BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 65, No. 5, 2017
DOI: 10.1515/bpasts-2017-0079

Mimicking speaker's lip movement on a 3D head model using cosine function fitting

I. LÜSI[1] and G. ANBARJAFARI[1,2]

[1] iCV Research Group, Institute of Technology, University of Tartu, Tartu 50411, Estonia
[2] Department of Electrical and Electronic Engineering, Hasan Kalyoncu University, Gaziantep, Turkey

Abstract. Real-time mimicking of human facial movement on a 3D head model is a challenge which has attracted attention of many researchers. In this research work we propose a new method for enhancing the capturing of the shape of lips. We present an automatic lip movement tracking method which employs a cosine function to interpolate between extracted lip features in order to make the detection more accurate. In order to test the proposed method, mimicking lip movements of a speaker on a 3D head model is studied. Microsoft Kinect II is used in order to capture videos and both RGB and depth information are used to locate the mouth of a speaker followed by fitting a cosine function in order to track the changes of the features extracted from the lips.

Key words: 3D lip movement modelling, mathematical modelling, depth information analysis, cosine function fitting, human-computer interaction.

1. Introduction

The automatic detection of facial features is still a work in progress. Even though numerous methods have been developed, various problems remain [1–6]. Those issues are caused by lighting condition changes, noise and the different characteristics of human faces [7–10].

Seo et al. proposed a face detection and facial feature extraction method that made use of colour snakes [11]. The proposed method consisted of three main steps, which were face region estimation by using the characteristics of skin colour, template matching and then facial feature extraction. The classic energy model was changed by utilising the pixel intensity and adding weights to the energy equation of a node. This approach stopped the contour from randomly converging into a point or a line. However this being solely based on RGB information, a template matching method was proposed to make the method less dependable on lighting conditions. The template was matched to vertical edges obtained from edge detection. The template was scaled according to the results obtained by facial area extraction. This combined method yielded good results in differently illuminated environments.

Scatter difference criterion was used for facial feature extraction in [12]. This was a generalisation of the maximum scatter difference (MSD) algorithm [13]. Instead of calculating the criterion based on the maximum and minimum values of the scatter matrix, the optimisation was achieved by combining an orthonormal basis matrix with scatter matrices and finding extreme values of their traces. This way a binary criterion became able to distinguish a greater number of objects. Also the algorithm gained in speed and stability.

Qi et al. introduced an improved facial feature detection based on blocking [14]. Their proposed algorithm benefitted from the usage of constant facial organ relations in computational efficiency and accuracy. The face area was divided into a grid of 3-by-3. For most humans a specific block contained the nose, another one the left eye, and so on. The method searched for a set of 9 blocks achieved from a binary representation of the image that best corresponded to the characteristics of human face.

In speech analysis based on video feed, two major branches have been developed: one is to temporarily detect important points of the lips, like corners, the other one is to analyse the overall change in the mouth area and shape. Both methods can be used to decide whether a person is speaking or quiet at the moment.

A statistical approach for lip activity detection was introduced in [15]. This method aimed at distinguishing visual silence from visual speech. The authors noted that opening of the mouth produced a radical increase of low-intensity pixels. To make the algorithm distinguish between speech and non-speech a video specific threshold was defined using Neyman-Pearson theorem [16]. Accordingly, a hypothesis of lip activity was proposed and by using a statistical algorithm, the likelihood of the hypothesis was calculated. For confirmed lip activity both the increased number of low-intensity pixels (smaller than said threshold) and a large variance in their number had to be detected.

A method for analysing planning meetings and determining the speaker was proposed in [17]. The face areas of different people were extracted by using individual skin colour filters. After that the mouth area was found by using a template and calculating the probabilities. The focus of the paper was merging the RGB and vocal information. The authors decided to only track whether a person changed the position of their mouth or their head. The movement was extracted by comparing the

[*] e-mail: shb@icv.tuit.ut.ee

Manuscript submitted 2016-07-19, revised 2017-03-19, initially accepted for publication 2017-04-19, published in October 2017.

mouth area or the face area between two subsequent frames in the video. The results showed that this kind of method worked well for analysing the dynamics of the meeting and what kind of role each speaker played in it.

Caplier suggested a template based method for mouth transformation tracking [18]. Only mouth area was captured by using a camera that was attached to a helmet. The line between lips was extracted as a dark line in the middle of the mouth area and all the landmark points were computed based on that. By analysing a set of closed and opened mouths separately, two templates were comprised calculating the average of both options. After that the landmark points were used to deform the template according to the lips under observation. Then Kalman filter [19] was initialised and later it was used to track movements. If the mouth changed its state the template was chosen again and modified, and Kalman filter was reinitialised.

The shortcomings of these methods are that they tend to be either computationally very complex or quite angular in essence. This happens because a discrete set of points will always have missing data when approximating a curved object. This can be reduced by increasing the number of tracked points, however this can lead to higher computational complexity or require a database with a large number of accurate labels for training. To overcome these issues a simple, yet accurate interpolation is necessary.

Barmpoutis [20] presented a framework for the reconstruction of a 3D model of the human body. The reconstruction was performed real-time, while the subject moved in front of the camera arbitrarily. The information was captured from a single RGB and depth camera, which the author had chosen to be Microsoft Kinect II. The parametric model used the Cartesian tensor basis and b-spline basis. Due to positive nature of the volume of the human body, a positive-definite tensor spline model was employed to approximate arm, forearms, thighs, legs, torso using an energy-driven model. In this space a Riemann metric was defined to set constraints on the model and make it more accurate. The modelling of head and hands was done by using the RGB stream. The RGB stream was associated with the 3D mesh by using texture mapping. Based on that a skeletal model was formed and the points were divided into body regions. For each body region a tensor spline was estimated by mapping every point p onto a unit circle and then minimising an energy function between the spline and the corresponding point in the circle. Due to skeletal misfits the data varied significantly between frames. To solve the problem a robust energy function was proposed.

The contribution of this paper is the real-time application of the tracking lip movements for 3D head model that can be used in teaching a language. The proposed method benefits from RGB and depth information obtained by Microsoft Kinect II and the extraction and tracking of lip-corners by introducing a new cosine function fitting algorithm. This paper is organised in the following structure. In Section 2 a description of the head-model used is given. In the Section 3 the preliminary work to his research is described and the proposed method for extracting the information about mouth movements is described in detail. In Section 4 the experimental results are presented and discussed and a conclusion is drawn in Section 5.

2. The adopted 3D avatar

The 3D head model constructed by the Institute of Cybernetics in the Tallinn University of Technology has modules that mimic the Estonian speech. The model was designed to synthesise audio-visual Estonian speech based on text. This text-to-speech avatar has been embedded with the connections between different lip and mouth movements and the corresponding speech and text. The model can be seen in Fig. 1.

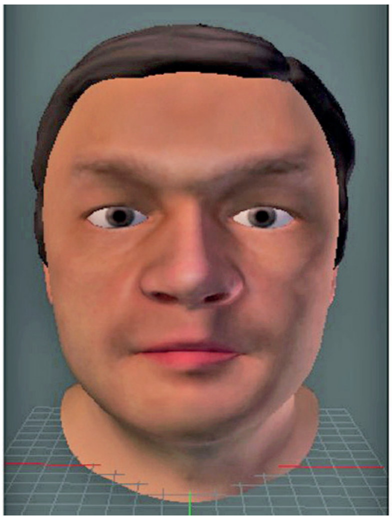

Fig. 1. The 3D head model used in this research

The head model consists of the skin and the structure of 82 bones, including 12 bones of the mouth. In this context bones refer to a collection of vertices that move synchronously. The mouth movements are set to move along with the movement of the chin. This looks quite natural, although the width of the mouth stays the same and the middle point of the upper lip does not change either. In this research in order to make the movement of the lips more natural the changes in upper-lip middle point position and in the mouth corner positions are tracked.

3. The proposed method and configuration

3.1. General description. In this section the setup of the algorithm and the overall scheme is explained. Also all the classical methods used in this research are further explained. The proposed algorithm consists of three steps. First the facial area followed by mouth area is extracted from the frame. Second, the landmark points on the mouth, the nose and the chin are found. During the initialisation process the distances between nose and chin and mouth corners are calculated and a geometric transform between the model and the face is found. For the second step the mouth area is tracked and the important points of the lips, the nose and the chin are found. Third, the avatar is moved according to the changes of mouth shape and chin movements.

A Microsoft Kinect II camera is used in this work because it is a relatively accurate off-the-shelf multimodal acquisition device, making it perfect for research purposes. Since Microsoft Kinect II [21] uses time-of-flight technology for calculation of depth, it is much more accurate than its predecessor, resulting in only 1 mm of error in approximately 1 m range.

3.2. Mouth area detection. A widely used method for separating the face from background is the Viola-Jones algorithm [22]. This algorithm's robustness and high accuracy make it perfect for real-time face detection. In this work the algorithm is adopted with the help of depth information so that the closest person to the visual acquisition system is chosen for further processing. Thus if there is someone moving in the background, they will be ignored. In order to achieve proper and accurate mouth area detection, a cornucopia of constraining parameters, which include narrowing down the area to the lower third of the face, and defining minimum and maximum sizes in relation to the face area is used. Viola-Jones is able to find the mouth area accurately. However, since the focus of the work is on the upper lip, the detected area is translated upwards about one fifth of its height and these results are used for further processing.

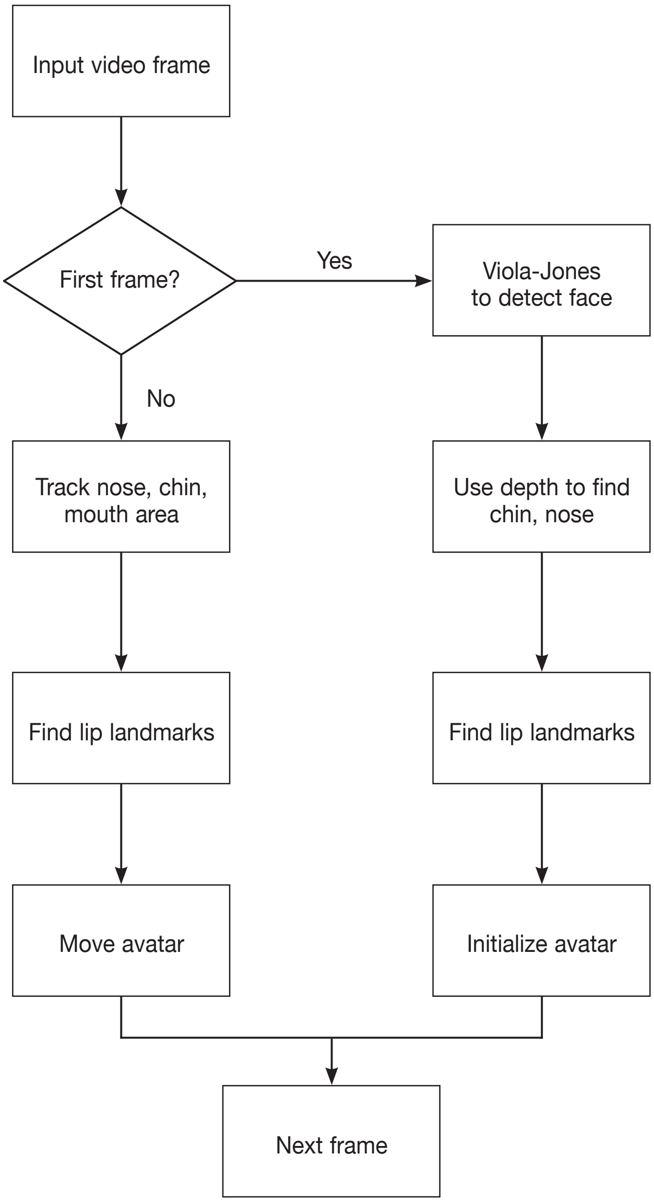

Fig. 2. General scheme of the proposed method

3.3. Important lip point extraction. For the extraction of the upper-lip middle point that is located just below the philtrum, the first Canny edge detector [23] is applied to the mouth region. Canny algorithm is used on grey-scale mouth region image and it consists of five major steps. First, Gaussian filter is applied to avoid false edge detection caused by noise. Second, the intensity gradient of the image is calculated. Third, the non-maximum suppression is used, which means that all the values that are not local maximum are set to zero. Fourth, a double-thresholding step is applied to extract edges and, finally, weak edges and edges that are not connected to strong edges are removed. To make finding the upper-lip line easier, only upper-half of the edge image is used.

Seeing how thoroughly this edge detector normalises the image, it would be difficult to enhance its performance using simple methods such as histogram equalisation, as the same result can be achieved by adjusting the thresholds of this detector. Depending on the chosen parameters either too many edges and noise will be extracted, or there will be a lot of missing data. Since there is no perfect output, in order to automatically choose which edges should be kept and how the missing values should be filled, a cosine function is used.

To easily obtain the location of edge-pixels from the binary output of the edge detector, an algorithm proposed by Suzuki [24] is used. It analyses the topological structures of the image by investigating neighbourhoods of pixels and their connectivity with other pixels. By doing that, one is able to classify all the pixels as background, components or holes and give the outlines of said components as an output.

Extracted contours are merged and upper-lip line is found from the achieved set of points. After that the local minimal between two highest peaks on the line is chosen as the sought after landmark point.

However, these contours cannot be used for mouth corner extraction, because the information changes and is quite often incomplete, failing to capture the entire upper arc. So Shi- Tomasi [25] corner detection method is used. This is a slightly modified version of the Harris corner detector [26]. As a result ,still more corners than needed are obtained and a lot of them are often together in a clutter.

To choose between the appropriate corners and which edges to keep, a cosine function is used. Due to the fact that the upper-lip outline often resembles a cosine, the most fitting one is found based on the output of the contours function. The function is defined by four main parameters: x and y-coordinate of the zero point, period of the cosine and the height of the function (by default it is 1). In order to obtain a computationally easy, but fairly accurate result a moderate number of different values are tried for the cosine and the best one is chosen. In case of closed mouth the corners can be easily estimated by using the darkest line in the middle of the mouth area and finding its intersection with the cosine.

In this work an orthonormal basis transformation is first performed. Let (i, j) be the coordinates of a random pixel, (where i denotes the row and j denotes the column index), (x, y) coordinates of the same pixel in the new system and (x', y') be the zero point of the new axis. Then the coordinates are calculated as follows:

$$x = j - x' \qquad y = y' - i \tag{1}$$

Since the mouth area is not large in size, the cosine fit is achieved by trying out a number of values for parameters and then choosing ones that fit best. The general formula for the cosine is:

$$y' = a\cos(bx) + c \tag{2}$$

where y' is the estimated y-coordinate of the pixel based on the cosine, a is the amplitude of the cosine function, b describes the period of the function and c is the y-axis intersection. The significance of these parameters on the mouth can be stated as a represents the height of the mouth, b indicates the width and c is an additional parameter to achieve a better fit and for the model to be less rigid.

In the fitting process a range of possible values for each parameter is calculated based on the size of the mouth area rectangle. Then a step is calculated for each value and a collection of possible parameter values is formed. In order to find the cosine that is the best fit for the current frame, for each value of x the Euclidean distance between the actual y-coordinate and the predicted y' is calculated:

$$d = |y - y'| \tag{3}$$

If the distance is smaller than a set threshold, which in this work it is set to be one fifteenth of the image height, then the point is considered to be on the cosine and this contributes to the accuracy measure calculation of the current cosine. Based on the intersection points of this line and the cosine, appropriate corners are chosen.

We use this kind of fitting method instead of the mean-square error optimization, because the aim is to find a part of data that the function suits best and fit it, instead of finding the best overall fit for the whole set of datapoints.

However to predict the corners, more information is needed, as cosine is a continuous function and due to the incompleteness of the contours the corners cannot easily be chosen. To fix that the horizontal middle line of the mouth is found.

3.4. Tracking the mouth in the next frame. Even though Viola-Jones algorithm is very fast and accurate when it comes to finding the face, its performance is much worse when used with mouth. Thus to achieve a proper real-time application in this research the depth information is used for tracking. As the nose is always very easily distinguishable in the depth image a simple approach to track the tip of the nose is used. Since this work assumes only little movement of the head, this is enough for a quite accurate estimation of the location of the mouth.

The accuracy of fitting the cosine depends on how small the steps between different values of the parameters are. However trying to fit ten or so values for each parameter can be slow. Also, it is much easier to track a general mouth area, than to track the exact area of the mouth. In this paper the edge information and contours are extracted from each frame. However, no attempt to find the upper-lip line in the contours is made in the tracking step. Instead the cosine is refitted in this new area using a small range of parameters around their values in the previous frame.

3.5. The lower lip. The lower lip is very different from the upper lip, as the transition into the skin is often not very well defined and gradual in essence. Since the avatar has been adapted to move the lower lip very naturally by moving the chin bone, in this work the exact tracking of lower-lip is ignored. Instead for each frame the lower chin point is found by using Microsoft Kinect II. The vertical movement of the chin can be easily calculated, as the transition from chin to neck is easily distinguishable in the depth image obtained from Microsoft Kinect II. In order to keep the proposed algorithm running fast and smoothly in real-time scenarios, only a specific area below the tracked mouth area is prodded for a possible location of the chin. Thus it is easy to obtain the 3D location of the chin from each frame. The location of the bones in the lower lip is further enhanced by the movement of lip corners and reflects the form of the upper lip.

3.6. Moving the model. In this work only the mouth of the model is moved. To easily calculate the changes in the avatar the nose is used as an anchor. In every frame the nose is found as the closest point to the camera in a small area above the mouth. The coordinates are transformed into the domain that corresponds to the avatar, by considering the location of the nose. The upper-lip points are moved in accordance to the changes in the cosine and the lower-lip is moved by moving the chin bone. Figure 3 illustrates the simulation of a speaker's mouth on the avatar.

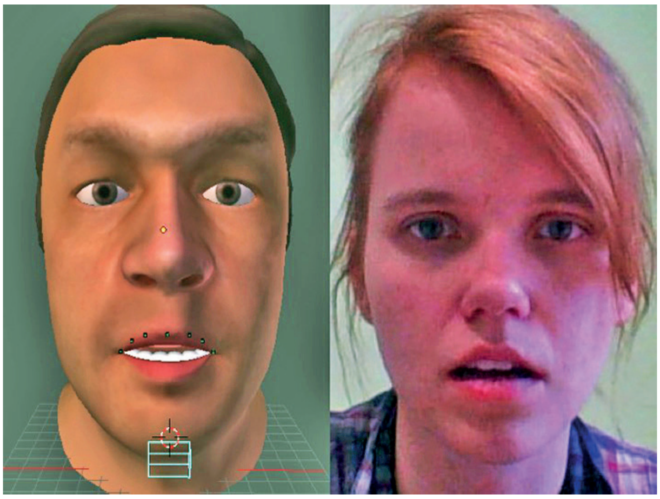

Fig. 3. The mouth position (right) mimicked by the avatar(left)

4. Experimental results and discussions

4.1. Obtaining images. The images used as input in this method were obtained by using Microsoft Kinect II. This camera was chosen due to the possibility of multimodal information. In this paper the RGB and depth information were used. For all the tests the speaker sat approximately a meter from the device and talked straight towards it. The recording was conducted on the controlled environment in order to exclude the effect of illumination variation. In this test dim natural light was chosen, as Microsoft Kinect II camera has a tendency to adjust the brightness of output, sometimes resulting in washed out images. It is important to note that illumination can be always corrected by using available state-of-the-art techniques [27, 28], however in this work illumination has been kept unchanged.

4.2. Initialisation. During the initialisation stage Viola-Jones algorithm finds the face and the mouth area accurately. However this is usually shifted a bit so after shifting it one sixth of the window, the upper arc will be displayed in the images with a much higher probability. The kind of contours extracted from the images is displayed in Fig. 4. These are good representations of the mouth, because they contain most of the information about the upper lip. However these representations become inaccurate if the mouth is opened and teeth are visible. Then the upper-lip contour points become more difficult to extract and often false results are given. The cosines are also displayed in Fig. 4 along with the middle line of the mouth.

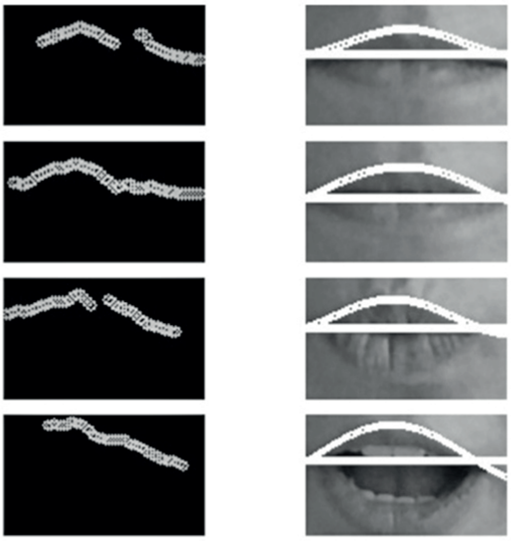

Fig. 4. Extracted contours (left), fitted cosines with middle line of the mouth (right)

To illustrate the naturalness of this method, about 144 samples were recorded of Estonian speaking subject pronouncing nonsense vocal combinations from the language. The results were compared to the output of an accurate 3D scanner that was used to record the same subjects saying the same sentences. However, due to the nature of this 3D scanner the samples could not be recorded simultaneously, thus eliminating the possibility of accurate mathematical comparison between the two. However, similar visual results were achieved by viewing the avatar and plotting the graphs of movements, having very similar patterns and amplitudes. This method has not been compared to any state-of-the art active appearance models, as it is a system with no training set and thus is less robust. In the future the authors wish to adopt an active active appearance model (AAM) system [29, 30] by adding cosine constraint to it in order to improve its accuracy and naturalness.

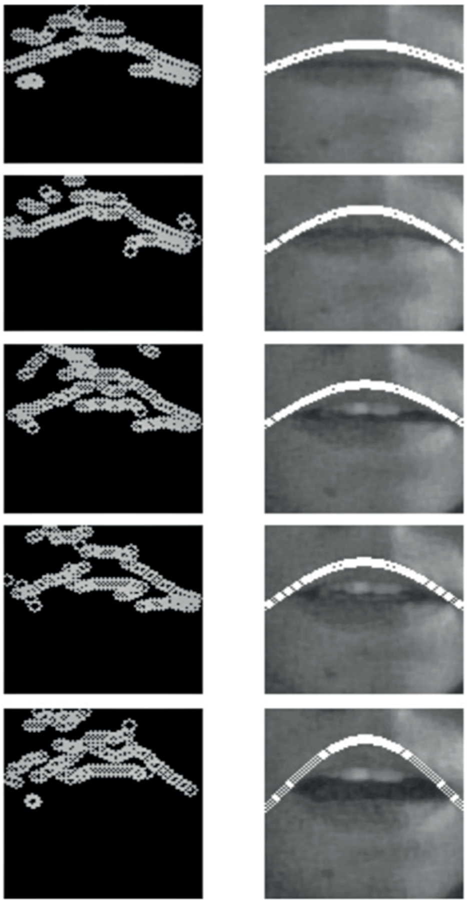

Fig. 5. Contours extracted during tracking (left) and the corresponding cosines (right). These images are from frames 1, 2, 44, 106 and 144, respectively, of a random video

In the case of tracking the cosine fitting becomes accurate for the open mouth too as more information is used. Also the mouth area tracking is very stable as it is adjusted by the location of the nose and the chin. The results of this tracking can be seen in Fig. 5. The images are taken from a video sequence, while the subject was talking at the camera. The fitting error between the estimated cosine function and the manually indicated lip edges highly depends on the way that one can draw such manual line. Thus, introducing any numerical error value can be very subjective. However, the track of mouth movements are very accurate and the avatar mimicked the same shape of mouth in these experimental results.

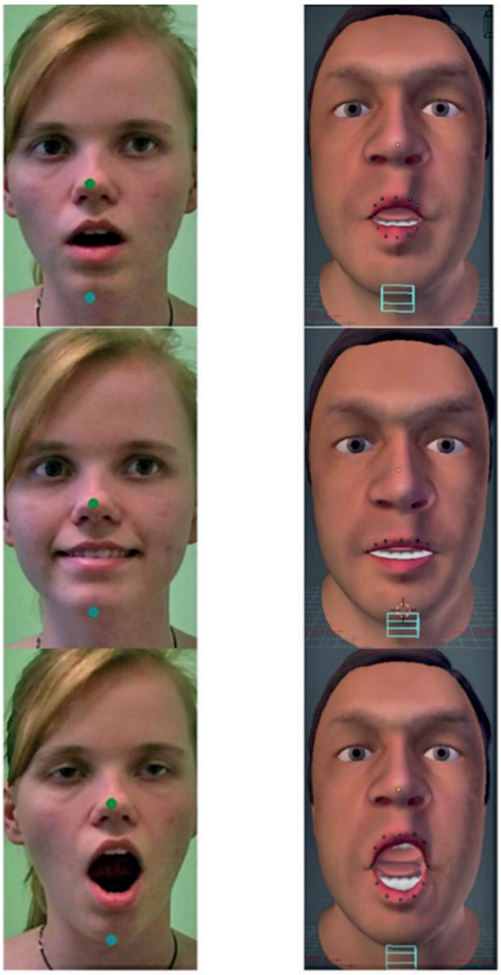

Fig. 6. Face with nose and chin marked (left) and the mouth mimicked on avatar (right)

Using the depth information the tip of the nose can be found very accurately. The location tends to flicker slightly because the depth info is not very stable. However, this problem can be countered by setting simple constraints and thus does not affect the movement of the avatar. The tip of the chin is also extracted using the depth info and in this work only y-coordinate is used, as the x-coordinate is set the same as the nose tip x-coordinate. The nose tips also flickers, but the constraints stop it from showing up in the movement of the avatar. The faces, with nose and chin tips and the corresponding 3D models can be seen in Fig. 6.

The mouth moves extremely well vertically. However, when mouth width changes, the bones rotate somewhat unnaturally, which shows that the model cannot only be controlled by moving the bones around numerically, but they should also be rotated for a more natural mimicking. The solution would be to modify the bone structure of the model and set additional constraints.

5. Conclusion

In this paper a very simple yet effective method was proposed for improving real-time video-based mouth movements of 3D avatar applications. The greatest challenges of an Estonian speaking avatar are set by the unique combination of vocals and sounds in the Estonian alphabet, which raises the necessity for accurate curvature and shape of the mouth. To achieve accurate results the upper-lip RGB and depth information from Microsoft Kinect II camera was processed and a the help of a cosine function was used. By first initialising the model in a relaxed state and later tracking changes, realistic mouth movement was achieved. Microsoft Kinect II is an excellent tool for simple 3D applications, however, it produces different types of depth noise that for better accuracy should be analysed and mathematically modelled. This opens a new challenge by introducing possibility of further research for finding an alternative for Microsoft Kinect II.

Acknowledgements. This work is supported by an Estonian Research Council Grant (PUT638), the Scientific and Technological Research Council of Turkey 1001 Project (116E097), the Estonian Centre of Excellence in IT (EXCITE) funded by the European Regional Development Fund and the European Network on Integrating Vision and Language (iV\&L Net) ICT COST Action IC1307, and the Estonian Centre of Excellence in IT (EXCITE) funded by the European Regional Development Fund. The authors would like to thank the Institute of Cybernetics at Tallinn University of Technology for providing the avatar for conducting the experimental results. The authors also gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU.

REFERENCES

[1] S.-H. Jeng, H.Y.M. Liao, C.C. Han, M.Y. Chern, and Y.T. Liu, "Facial feature detection using geometrical face model: an efficient approach", *Pattern Recognition* 31 (3), 273–282 (1998).

[2] Y. Wang, C.-S. Chua, and Y.-K. Ho, "Facial feature detection and face recognition from 2d and 3d images", *Pattern Recognition Letters* 23 (10), 1191–1202 (2002).

[3] M. Dantone, J. Gall, G. Fanelli, and L. Van Gool, "Real-time facial feature detection using conditional regression forests", *IEEE Conference on Computer Vision and Pattern Recognition,* 2578–2585 (2012).

[4] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T.S. Huang, "Interactive facial feature localization", *Computer Vision–ECCV 2012*, Springer, 679–692 (2012).

[5] J. Wang, R. Xiong, and J. Chu, "Facial feature points detecting based on gaussian mixture models", *Pattern Recognition Letters* 53, 62–68 (2015).

[6] J. Kim and J. Chung, "Untangling polygonal and polyhedral meshes via mesh optimization", *Engineering with Computers* 31 (3), 617–629 (2015).

[7] R.-L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face detection in color images", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (5), 696–706 (2002).

[8] S. Agrawal and P. Khatri, "Facial expression detection techniques: Based on viola and jones algorithm and principal component analysis", *5th International Conference on Advanced Computing & Communication Technologies*, 108–112 (2015).

[9] Y. Wu and Q. Ji, "Learning the deep features for eye detection in uncontrolled conditions", *22nd International Conference on Pattern Recognition,* 455–459 (2014).

[10] I. Lüsi, S. Escarela, and G. Anbarjafari, "Sase: Rgb-depth database for human head pose estimation", *Computer Vision– ECCV 2016 Workshops*, Springer, 325–336 (2016).

[11] K.-H. Seo, W. Kim, C. Oh, and J.-J. Lee, "Face detection and facial feature extraction using color snake", *Proceedings of IEEE International Symposium on Industrial Electronics*, 457–462 (2002).

[12] F. Song, D. Zhang, D. Mei, and Z. Guo, "A multiple maximum scatter difference discriminant criterion for facial feature extraction", *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 37 (6), 1599–1606 (2007).

[13] F. Song, D. Zhang, Q. Chen, and J. Wang, "Face recognition based on a novel linear discriminant criterion", *Pattern Analysis and Applications* 10 (3), 165–174 (2007).

[14] W. Qi, Y. Sheng, and L. Xian-Wei, "A fast mouth detection algorithm based on face organs", *2nd International Conference on Power Electronics and Intelligent Transportation System,* 250–252 (2009).

[15] S. Siatras, N. Nikolaidis, M. Krinidis, and I. Pitas, "Visual lip activity detection and speaker detection using mouth region intensities", *IEEE Transactions on Circuits and Systems for Video Technology* 19 (1), 133–137 (2009).

[16] J. Neyman and E. S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, Springer, 1992.

[17] Y. Xiong, B. Fang, and F. Quek, "Detection of mouth movements and its applications to cross-modal analysis of planning meetings", *International Conference on Multimedia Information Networking and Security*, 225–229 (2009).

[18] A. Caplier, "Lip detection and tracking", *Proceedings 11th International Conference on Image Analysis and Processing,* 8–13 (2001).

[19] R. E. Kalman, "A new approach to linear filtering and prediction problems", *Journal of Fluids Engineering* 82 (1), 35–45 (1960).

[20] A. Barmpoutis, "Tensor body: Real-time reconstruction of the human body and avatar synthesis from rgb-d"*, IEEE Transactions on Cybernetics* 43 (5), 1347–1356 (2013).

[21] L. Yang, L. Zhang, H. Dong, A. Alelaiwi, and A. El Saddik, "Evaluating and Improving the Depth Accuracy of Kinect for Windows v2", *Sensors Journal* 15 (8), 4275–4285 (2015).

[22] P. Viola and M.J. Jones, "Robust real-time face detection", *International Journal of Computer Vision* 57 (2), 137– 154 (2004).

[23] J. Canny, "A computational approach to edge detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 679–698 (1986).

[24] S. Suzuki et al., "Topological structural analysis of digitized binary images by border following", *Computer Vision, Graphics, and Image Processing* 30 (1), 32–46 (1985).

[25] J. Shi and C. Tomasi, "Good features to track", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 593–600 (1994).

[26] C. Harris and M. Stephens, "A combined corner and edge detector", *Alvey Vision Conference* 15, Citeseer, p. 50 (1988).

[27] H. Demirel, G. Anbarjafari, and M.N.S. Jahromi, "Image equalization based on singular value decomposition", *23rd International Symposium on Computer and Information Sciences*, 1–5 (2008).

[28] G. Anbarjafari, A. Jafari, M.N.S. Jahromi, C. Ozcinar, and H. Demirel, "Image illumination enhancement with an objective no-reference measure of illumination assessment based on gaussian distribution mapping", *Engineering Science and Technology* (2015).

[29] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models", *IEEE Transactions on Pattern Analysis & Machine Intelligence* 6, 681–685 (2001).

[30] Y. Chen, C. Hua, and R. Bai, "Sequentially adaptive active appearance model with regression-based online reference appearance template", *Journal of Visual Communication and Image Representation* 35, 198–208 (2016).