initialization has the following advantage.

- It is able to align landmarks in both seen/unseen appearance and identity variations.
- The scale, translation, and rotation of test images are adjusted to a reference face shape at the initial stage.

After the initialization step, the local region of 32×32-pixel size has been cropped taking current landmark position as center. Then the patch experts are applied to the cropped region to find the response images. An optimization technique and the learned shape constraint are now applied to fit the CLM model. In this paper, the quadratic programming method [18] is adopted as the optimization technique. The optimization technique predicts the best position of the landmarks. If the predict landmarks do not meet the desire one, Step 2-5 are repeated until it converged.

III. EXPERIMENTS

A. Experimental setup

For the experiment, two databases have been adopted namely CMU Multi-PIE database [19] and AR database [20]. The AR database consists of 126 individual; out of which 112 individual (58 men and 54 women) have the ground truth landmark points. This database contains different facial expressions, occlusions and illuminations conditions. For the experiment, only the frontal images which have ground truth landmark points are considered. A total of 896 numbers of images are considered for the experiment. The images wearing sunglasses are excluded from the evaluation setup. The Multi-PIE database consists of 337 individuals bearing different head poses, expressions, and illumination variations. For the experiment, only the frontal images (2526 images) are taken into considerations which have ground truth landmark points. The 130 ground truth landmark points are used to represent the face shape in the AR database, whereas the Multi-PIE database uses 68 ground truth points. For the training phase, we have taken 50% images from both the databases.

In order to calculate the precise eye center estimation, normalized eye localization error (N_{err}) [21] is adopted. Mathematically, it can be derived as

$$N_{err} = \frac{\max\left(d_{left}, d_{right}\right)}{d_{IOD}} \tag{5}$$

where, d_{left} and d_{right} are the distance between the ground truth and the detected position of the left and right eye, respectively. The d_{IOD} is the distance between the ground truth of the left and right eye.

The performance of the landmark localization method is measurement on the basis of Root Mean Square Error (RMSE) [8] and it can be derived as equation (6).

$$RMSE = \frac{\sum_{i=1}^{L}\sum_{j=1}^{N}\sqrt{(\mathrm{x}_{i,j} - \hat{x}_{i,j})^2 + (\mathrm{y}_{i,j} - \hat{y}_{i,j})^2}}{L \times N \times IOD_i} \tag{6}$$

where N is the number of landmarks, L is the total number of test images in a database and IOD_i is the inter-ocular distance of ith image. $x_{i,j}$ and $\hat{x}_{i,j}$ are the detected and ground truth point of jth landmark on ith image. The RMSE value is normalized by dividing it with interocular distance (IOD) to ensure the result is independent of the scale variation of faces [8].

Moreover, the detection rate is calculated to measure the precision of landmark localization. For the experiment, three threshold value 5%, 10% and 20% of the IOD distance is taken into consideration.

B. Results and discussion

In this section, the experimental results and analysis of the proposed method are discussed. The evaluations of different stages of the proposed method are conducted in this section. At first, the face detector experiment is performed. The second experiment is done to show the precision estimation of the eye centers. The third experiment is conducted to evaluate the landmark localization performance of the proposed method. The fourth experiment is done for evaluating the effect of precise eye center localization on landmark localization. The result of the proposed method has also been compared with other recent works [6, 12]. All the experiments are conducted on the images which are unseen at the training images (i.e. rest 50% of the databases). Both train and test experiments are performed on PC with processor Intel(R) Core(TM) i5-4590, 3.30 GHz, 16GB RAM, and Windows 7 based MATLAB (2017b) platform.

1. Comparison performance of the different face detectors

In this experiment, the performance of the face detector is conducted on both AR and Multi-PIE databases. The experiment is done to compare the performance between using single VJ face detector and using three VJ face detector implemented in OpenCV. It is observed that using three VJ face detector from OpenCV shows 100% accuracy. Single VJ face detector fails to detect the face region in some images where slight pose variation occurs. Thus, in this paper, the three VJ face detector is used for face detection.

TABLE I. FACE DETECTOR PERFORMANCE BETWEEN SINGLE AND MULTIPLE FACE DETECTORS

Database	*Methods*	*Total image*	*Detected image*	*Detection rate (%)*
AR Database	Single VJ face detector	896	886	98.88%
	Three VJ face detectors from OpenCV	896	896	100%
Multi-PIE database	Single VJ face detector	2526	2520	99.76%
	Three VJ face detectors from OpenCV	2526	2526	100%

2. Evaluation of eye center estimation performance

In this section, the performance of eye center estimation is performed on both the databases. The experiment is performed for three different normalized eye localization errors. These three normalized error signifies of localizing of the pupil ($N_{err} < 0.05$), iris ($N_{err} < 0.10$) and eye ($N_{err} < 0.25$). It is observed from the experiment that the eye center estimation method shows good accuracy for eye localization. However, for pupil and iris localization, the proposed method shows moderate performance because it fails to localize precisely in the images where the iris is not visible properly.

TABLE II. EYE LOCALIZATION PERFORMANCE FOR DIFFERENT NORMALIZED ERRORS

Database	$N_{err} < 0.05$	$N_{err} < 0.10$	$N_{err} < 0.25$
	Accuracy (%)	*Accuracy (%)*	*Accuracy (%)*
AR database	83.53	93.96	96.48
Multi-PIE database	86.87	95.16	97.24

3. Evaluation of landmark localization performance

The experiment of the proposed landmark localization method is performed in this section. The experiment is done for evaluating the RMSE value and for three detection rates on both the databases. It is observed that both AR and Multi-PIE models are able to localize the landmark points with low RMSE value and high detection rate.

TABLE III. LANDMARK LOCALIZATION PERFORMANCE FOR RMSE AND DETECTION RATES

Database	*Models*	*Threshold for landmark localization*			
		RMSE	*5% IOD*	*10% IOD*	*20% IOD*
AR database	AR model	.0612	77.86	90.89	96.32
Multi-PIE database	Multi-PIE model	.0486	87.56	94.86	98.30

4. The effect of precise eye center localization on landmark localization

In this section, the effect of precise eye center estimation on the landmark localization is calculated. The analysis is performed for three different normalized eye localization errors as listed in Table IV. It is observed the landmark localization shows less RMSE and high detection rate for lower normalized eye localization error and it gradually deteriorate with an increase in normalized eye localization error. Thus, it can be concluded from the analysis that the precise eye center localization plays a major role in proper landmark localization.

TABLE IV: EFFECT OF PRECISE EYE LOCALIZATION ON LANDMARK LOCALIZATION PERFORMANCE

Models	*Normalized eye localization error*	*Accuracy for landmark localization*			
		RMSE	*5% IOD*	*10% IOD*	*20% IOD*
AR model	$N_{err} < 0.05$	.0403	82.86	94.89	98.32
	$N_{err} < 0.10$	.0448	81.86	92.15	98.02
	$N_{err} < 0.25$	.0612	77.86	90.89	96.32
Multi-PIE model	$N_{err} < 0.05$	.0286	89.42	96.43	99.29
	$N_{err} < 0.10$	.0312	88.76	95.15	98.86
	$N_{err} < 0.25$	.0486	87.56	94.86	98.30

5. Comparison performance analysis

In this section, the comparison of the proposed landmark localization performance with other methods is conducted on both the databases. For the experiment, seventeen numbers of landmark points have been considered to maintain the uniformity between the databases as well as for earlier works [6, 12]. These points are basically the mouth corners (4 points), the eyebrow corner (4 points), the eye centers and corners (6 points) and the nose tip and sides (3 points) [8]. These points show stable and reliable results for face recognition and tacking with various facial expressions. The comparison is performed in terms of RMSE value with the previous works [6, 12] as shown in Fig 5. It is observed that the RMSE value of the proposed method is less as compared to the other methods on both the database.

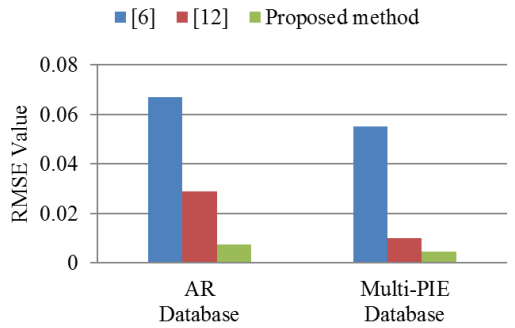


Fig. 5. Comparison results with other methods in terms of RMSE value

Fig. 6 shows the qualitative results of the proposed method on both the databases. It can be observed that our method is successfully able to localize landmark in varied facial appearance. However, its failure occurs when the eyes are not visible properly or face in large deformation.

Fig. 6. Sample of success (first and second row) and failures (last row) on AR database (first column) and Multi-PIE database (second column)

IV. CONCLUSION

In this paper, an eye center guided constrained local model is proposed to localize the facial landmark points. Unlike the AAMs and CLMs, the proposed CLM approach handle the initialization problem taking eye centers as references to the mean face shape. Thus, the method initially finds the eye centers using the multi-scale iris shape feature and initialize the other facial points taking eye center as a reference to the mean face shape. After initialization, the other steps of CLM approach are applied for landmark localization. The performance of eye center estimation and landmark localization are evaluated on AR and Multi-PIE databases. It is observed that the eye center estimation method shows good accuracy for eye localization. Also, the proposed CLM approach shows less RMSE and high detection rate for landmark localization. The effect of the precise eye center estimation on landmark localization has been analyzed in this paper. It is observed that the for lesser normalized eye localization error shows high accuracy in landmark localization by the proposed method. The performance of the proposed method is also compared with the other method. The experimental results show that the proposed method shows lesser RMSE value compare to the other methods. The future work is to extend the proposed method to localize the facial landmarks under different head poses.

ACKNOWLEDGMENT

This research work has been carried out in the Speech and Image Processing Lab, NIT Silchar, India, and is supported by Visvesvaraya Ph.D. Scheme of MietY, Government of India (Ref No.: PhD-MLA/4(74)/2015-16).

REFERENCES

[1] O. Çeliktutan, S. Ulukaya, and B. Sankur, "A comparative study of face landmarking techniques," *EURASIP Journal on Image and Video Processing*, 2013(1), 13.

[2] H. Gao, H. K. Ekenel, and R. Stiefelhagen, "Pose normalization for local appearance-based face recognition," *International Conference on Biometrics* (pp. 32-41). Springer, Berlin, Heidelberg, 2009.

[3] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A comprehensive survey to face hallucination," *International journal of computer vision*, 2014, 106(1), 9-30.

[4] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 23, no. 6, pp. 681-685, June 2001.

[5] X. Zhao, S. Shan, X. Chai, and X. Chen, "Locality-constrained active appearance model," In *Asian Conference on Computer Vision* (pp. 636-647). Springer, Berlin, Heidelberg, 2012.

[6] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," In *European conference on computer vision (pp. 679-692). Springer, Berlin, Heidelberg* (2012, October).

[7] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 2879-2886).

[8] A. Liang, W. Liu, L. Li, M. R. Farid, and V. Le, "Accurate facial landmarks detection for frontal faces with extended tree-structured models," In *Pattern Recognition (ICPR), 2014 22nd International Conference on* (pp. 538-543).

[9] D. Cristinacce and T. F. Cootes, "Feature detection and tracking with constrained local models," In *Bmvc(Vol. 1, No. 2, p. 3)*, (2006, September).

[10] Y. Wang, S. Lucey, and J. F. Cohn, "Enforcing convexity for improved alignment with constrained local models," In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1-8).

[11] H. Li, K. M. Lam, M. Y. Chiu, K. Wu, and Z. Lei, "Efficient likelihood Bayesian constrained local model," In *Multimedia and Expo (ICME), 2017 IEEE International Conference on* (pp. 763-768).

[12] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable model fitting by regularized landmark mean-shift," *International Journal of Computer Vision*, *91*(2), 200-215, (2011).

[13] H. Kim, J. Jo, K. A. Toh, and J. Kim, "Eye detection in a facial image under pose variation based on multi-scale iris shape feature," *Image and Vision Computing*, 57, 147-164, 2017.

[14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Computer Vision and Pattern Recognition*, 2001, Proceedings of the 2001 IEEE Computer Society Conference, Vol. 1, pp. I-I; 2001.

[15] B. S. Kim, H. Lee, and W. Y. Kim, "Rapid eye detection method for non-glasses type 3D display on portable devices," *IEEE Transactions on Consumer Electronics*, 56(4); 2010.

[16] V. Vapnik, "The nature of statistical learning theory," *Springer science & business media*; 2013.

[17] C. Zhang and Z. Zhang, "A survey of recent advances in face detection," *Technical report, Microsoft Research*, 2010.

[18] X. Yan, "Constrained Local Model for Face Alignment, a Tutorial," https://sites.google.com/site/xgyanhome/home/projects/clm-implementation, (2015).

[19] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, *28*(5), 807-813, (2010). Access date: 29/11/2017.

[20] A. M. Martinez, "The AR face database," *CVC Technical Report #24*, (1998). Access date: 09/10/2017

[21] Jesorsky O, Kirchberg KJ, Frischholz RW, Robust face detection using the Hausdorff distance, In International Conference on Audio-and Video-based Biometric Person Authentication, Springer, Berlin, Heidelberg, 2001. pp. 90-95