ISSN 2300-5963 ACSIS, Vol. 11
DOI: 10.15439/2017F425

Robust face model based approach to head pose estimation

Krzysztof Fornalczyk
Lodz University of Technology,
Institute of Information Technology,
Wolczanska 215, 90-924, Lodz, Poland
Email: krz.fornalczyk@gmail.com

Adam Wojciechowski
Lodz University of Technology,
Institute of Information Technology,
Wolczanska 215, 90-924, Lodz, Poland
Email: adam.wojciechowski@p.lodz.pl

***Abstract*—Head pose estimation from camera images is a computational problem that may influence many sociological, cognitive, interaction and marketing researches. It is especially crucial in the process of visual gaze estimation which accuracy depends not only on eye region analysis, but head inferring as well. Presented method exploits a 3d head model for a user head pose estimation as it outperforms, in the context of performance, popular appearance based approaches and assures efficient face head pose analysis. The novelty of the presented approach lies in a default head model refinement according to the selected facial features localisation. The new method not only achieves very high precision (about 4°), but iteratively improves the reference head model. The results of the head pose inferring experiments were verified with professional Vicon motion tracking system and head model refinement accuracy was verified with high precision Artec structural light scanner.**

I. Introduction

ORIENTATION and movement of human head allow not only to interpret person intentions, but become a part of nonverbal communication as well. It might be exploited in sociological children behaviour monitoring [4], [5], distant computer interface control [6], [31], crowdsourcing systems [7], [8] or in cognitive computation researches [9], [29]. Though very intuitively and naturally accomplished by humans, the problem of head pose estimation is still a challenging problem for current computer systems.

From the computer vision point of view head pose estimation is a process of evaluating head position and orientation from digital images. Eye tracking process which the authors studied in previous researches [1], definitely requires inferring head position and orientation as it considerably affects gaze tracking precision. Moreover psychological investigations show that both head pose and eye direction are strongly correlated and influence person's gaze prediction [2]. In such context (gaze controlled interaction) head pose tracking can be considered relatively to the view direction of the camera rather than global coordinating system.

Current state of art methods [3], [23], targeting at RGB camera image head pose inferring, claim to achieve head orientation angular precision of about 5° for individual axes: pitch, yaw, roll. Proposed head 3d model based solution can not only obtain comparable estimation precision, but also reveals high performance due to robust head image to model assignment - aligned just with a few face predominant key features. The method can start without prior knowledge of user head dimensions and iteratively, during the process of alignment, refines considerably default head model parameters.

II. Head pose estimation methods

Geometrical based approaches to head pose estimation rely on cues such as deviations of the head from bilateral symmetry [10]. They consider both a head shape and a configuration of local features to estimate its pose. Features search-space can be effectively reduced by using knowledge of human face structure. The key aspect seems to be a proper selection and in-face localisation of the face fiducial points.

Horprasert et al. [13] proposed selection of 5 feature points (outer eyes and outer mouth corners and tip of the nose) for reconstruction of head pose. Authors suggested geometrical analysis of vectors connecting feature points: face normal vector rotation and feature points spanned vectors affine projection. Gee et al. [12] and Wang et al. [17] analysed eyes' crossing line and mouth vertically crossing line interrelation. They provided an approach where head pose expectation maximisation was obtained due to eyes and mouth lines perspective convergence analysis. More recently Baltrusaitis et al. [23] considered conditional local neural fields (CLNF) [24] for face features detection and applied their orthographic projection on the camera image plane. Subsequently PnP problem solution was used for an appropriate head pose estimation. Though authors claim that their method head inferring precision varies between 2.8° - 6° (depending on a tested dataset) 3d head model, obtained in the reprojection stage, was not verified for its precision. There are also auxiliary sensors considered for supporting camera view head pose estimation. Morency et al. [25] suggested additional inertial and magnetic sensor drift reduction and Funes et al. [30] used depth data for head orientation analysis.

The geometric methods are fast and simple however their difficulty lies in detecting the features with high precision and accuracy. The process might be even more challenging when features become outlying or missing. The most frequent approaches for in-image face detection, relies on active appearance model (AAM) [28] or active shape model (ASM) [27]. Competitive local approaches, constrained local model (CLM)

TABLE I
INITIAL HEAD SELLION RELATED COORDINATES OF AN AVERAGE ANTHROPOMETRIC HEAD MODEL FEATURE POINTS

Point name	X[mm]	Y[mm]	Z[mm]
Sellion	0	0	0
Right eye	65,5	-5	20
Left eye	-65,5	-5	20
Right ear	77,5	-6	100
Right ear	-77,5	-6	100
Nose	0	-48	-21
Stomion	0	-75	-10
Menton	0	-133	-0

[18] and constrained local neural fields (CLNF) [24], claim to obtain higher precision of landmark detection, especially in strictly constrained, restricted environment (inconvenient light, partial face occlusion). Highly efficient local analysis may also base on gradient templates [29].

Face feature points, retrieved within face analysis, can be subsequently structured in a face or head reference model. 3d model is used for optimizing temporal face features spatial positioning and face image features alignment. For example Kazemi et al. [14] suggested regression trees for one millisecond face alignment. The head model can be reconstructed basing on real head of the user [11] or it can be constructed basing on default average anthropological measures.

On contrary to previous approaches, presented solution coherently refines 3d model while head pose estimation and uses it directly for better head inferring. Elaborated method preserves and in some scenarios outperforms, state-of-the-art methods accuracy.

III. METHOD

Presented method consists of two main steps: head pose estimation and head model refinement. The head model consists of 8 points: sellion, eyes outer corners, ears, top of the nose, stomion (center of mouth) and menton. Points locations are presentened in image 1.

At the beginning, the average anthropometric head model was retrieved. This head model corresponds to the head shape of averaged male adult and was based on anthropometric data collection [20], [21]. The initial values of points coordinates (sellion related) are presented in the table I.

A. Head pose estimation

Head pose estimation stage is divided into 2 substeps: localizing facial landmarks (points corresponding to the 8 selected head model points) and head pose calculation.

Facial landmarks detection method should work in real time and reliably calculate landmarks positions, even in difficult lighting conditions. In our tests, we decided to use the method described in [14] (implemented in [15]), however it is possible to replace it with other methods satisfying mentioned requirements. We decided to use this method, because it performs well even in poor lighting conditions and can deal with long hair, glasses and different skin colors as well. Additionaly it provide more precise results than other currently used methods, such as Supervised Descent Method [32] or Face Alignment by Explicit Shape Regression [33]. The method can retrieve up to 68 face fiducial points from which several facial landmarks should be selected. Though 8 specific points (fig. 1) were selected some substitutions are possible (for example - eyes inner corners instead of outer corners), as well as it might be necessary to change the number and coordinates of the initial points.

Once the facial landmarks are detected, it's possible to calculate algebraically the head position and rotation. We decided to use classic solution of PnP problem - iterative method based on Levenberg-Marquardt optimization. This method is implemented in OpenCV library. Obviously, the calculated pose is not perfect. The most important reasons of pose inaccuracies are: not accurate head model (initial head model was based on averaged anthropometric values) and facial landmarks detection imprecision. First problem is handled during head model refinement step and second problem is addressed in the next substep.

All, currently available, facial landmarks localization methods produce some errors. Usually it's not a real problem, since subpixel accuracy is not required in most use-cases. It is important to note, that usually most of the detected points, are localized with high accuracy and only some points contains errors big enough to produce meaningful inaccuracies in head rotation and translation estimation process. Due to this fact, we decided to use easy method to improve accuracy of our system. After initial head pose calculation, we calculate head model points reprojections. Next, for each point, the reprojection error (distance between detected and reprojected points) is calculated. After that, we repeat calculation of rotation and translation, using all points except the point with the biggest error. The calculated values are final head rotation and translation values. The rationale of this decision is that the removed point most likely contains the biggest localization error and generally makes it much harder (or even impossible) to find good PnP solution. Obviously the removed point, can be perfectly fine, but in this situation it's possible to calculate descent solution from only 7 (instead of 8) points. It's quite easy to note, that this approach is a bit similar to RANSAC method [16]. RANSAC method solves PnP problem multiple times for different random points selected from all 8 points, which makes it robust in precision, but relatively slow (because it tests all 8 possibilites, assuming only 1 outlier). Our method solves PnP problem only twice, which makes it much faster than RANSAC approach. Since we already know which point burdens solving PnP problem, we can achieve optimal solution without testing all other possibilities. In contrast to RANSAC, our method is not based on randomness, which makes testing, evaluating and debugging much easier.

The overall algorithm of head pose estimation, for single frame, is presented in 1.

Algorithm 1 Head pose estimation algorithm

Require: $head\_model$ - model of head (eight 3D points - P_i, where $i = 1, 2, ..., 8$), at the beginning it's initial head model from table I, after each head pose refinement points are adjusted;

Detect and assign facial landmarks:
$landmarks\ L_i \leftarrow detected\ face\ landmarks$

Get rotation (R) and translation (t), solving PnP problem using $landmarks$ as projected points and $head\_model$ as 3D points:
$R, t = solvePnP(landmarks - L_i, head\_model - P_i)$;

Calculate reprojection L'_i of each point P_i of head_model:
$L'_i \leftarrow K[R|t]P_i$
{K - camera intrinsic (focal length and principal point) parameters matrix, P_i - 3D point of the head model}

Calculate error of each reprojection (using $landmarks$):
$e \leftarrow distance(L'_i, L_i)$
$\{(L'_i, L_i)$ - correpsonding_landmarks for $i = 1, 2, ..., 8\}$

Find which point P_k, where $k \in (1, 2, ..., 8)$ produces the biggest reprojection error e, where L'_k - corresponding facial landmark point
$head\_model' \leftarrow all\ points\ from\ head\_model\ except\ P_k$
$landmarks' \leftarrow all\ points\ from\ landmarks\ except\ L'_k$

Solve PnP using $landmarks'$ and $head\_model'$ $R', t' = solvePnP(landmarks', head\_model')$;

R', t' - final result of algorithm

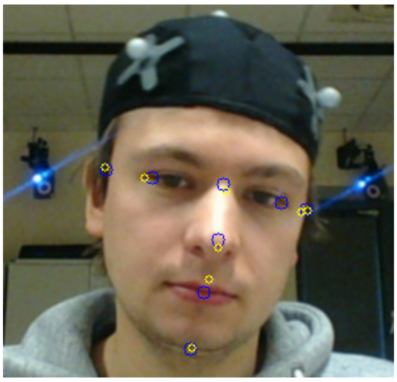

Fig. 1. Detected facial landmarks L_i (blue circles), reprojected head model points L'_i (yellow circles)

B. Head model refinement

As already mentioned, inaccurate head model is one of the most important reasons of head pose inferring. Of course using model from digital 3D scanner is not available in most of cases, therefore we had to find another solution. Analysing results of facial landmarks detection and reprojected head model points it was easy to note that detected landmarks L_i are much more accurate than reprojected points L'_i. Discrepancies were presented in image 1, especially big differences between left eye corner points and mouth center should be noted. Based on that fact we decided to create method which adjust and refine head model points according to the results of facial landmarks detection (i.e. minimize reprojection error). A classic approach for such problem is bundle adjustment method [19], which optimizes both - camera poses and points positions. However, this method tends to be to slow for real time applications and can be replaced with our much simpler approach. Additionally in our method, we can easily use information from facial landmarks detector.

This part of our method is executed only in every n-th (usually 175) frame. The algorithm operates on all frames since last head model refinement.

The result of algorithm 2 (even after single iteration) retrieves more precise face model. This face model was afterwards used in the head pose estimation stage, which results in overall higher accuracy. The biggest refinement takes place in first and second iteration of algorithm. For next iterations, the changes are much smaller, therefore recommended number of iterations is 3 to 5.

Algorithm 2 Head model refinement algorithm

```
v_array ← array of 8 zero vectors(0, 0, 0)
for each frame since last head_model refinement iteration do
    for each facial landmark L_i do
        l ← line through camera (point(0, 0, 0)) and detected landmark L_i; {line should be in parametric form - P = P_0 + t·v⃗}
        R ← rotation calculated for this frame;
        t ← translation calculated for this frame;
        Create [R|t] 3x4 transformation matrix, calculate its pseudo inverse transform and apply it to line l
        P_i ← head model point (corresponding to processed facial landmark point L_i)
        P'_i ← P_i projection on line l {Note that P'_i is the nearest point from P_i that is on line l}
        w⃗ ← P'_i − P_i {w⃗ is the correction vector - if it would be added to P_i the resulting point would give perfect reprojection (for this frame)}
        v_array[i] ← v_array[i] + w⃗
    end for
end for
N ← number of frames used in head model refinement
for each head_model point {P_i} do
    hp = head_model_points[i]
    hp = hp + v_array[i]/N
    head_model_points[i] = hp
end for
```

The resultant model was also interesting on its own - for example it can be part of face reconstruction or recognition system, that's why we decided to check the accuracy of head model refinement. For this purpose we compared results with ground truth data from 3d digital scanner 2.

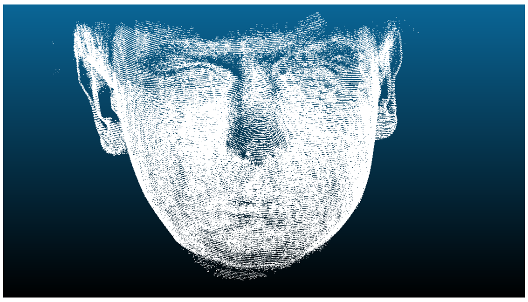

Fig. 2. Digital scan of author's head, acquired using Artec Eva [22] scanner

Fig. 3. Images from recorded movies. Note motion tracking system markers on the head. On (4) we can see markers used to synchronisation. In the background - 2 IR cameras used by motion tracking system.

IV. Experiments Results

The verification of the method accuracy was performed with a professional ground truth data. Head movements of user were recorded simultaneously by camera assigned to the laptop monitor and by professional Vicon passive tracking system. Tests were performed and recorded with standard webcam camera of 640x480 resolution. The precision of the reference Vicon tracking system is very well known and reach 0.5 mm within large (several cubic meters) tracking volume. Both data streams, video camera and head movement tracking system were synchronised. Tests were performed in changing lighting conditions (strong light from one side, blinking light in different colors, etc.). The distance between subject was in range 30-100cm, simulating average laptop usage. An example frames from movies are shown in image 3.

The results of head pose estimation and head model refinement algorithm are presented separately in two independent subsections. In both cases the results are dependent on frequency of head model refinement and maximum number of allowed refinements.

A. Results of head pose estimation

Results of head pose estimation were compared with ground truth data from reference Vicon motion tracking system. In the table II we can find the influence of different parameters on average head rotation angle differences. It can be noticed

TABLE II
Results of head pose estimation errors. x, y, z axes represent respectively pitch, yaw, roll head rotations;

Head model	Remove point with biggest reprojection error	Frames between refinements	X[°]	Y[°]	Z[°]	avg[°]
Initial model	no	-	6,07	5,69	2,45	4,74
	yes	-	4,94	6,05	2,91	4,64
Initial model + refinements	yes	100	6,46	5,25	5,23	5,65
	yes	175	2,87	5,37	2,53	3,59
	no	175	4,14	4,95	2,43	3,84
	yes	250	3,40	5,47	2,53	3,80
	no	250	4,44	5,10	2,51	4,02
	yes	500	4,76	5,46	2,52	4,25

that only selected results have been provided (for discrete set of configuration parameters). The best achieved configuration considered removing the most erroneous head model point with refinement for every 175-th frame. Then the average head orientation error was 3.59 degrees but individual axes reached an error of about 2.5 degree (Z axis representing $roll$ head rotation). The average frames per second value during performing tests was 13.66, which in our opinion is enough for real-time applications.

B. Results of head model refinement

Table III presents exemplary results of head refinement process. Provided parameters were obtained for the least head orientation error described in previous section. The results were obtained with exclusion of the point with the biggest reprojection error, with assignment of number of frames between refinements to 175 and maximum number of refinements set to 5. The overall head model correction reaches only about 10% and is mainly affected by the noise of ears imprecise positioning. This is most likely caused, by the imperfection of tested facial landmarks detection method, especially when ears feature points are not visible. Assuming exclusion of ears feature points, from the head model, its considerable improvement can be presumed.

V. Conclusions

Performed experiments revealed unquestionable high average accuracy of head pose estimation. Average orientation error oscillates around $4°$ what situates presented method at a comparable position with the state-of-the-art approaches. Consecutively corresponding 3d head model robust refinement was performed but still a considerable space for further improvements in this field are noticeable. The refinement process assures much faster, than corresponding bundle adjustment process, face image alignment and resulting head pose interactive (real-time - average FPS 13.66) estimation.

The best results were obtained for relatively seldom refinement (every 175-th frame) and the worst head model point exclusion appeared to improve the final estimation results. According to conducted experiments further exclusion of the worst points, in subsequent iterations, can improve the results

TABLE III
RESULTS OF HEAD MODEL REFINEMENT

Point	Error of initial model[mm]				Error of refined model[mm]			
	x	y	z	sum	x	y	z	sum
Sellion	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Right eye	**-19,62**	**-0,04**	**2,76**	**22,43**	**0,56**	**-2,33**	**0,29**	**3,18**
Left eye	**17,01**	**0,93**	**4,39**	**22,35**	**0,51**	**-0,35**	**5,51**	**6,38**
Right ear	**8,99**	**26,34**	**12,16**	**47,50**	**16,19**	**25,30**	**12,64**	**54,14**
Right ear	**-12,60**	**27,96**	**6,43**	**47,00**	**-22,58**	**26,02**	**11,93**	**60,55**
Nose	**-3,20**	**27,86**	**30,00**	**61,07**	**-0,65**	**13,19**	**35,62**	**49,48**
Stomion	**-1,97**	**3,22**	**3,68**	**8,88**	**-2,10**	**-3,01**	**6,07**	**11,19**
Menton	**-1,17**	**-0,93**	**22,16**	**24,28**	**-0,45**	**3,18**	**25,16**	**28,81**
Sum	**-**	**-**	**-**	**233,52**	**-**	**-**	**-**	**213,75**

only within one or two further steps - if the points misalignment is relatively big. If not, further *bad points* removement results in perceiveable worse general method precision.

The proposed head model refinement corrects the average head orientation error by about 20% - for considered method coefficients average head orientation error decreased from 4.64° to 3.59°.

Further perspectives of the method improvements encompass: filtering components of the refinement vectors specified in the algorithm 2, introducing certain anthropometric constraints of the head model modifications as not to allow extensive model degeneration while refinement (in this context symmetry of the face might be considered for further face feature points stabilization and for head model refinement stabilization as well) and calculating head position using information about facial featues detection uncertainty (for example - usually ears are detected with bigger error than nose or eyes).

REFERENCES

[1] A. Wojciechowski, and K. Fornalczyk, "Single web camera robust interactive eye-gaze tracking method", *Bulletin of the Polish Academy of Sciences,* vol. 63 no.4, pp. 879, 2015.

[2] S. Langton, H. Honeyman, and E. Tessler, "The influence of head contour and nose angle on the perception of eye-gaze direction", *Perception and Psychophysics*, vol. 66, no. 5, pp. 752-771, 2004.

[3] E. Murphy-Chutorian, and M. M. Trivedi, "Head pose estimation in computer vision: A survey", *IEEE transactions on pattern analysis and machine intelligence* vol. 31 no.4, pp. 607-626, 2009.

[4] J. M. Rehg, G. D. Abowd, A. Rozga, M. Romero, M. A. Clements, S. Sclaroff, I. Essa, O. Y. Ousley, Y. Li, K. Chanho, H. Rao, J. C. Kim, L. L. Presti, J. Zhang, D. Lantsman, J. Bidwell, and Z. Ye, "Decoding Children's Social Behavior", *Computer Vision and Pattern Recognition (CVPR)*, pp. 3414-3421, 2013.

[5] P. Kucharski, P. Łuczak, I. Perenc, T. Jaworski, A. Romanowski, M. Obaid and P. W. Woźniak, "APEOW: A personal persuasive avatar for encouraging breaks in office work", *Proc. of the 2016 FedCSIS Conf., Eds. M. Ganzha, L. Maciaszek and M. Paprzycki*, IEEE, ACSIS, Vol. 8, pages 1627-1630, 2016.

[6] D. Rozado, A. El. Shoghri, and R. Jurdak, "Gaze dependant prefetching of web content to increase speed and comfort of web browsing", *Int. J. of Human-Computer Studies* vol. 78, pp. 31-42, 2015.

[7] C. Chen, P. Wozniak, A. Romanowski, M. Obaid, T. Jaworski,, J. Kucharski, K. Grudzień, S. Zhao, M. Fjeld, "Using Crowdsourcing for Scientific Analysis of Industrial Tomographic Images", *ACM Trans. on Intel. Syst. and Tech.*, Vol. 7 Issue 4, art no. 52, 25p., 2016.

[8] I. Jelliti, A. Romanowski, K. Grudzień, "Design of Crowdsourcing System for Analysis of Gravitational Flow using X-ray Visualization", *Proc. of the 2016 FedCSIS Conf., Eds. M. Ganzha, L. Maciaszek and M. Paprzycki*, IEEE, ACSIS, Vol. 8, pages 1613-1619, 2016.

[9] Q. Zhao, and Ch. Koch, "Learning saliency-based visual attention: A review". *Signal Processing*, vol. 93 no. 6, pp. 1401-1407, 2013.

[10] H. Wilson, F. Wilkinson, L. Lin, and M. Castillo, "Perception of head orientation", *Vision Research*, vol. 40, no. 5, pp. 459-472, 2000.

[11] M. Kowalski, and W. Skarbek, "Online 3D face reconstruction with incremental Structure From Motion and a regressor cascade", *Symp. on Photonics Applications in Astronomy, Communications, Industry and High-Energy Physics Experiments. Int. Soc. for Opt. and Phot.*, 2014.

[12] A. Gee, and R. Cipolla, "Determining the gaze of faces in images", *Image and Vision Computing*, vol. 12, no. 10, pp.639-647, 1994.

[13] T. Horprasert, Y. Yacoob, and L. Davis, "Computing 3-d head orientation from a monocular image sequence", *Proc. Int. Conf. Automatic Face and Gesture Recognition*, pp. 242-247, 1996.

[14] V. Kazemi, and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867-1874, 2014.

[15] Dlib C++ Library., http://dlib.net/

[16] M. Fischler, and R. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Comm. of the ACM*, vol. 24 no. 6, pp. 381-395, 1981.

[17] J. G. Wang, and E. Sung, (2007). "EM enhancement of 3D head pose estimated by point at infinity", *Image and Vision Computing*, vol. 25 no. 12, 1864-1874.

[18] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3444-3451, 2013.

[19] R. Hartley, and A. Zisserman, "Multiple view geometry in computer vision", 2nd edition, *Cambridge Univ. Press*, 2004.

[20] Static adult human physical characteristics of the head., https://en.wikipedia.org/wiki/Human_head#/media/File:HeadAnthropometry.JPG

[21] A head-and-face anthropometric survey of U.S. respirator users., https://www.nap.edu/resource/11815/Anthrotech_report.pdf

[22] Artec Eva laser scanner., https://www.artec3d.com/3d-scanner/artec-eva

[23] T. Baltrusaitis, P. Robinson, L. P. Morency, "Openface: an open source facial behavior analysis toolkit", *App. of Comp. Vision*, p. 1-10, 2016.

[24] T. Baltrusaitis, P. Robinson, L. P. Morency, "Constrained local neural fields for robust facial landmark detection in the wild", *Proc. of the IEEE Int. Conf. on Comp. Vision Work.*, p. 354-361, 2013.

[25] L. P. Morency, J. Whitehill, and J. Movellan, "Generalized adaptive view-based appearance model: Integrated framework for monocular head pose estimation", *Automatic Face and Gesture Recognition, 8th IEEE International Conference on. IEEE*, p. 1-8, 2008.

[26] N. Wang, X. Gao, D. Tao, and X. Li. "Facial feature point detection: A comprehensive survey", *CoRR*, 2014.

[27] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graha, "Active shape models-their training and application", *Computer vision and image understanding*, vol. 61 no. 1, pp. 38-59, 1995.

[28] G. J. Edwards, Ch. J. Taylor and T.F. Cootes, "Interpreting face images using active appearance models", *Automatic Face and Gesture Recognition, Proc. Third IEEE Int. Conf. on. IEEE*, pp. 300-305, 1998.

[29] R. Staniucha, and A. Wojciechowski, "Mouth features extraction for emotion classification", *Computer Science and Information Systems (FedCSIS), 2016 Federated Conference on. IEEE*, pp. 1685-1692, 2016.

[30] K. A. Funes, "3D Gaze Estimation from Remote RGB-D Sensors", PhD Thesis, Ecole Polytechnique Federale de Lausanne, 2015.

[31] M. Kowalczyk, and P. Napieralski, "An Effective client-side object detection method on the Android platform", *Journal of Applied Computer Science*, vol. 23, pp. 29-38, 2015.

[32] X. Xiong, and F. Torre, "Supervised Descent Method and its Applications to Face Alignment", *Comp. Vision and Pattern Rec.*, 2013.

[33] X. Cao, Y. Wei, F. Wen and J. Sun, "Face Alignment by Explicit Shape Regression", *International Journal of Computer Vision*, vol. 107, pp. 177-190, 2014.