Vol.2, No.6, 425-430 (2009) **JBiSE**
doi:10.4236/jbise.2009.26061

Unsupervised human height estimation from a single image

Ye-Peng Guan[1,2]

[1]School of Communication and Information Engineering, Shanghai University, Shanghai, China; [2]Key Laboratory of Advanced Displays and System Application, Ministry of Education, Shanghai, China.
Email: ypguan@shu.edu.cn

Received 15 May 2008; revised 16 June 2009; accepted 17 June 2009.

ABSTRACT

The single image containing only a human face not previously addressed in the literature is employed to estimate body height. The human face especially the facial vertical distribution possesses some important information which strongly correlates with the stature. The vertical proportions keep up relative constancy during the human growth. Only a few facial features such as the eyes, the lip and the chin are necessary to extract. The metric stature is estimated according to the statistical measurement sets and the facial vertical golden proportion. The estimated stature is tested with some individuals with only a single facial image. The performance of the proposed method is compared with some similar methods, which shows the proposal performs better. The experimental results highlight that the developed method estimates stature with high accuracy.

Keywords: Human Height Estimation; Golden Proportion; Facial Proportion; Feature Extraction; Projection Model

1. INTRODUCTION

Human height estimation has many important applications such as soft-biometrics and human tracking [1]. In the first case, the stature can be used to rule out the possibility that a particular person is the same person from the surveillance cameras [2,3]. In the latter case, it can be exploited to distinguish among a small set of tracked people in the scene [1,4,5,6,7]. The stature, therefore, may become a very useful identification feature. In the cases with two or more images, the stature estimation by stereo matching is computationally expensive and there exists ambiguity being resident in the stereo correspondence, which is not overcome efficiently so far [8,9,10]. In the case of only one image available, the stature measurement from single image has to be performed. Many single view based approaches are proposed based on some geometric structures or models [1,2,3,4,5,6, 7,11,12,13,14,15,16]. In [1,2,3,4,5,6,7,12,13,14,15,16], plane metrology algorithms based on vanishing points and lines are developed to measure distances or lengths on the planar surfaces parallel to the reference plane. Any slight inaccuracy in measuring vanishing points will result in large errors [2,17]. Reference points defining the top and bottom of the object should be clear and unambiguous. Besides, reference objects must be in the same plane as the target object. Moreover, the full body of the user must be visible in measuring the stature. In [11], BenAbdelkader and Yacoob incorporated some certain statistical properties of human anthropometry into the stature estimation. It would be difficult to obtain the stature when some anthropometric values such as acromion and trapezius (or neck) are not available. Besides, if the whole body and/or the facial plane are at an angle with respect to the camera, the weak perspective assumption cannot be exploited.

Although many single view based approaches have been proposed, there exist some problems in the literature. In order to improve both validation and automation of the procedure, we have developed an approach to estimating the stature with the following advantages. 1) The used image contains only the human face not previously addressed in the literature. The human face especially the facial vertical distribution owns some important information which strongly correlates with the stature [18]. The facial vertical proportion used is based on the golden proportion [19,20]. A number of investigators have commented on the relative constancy of the facial vertical proportions during the human growth [21]. 2) Only a few facial features are necessary to extract, and the procedure is processed automatically by image analysis operators. 3) The extracted facial features and the facial vertical proportions are used together to estimate the stature. The estimated result is tested with some individuals with only a facial image, showing high esti-

mation accuracy, which validates the developed approach to be objective, and can be taken as an automated tool for estimating the stature. The potential applications of this work include biometric diagnoses, user authentication, smart video surveillance, human-machine interface, human tracking, athletic sports analysis, virtual reality, and so on.

The rest of the paper is organized as follows: Section 2 describes the facial proportions used for estimating the stature. Section 3 discusses the face detection and the facial features extraction. Section 4 describes the stature estimation based on a calibrated camera. Experimental results and evaluation of the performance are given in Section 5 and followed by some conclusions and future works in Section 6.

2. FACIAL PROPORTIONS

All living organisms including humans are encoded to develop and confirm to a certain proportion [20]. The human face especially the facial vertical proportion owns some important information which correlates with the stature [18]. The facial vertical proportions include the golden proportion [19,20] and the facial thirds method [22,23]. The facial golden proportion is approximately the ratio of 1.618 to 1 as shown in **Figure 1**. It states that the human face may be divided into a golden proportion distribution by drawing horizontal lines through the forehead hairline, the nose, and the chin, or through the eyes, the lip, and the chin.

The facial thirds method states that the face may be divided into roughly equal a third by drawing horizontal lines through the forehead hairline, the eyebrows, the base of the nose, and the edge of the chin as shown in **Figure 2**. Besides, the distance between the lip and the chin is double the distance between the base of the nose and the lip [22,23].

The golden proportion and the facial thirds are similar to each other. The former specifies a larger number of proportions than the latter. They used some different features so that they cannot be directly compared [19]. By experiments it shows that the accuracy of stature estimated by the golden proportion is more consistent with the ground-truth data than that of by the facial thirds.

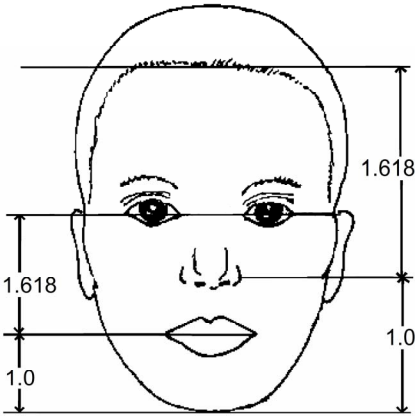

Figure 1. Facial vertical golden proportion

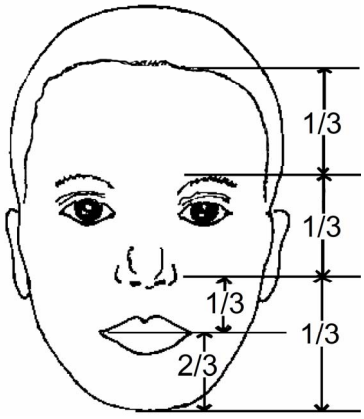

Figure 2. Facial vertical thirds proportion

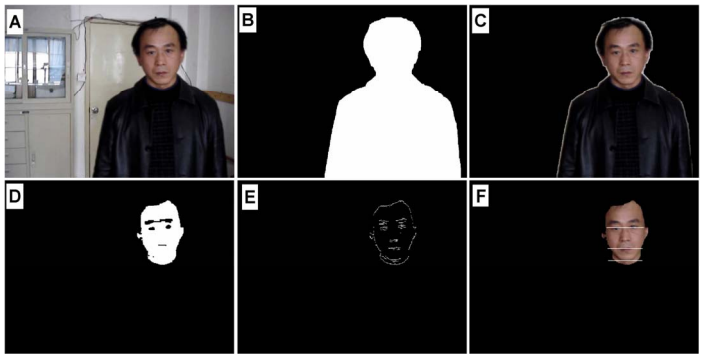

Figure 3. Face detection and facial features extraction results. A: original image. B: extracted binary foreground. C: extracted foreground object. D: extracted binary blob. E: resulting image from horizontal edge detection. F: horizontal locations of the eyes, the lip and chin lines.

3. FOREGROUND DETECTION AND FACIAL FEATURE EXTRACTION

Robust and efficient extraction of foreground object from image sequences is a key operation. Many algorithms have been developed [24,25,26]. The algorithm proposed in [26] is employed to extract foreground blob as shown in **Figure 3B**. After extracting the foreground objects, accurate facial features extraction is important for reliable estimation of the stature. A number of methods have been developed for extracting facial features [27,28,29]. Among the facial features such as the eyes, nose, lip, chin, and so on, the eyes are one of the most important facial features [30,31]. Since effective automatic location and tracking of a person's forehead hairline is difficult, we select the eyes, lip, and the chin as the facial features used in the study.

The first step consists of locating the facial region to remove irrelevant information. Human skin color, though differs widely from person to person, is distributed over a very small area on a C_bC_r plane [32,33]. This model is robust against different types of skin, such as those of people from Europe, Asia and Africa. The skin tone pixels are detected using the C_b and C_r components. Let the thresholds be chosen as $[C_{b1}, C_{b2}]$ and $[C_{r1}, C_{r2}]$, a pixel is classified to skin tone if the values $[C_b, C_r]$ fall within the thresholds. Each pixel in the C_b and C_r layer which does not meet the range $[C_b, C_r]$ is set to zero. In some cases, the obtained mask has concavities or spikes as shown in **Figure 3D**, which affects the facial features location. We use the algorithm in [34] to process this problem.

There are many fairly long horizontal edges near the facial features. In order to make the edge detector behaves more stable, we transform the intensity of the image into a second derivate and then horizontally project it to determine the horizontal positions of facial features [34] (seen from **Figure 3E**). The positions of peaks in the horizontal projection curve correspond with the horizontal facial features including the eyes, nose and lip. Horizontal transition is stronger at the lip than at the eyes in some cases. The lip can be detected in hue/saturation color space [35]. We detect the peak with a maximal value above the lip as the horizontal position of eyes. In the meantime, the chin is automatically located between the lip and the neck. The horizontal location of the eyes, lip and chin is given in **Figure 3F**.

4. STATURE ESTIMATION

Stature estimation is discussed here to highlight the use of the extracted facial features. Assume the person stands or walks on a plane and a camera is calibrated with respect to this plane. We compute 3D position of the extracted facial features according to the golden proportion. Since the facial vertical proportions keep relative constant during the human growth [21], the 3D position of the extracted facial features can be determined if a certain length or distance is known.

More detailed metric description of the head and face was recognized quite early. Major U.S. surveys, those in which large numbers of measurements have been made on samples of a thousand or more individuals, have been carried out on military personnel [18]. The measurement device could provide a sensitivity of less than 0.01 mm in each axis, and the accuracy with the order of 0.1 mm could be achieved [36]. Shiang [18] has made extensive 3D statistical work of human head and face also. According to the measurement sets the metric stature can be estimated based on the calibrated camera.

The camera model used is a central projection. Effect such as radial distortion can be removed and is not detrimental to the method. The camera perspective projection model can be represented by a 3×4 matrix M. The image coordinates (u_i, v_i) of a point P_i expressed in a homogenous coordinate system are given as follows:

$$s_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \quad (1)$$

When estimating the stature, we assume that depth difference among the eyes, lip and the chin is negligible. For simplifying computation, assume that the 3D coordinates of the chin point (u_1, v_1) is (X, Y, Z) whose horizontal line intersects with the vertical line through the center of the eyes. The coordinates of the lip point (u_2, v_2) is $(X, Y, Z+h)$ whose horizontal line intersects with the same vertical line as above. According to the golden proportion the coordinates of the center of the eyes point (u_3, v_3) is $(X, Y, Z+2.618h)$. If h, the height between the chin and the lip is known, (1) can be used to infer the 3D coordinate of the chin point. Expending the (X, Y, Z) gives as follows:

$$A \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = B \quad (2)$$

where,

$$A = \begin{bmatrix} u_1m_{31} - m_{11} & u_1m_{32} - m_{12} & u_1m_{33} - m_{13} \\ v_1m_{31} - m_{21} & v_1m_{32} - m_{22} & v_1m_{33} - m_{23} \\ u_2m_{31} - m_{11} & u_2m_{32} - m_{12} & u_2m_{33} - m_{13} \\ v_2m_{31} - m_{21} & v_2m_{32} - m_{22} & v_2m_{33} - m_{23} \\ u_3m_{31} - m_{11} & u_3m_{32} - m_{12} & u_3m_{33} - m_{13} \\ v_3m_{31} - m_{21} & v_3m_{32} - m_{22} & v_3m_{33} - m_{23} \end{bmatrix} \quad (3)$$

$$B = \begin{bmatrix} m_{14} - u_1 m_{34} \\ m_{24} - v_1 m_{34} \\ m_{14} - u_2 m_{34} - (u_2 m_{33} - m_{13})h \\ m_{24} - v_2 m_{34} - (v_2 m_{33} - m_{23})h \\ m_{14} - u_3 m_{34} - 2.618(u_3 m_{33} - m_{13})h \\ m_{24} - v_3 m_{34} - 2.618(v_3 m_{33} - m_{23})h \end{bmatrix} \quad (4)$$

From (2), the linear least-squares solution is given by

$$[X,\ Y,\ Z]^{\mathrm{T}} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}B \quad (5)$$

Once person's head-top point (u_4, v_4) are known, we can rearrange (2) to estimate the stature as (6) using the coordinates (X, Y). It is found that (6) as a function of v_4 can get a more stable estimation of the stature.

$$H = \frac{m_{24} - v_4 m_{34} - (v_4 m_{31} - m_{21})X - (v_4 m_{32} - m_{22})Y}{v_4 m_{33} - m_{23}} \quad (6)$$

5. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed method, we have done experiments with some individuals. The experiments are performed with a CCD camera which produces 640×480 pixels image sequences. The camera is mounted overhead which look down at an oblique angle to capture human face. In the experiment, the parameters of skin segmentation are fixed for all images as follows: C_{b1}=83, C_{b2}=127, C_{r1}=140, C_{r2}=175. The parameter h used in (4) is selected as 44.96 mm.

The experimental setup includes a wall screen with a maximum size of 2.4 m × 4 m which is parted into 2×4 intended panels pointing by the users. The users' position is about 2 m~4.5 m away from the screen and the user can walk freely in the experimental room. The room size is about 3 m×5 m. The experiment is performed as following way. We capture the pointing person as she/he is pointing at the intended panel. We extract the pointing user's face and estimate his stature. **Figure 4(a)** gives some input video images as a user pointing at the panels. The tested user's face is shown in some light blue pixels superimposed on the original images. The estimated stature (unit: mm) is given in the image also. **Figure 4(b)** shows the stature as the user pointing at the panels at different locations. The symbol in the legend refers to some different statures: the real human height (real H), the thirds proportion based estimated height (thirds based H), and the golden proportion based estimated height (the proposed H).

The standard derivation σ of the estimated stature by the proposal is 6.744 mm, and the maximal deviation from the real height is 17.80 mm which is accurate within 3σ. Correspondingly, the σ of the thirds based estimated height is 10.4188 mm, while the maximal deviation from the real height is 36.9 mm which is out of 3σ. The proposed method outperforms the thirds based method in estimating the stature.

The deviation is due to the fact that relative error increase with the distance between the camera and the user since the pixel size is proportional to the view angle, which means smaller resolution from a larger distance. In the experiment a pixel difference corresponds to about 5.26 mm resolution ambiguity from the distance of 3 m. Besides, inaccurate camera calibration affects the measurement results. Another factor is human gesture. Human gesture involves periodic up-and-down displacement. Some other factors such as occlusion, and face orientation may affect the estimation result. This problem can be avoided by using multiple cameras and a camera with the best view is used to estimate the stature.

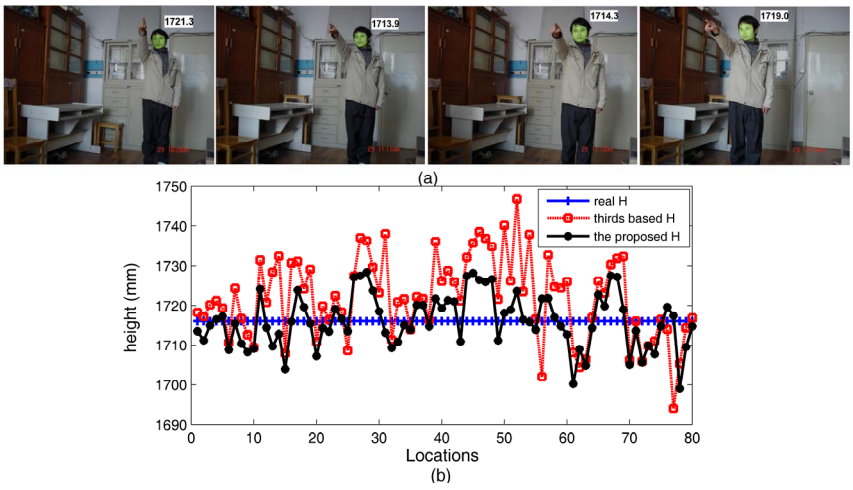

Figure 4. Height results. (a) The person pointing at the panels and the extracted face shown in light blue pixels superimposed on the original images. The estimated height is given based on the proposal also. (b) Heights at different locations.

Table 1. Height estimation results.

Algorithm	Tested object											
	1			2			3			4		
	AM	σ	*MD*	*AM*	σ	*MD*	*AM*	σ	*MD*	*AM*	σ	*MD*
Kim[7]	1817.2	5.12	10.77	1708.2	8.56	15.34	1668.5	6.45	15.36	1653.6	6.75	15.75
Wang[12]	1820.4	7.35	20.65	1722.4	10.56	26.74	1685.4	8.34	23.74	1664.5	9.26	26.51
Lee[15]	1818.5	6.45	17.34	1720.6	9.78	22.58	1682.8	7.67	22.46	1663.7	8.78	25.12
Proposed	**1813.6**	**4.89**	**8.90**	**1715.4**	**8.78**	**15.56**	**1672.3**	**6.16**	**15.13**	**1647.3**	**6.69**	**14.90**

We hope to discuss it in the future. The processing speed of the proposal is roughly 15frames/s for a single object and frontal face in the scene.

To verify the effectiveness of the mentioned approach, we have performed experiments with some moving individuals. The performance of the proposal for four tested individuals with known stature is compared with that of some similar methods [7,12,15] shown in **Table 1**. **Table 1** summarizes the average measurement (*AM*) stature, the standard derivation (σ), and the maximal derivation (*MD*) for each tested object. It is clearly seen that the developed approach performs better.

6. CONCLUSIONS AND FUTURE WORKS

We have developed unsupervised single view based method for robust and real-time estimating the stature. The image contains only a face or upper body little discussed in the literature. Only a few facial features such as the eyes, lip and chin are necessary to extract. The metric stature is estimated according to the statistical measurement sets and the facial vertical golden proportion. The estimated stature is tested with some individuals with only a facial image, showing high accuracy, which validates the proposal to be objective, and can be taken as an automated tool for estimating the stature. Extension to un-calibrated scenario case would be developed in the future.

7. ACKNOWLEDGEMENTS

This work in part is supported by the National Natural Science Foundation of China (Grant No. 60872117).

REFERENCES

[1] C. Ben-Abdelkader, R. Cutler, and L. S. Davis, (2002) Person identification using automatic height and stride estimation, Proceedings of 16th International Conference on Pattern Recognition, **4**, 377–380.

[2] A. Criminisi, (2002) Single-view metrology: Algorithms and applications, Proceedings of 24th DAGM Symposium on Pattern Recognition, 224–239.

[3] N. Saitoh, K. Kurosawa, and K. Kuroki, (1999) A study on height measurement from a single view, Proceedings of International Conference on Image Processing, **3**, 523–526.

[4] A. Bovyrin and K. Rodyushkin, (2005) Human height prediction and roads estimation for advanced video surveillance systems, Proceedings of IEEE Conference on Advanced Video and Signal-Based Surveillance, 219–223.

[5] C. S. Madden and M. Piccardi, (2005) Height measurement as a session-based biometric, Proceedings of Image and Vision Computing New Zealand, 282–286.

[6] D. De Angelis, R. Sala, A. Cantatore, P. Poppa, M. Dufour, M. Grandi, and C. Cattaneo, (2007) New method for height estimation of subjects represented in photograms taken from video surveillance systems, International Journal of Legal Medicine, **121(6)**, 489–492.

[7] D. Kim, J. Lee, H. -S.Yoon, and E. -Y. Cha, (2007) A non-cooperative user authentication system in robot environments, IEEE Trans. Consumer Electronics, **53(2)**: 804–811.

[8] L. Zhang, (2006) Fast stereo matching algorithm for intermediate view reconstruction of stereoscopic television images, IEEE Trans. Circuits and Systems for Video Technology, **16(10)**, 1259–1270.

[9] F. Remondino, S. F. El-Hakim, A. Gruen, and L. Zhang, (2008) Turning images into 3-D models, IEEE Signal Processing Magazine, **25(4)**, 55–65.

[10] W. Xiong, H. S Chung, and J. Jia, (2009) Fractional stereo matching using expectation-maximization, IEEE Trans. Pattern Analysis and Machine Intelligence, **31(3)**, 428–443.

[11] C. Ben Abdelkader and Y. Yacoob, (2008) Statistical body height estimation from a single image, Proceedings of 8th IEEE International Conference on Automatic Face and Gesture Recognition, 1–7.

[12] R. Wang and F. P. Ferrie, (2008) Self-calibration and metric reconstruction from single images, Proceeding of International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 639–644.

[13] L. Fengjun, Z. Tao, and R. Nevatia, (2002) Self-calibration of a camera from video of a walking human, Proceedings of 16th International Conference on Pattern Recognition, **1**, 562–567.

[14] Y. Kida, S. Kagami, T. Nakata, M. Kouchi, and H. Mizoguchi, (2004) Human finding and body property estimation by using floor segmentation and 3D labeling, Proceedings of IEEE International Conference on Systems, Man and Cybernetics, **3**, 2924–2929.

[15] S. -H. Lee and J. -S. Choi, (2007) A single-view based framework for robust estimation of height and position of moving people, Proceedings of Pacific-Rim Symposium on Image and Video Technology, 562–574.

[16] E. Jeges, I. Kispal, and Z. Hornak, (2008) Measuring human height using calibrated cameras, Proceedings of Conference on Human System Interactions, 755–760.
[17] A. Criminisi, I. Reid, and A. Zisserman, (1999) Single view metrology, Proceedings of 7th International Conference on Computer Vision, 434–442.
[18] T. -Y. Shiang, (1999) A statistical approach to data analysis and 3-D geometric description of the human head and face, Proceedings of the National Science Council, Republic of China, Part B, Life Sciences, **23(1)**, 19–26.
[19] H. Gunes and M. Piccardi, (2006) Assessing facial beauty through proportion analysis by image processing and supervised learning, International Journal of Human-Computer Studies, **64(12)**, 1184–1199.
[20] Y. Jefferson, (2004) Facial beauty: Establishing a universal standard, International Journal of Orthodontics, **15(1)**, 9–22.
[21] C. E. Nasjleti and C. J. Kowalski, (1975) Stability of upper face height-total face height ratio with increasing age, Journal of Dental Research, **54(6),** 12–41.
[22] H. L. Obwegeser and L. J. Marentette, (1986) Profile planning based on alterations in the positions of the bases of the facial thirds, Journal of Oral and Maxillofacial Surgery, **44(4)**, 302–311.
[23] C. Sforza, A. Laino, R. D'Alessio, C. Dellavia, G. Grandi, and V. F. Ferrario, (2007) Three-dimensional facial morphometry of attractive children and normal children in the deciduous and early mixed dentition, The Angle Orthodontist, **77(6)**, 1025–33.
[24] R. Li, S. Yu, and X. Yang, (2007) Efficient spatio-temporal segmentation for extracting moving objects in video sequences, IEEE Trans. Consumer Electronics, **53(3)**, 1161–1167.
[25] S. Youn, J. -H. Ahn, and K. Park, (2008) Entrance detection of a moving object using intensity average variation of subtraction images, Proceedings of International Conference on Smart Manufacturing Application, 459–464.
[26] Y. P. Guan, (2008) Wavelet multi-scale transform based foreground segmentation and shadow elimination, The Open Signal Processing Journal, **1(6)**, 1–6.
[27] K. Sridharan and V. Govindaraju, (2005) A sampling based approach to facial feature extraction, Proceedings of 4th IEEE Workshop on Automatic Identification Advanced Technologies, 51–56.
[28] F. Song, D. Zhang, D. Mei, and Z. Guo, (2007) A multiple maximum scatter difference discriminant criterion for facial feature extraction, IEEE Trans. Systems, Man, and Cybernetics, Part B, **37(6)**, 1599–1606.
[29] Y. He, (2009) Real-time nonlinear facial feature extraction using Cholesky decomposition and QR decomposition for face recognition, Proceedings of International Conference on Electronic Computer Technology, 306–310.
[30] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, (2000) The feret evaluation methodology for face-recognition algorithms, IEEE Trans. Pattern Analysis and Machine Intelligence, **22(10)**, 1090–1104.
[31] A. M. Martinez, (2002) Recognizing imprecisely localized partially occluded and expression variant faces from a single sample per class, IEEE Trans. Pattern Analysis and Machine Intelligence, **24(6)**, 748–763.
[32] H. Wang and S. F. Chang, (1997) A highly efficient system for automatic face region detection in MPEG videos, IEEE Trans. Circuit System Video Technology, **7(4)**, 615–628.
[33] C. Garcia and G. Tziritas, (1999) Face detection using quantized skin colour regions merging and wavelet packet analysis, IEEE Trans. Multimedia, **1(3)**, 264–277.
[34] Y. Guan, (2007) Robust eye detection from facial image based on multi-cue facial information, Proceedings of IEEE International Conference on Control and Automation, 1775–1778.
[35] G. Chetty and M. Wagner, (2004) Automated lip feature extraction for liveness verification in audio-video authentication, Proceedings of Image and Vision Computing, 17–22.
[36] J. F. Annis and C. C. Gordon, (1988) The development and validation of an automated headboard device for measurement of three-dimensional coordinates of the head and face, Tech. Report, http://oai.dtic.mil/oai/oai.