Jan Čech
Center for Machine Perception
Department of Cybernetics, Faculty of Electrical Engineering
Czech Technical University in Prague

Master's Thesis Review

Prague, May 20, 2015

Title: Muscle Tension Analysis Based Facial Motion Capture
Author: Andrey Mramornov
Supervisor: Roman Berka, Ph.D.
Date received: May 14, 2015

The thesis presents an algorithm for animating a virtual 3D face, an avatar, by tracking landmark points detected in the image of a real person.

Chapter 1, the introduction, is well motivated and a revision of the related literature is nicely reviewed. The thesis is well structured and the presentation has a good flow. Chapter 2 summarizes a broader background, the face anatomy, type of muscles and their functionality, and the optical flow and Lucas-Kanade tracker. Technical part of the method follows in Chapter 3, where the muscle model and the tracking of the model parameters are described. Implementation details are given in Chapter 4. Experiments are presented in Chapter 5, followed by a conclusion in Chapter 6. A documentation as a guide to the developed software is included as an appendix.

The author completed a compact piece of software that tracks the facial landmarks and animate the virtual avatar, probably in real-time. The solution of modifying Blender code and incorporating Makehuman model was elegant.

However, there are several unclear points that should be clarified during the thesis defence.

A reader would obviously expect a more rigorous formulation of the problem. There should be clearly defined both the direct and the inverse problems. The direct problem is a relationship between the muscle actuation and the 3D face surface deformation. The landmark points detected in the image are projected by a camera in generally unknown pose with respect to the face. The inverse problem is estimating the camera pose and the muscle actuation parameters given the landmarks detected in the image.

The thesis deals with the notion of "muscle model", but it is never explained what it really means. Chapter 3 presents several muscle models where a single 3D vertex is displaced. However, intuitively a muscle controls a set of vertices, it naturally has a certain extent. How does the surface deform in the neighbourhood of the vertex? How is the muscle model bound to the 3D surface? The optimization procedure to find the contraction parameters is also unclear, since it is not described how the landmarks in the image correspond to the muscles or 3D surface points.

Mathematical notation is chaotic. All scalars, vectors, matrices are typeset with the same font. Several collisions of identifiers occurred, e.g. E stands for both an energy and an error statistic, P stands for both a 3D point and a camera matrix. Equations are mostly not numbered.

There is a mistake in the description of LK-tracker, in Sec. 2.2.1. The LK-tracker is an iterative procedure, where the brightness-constancy equation is linearised in every iteration, while the section described a single iteration only. Moreover, the tracking of the landmark points by a general LK-tracker is rather naive and clearly sub-optimal. Recently, there have

been reliable detectors/trackers that estimate facial landmarks in super real time, e.g. [1, 2, 3, 4]. Note that [3, 4] provides a 3D estimate of landmark points, which would probably ease the estimation of the muscle contraction parameters, would not it?

Experiments attempt to evaluate accuracy of the algorithm. However, it was not explained how the ground-truth muscle contraction was obtained, Eq. (14). This point should be clarified. Moreover, Chapter 5 only shows several plots without any discussion. Which units are of the optical flow error in Fig. 17? Which units are of the error in Fig. 22? Statistical significance of results on literally two subjects is likely not very high.

I miss a better qualitative demonstration of the results. I would strongly appreciate a video demonstrating the performance of the method. Simply, display side-by-side the input video of a face with tracked landmarks and the output rendering of the avatar, similar to Fig. 26 and Fig. 27. I would like to ask the author to prepare the video for the defence. The author cannot expect a reviewer installing and compiling all his software with several dependencies to get a quick intuition how the method performs.

The novelty of the work is not discussed and remained unclear. A reference to the papers recommended by the thesis advisor is missing. Why? The papers of Choe et al. are closely related indeed.

The thesis would benefit from proofreading. The text contains several language mistakes.

Considering all above, I suggest evaluating the thesis as

C – good.

Jan Čech, Ph.D.

References

[1] S. Ren, X. Cao, Y. Wei, and J. Sun, "Face alignment at 3000 FPS via regressing local binary features," in *Proc. CVPR*, 2014.

[2] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proc. CVPR*, 2014.

[3] A. Asthana, S. Zafeoriou, S. Cheng, and M. Pantic, "Incremental face alignment in the wild," in *Proc. CVPR*, 2014.

[4] J. Cech, V. Franc, and J. Matas, "A 3D approach to facial landmarks: Detection, refinement, and tracking," in *Proc. ICPR*, 2014.