An unorthodox view on the problem of tracking facial expressions

Magdalena Błażek, Maria Kaźmierczak
University of Gdansk
Bażynskiego Str. 4, 80-952 Gdańsk, Poland
Email: psymb@ug.edu.pl, Email: psymk@ug.edu.pl

Artur Janowski
University of Warmia and Mazury
in Olsztyn
Michała Oczapowskiego Str. 2, 10-719 Olsztyn, Poland
Email: artur.janowski@geodezja.pl

Katarzyna Mokwa, Marek Przyborski, Jakub Szulwic
Gdansk University of Technology
Gdańsk, Narutowicza Str. 11/12, 80-233 Gdańsk, Poland
Email: katarzyna.a.mokwa@gmail.com, Email: marek.przyborski@pg.gda.pl, Email: jakub.szulwic@geodezja.pl

Abstract—Recent developments in imaging cameras has opened a new way of analyzing facial expression. We would like to take advantage from this new technology and present a method of imaging and processing images of human face as a response to the particular stimuli. The response in this case is represented by the facial expressions and the stimuli are still images representing six basic emotions according to Eckmann. Working hypothesis of presented research, states that the new method of tracking facial expressions is more precise and distinctive enough to give characteristic description of the analyzed human face. The biggest advantage of the presented method, in the opinion of research team, is the fact that it uses remote sensing techniques and presents dynamics of the changes happening on the human face. Therefore, FMRI might not be required, which decreases the costs of experiments, additionally, method is less stressful for the examined persons and provides more natural reactions.

I. Introduction

SCIENTISTS all over the world are looking for new methods of tracking facial expressions to use in the field of psychology, neuroscience and affective computing. Emotion influence all modes of human activity, communication, interpersonal relations, family and business life. There are lots of empirical models of affect. Basically they are focused on analysis of speech, visual and biophysiological signals (see for example [1], [2], [3], [4], [5]. In this paper we would like to focus attention on visual methods. Typically, these methods are based on comprehensive description of the changes in facial expressions, discernment of action units for all visually distinguishable facial movements and mapping them with basic emotion (based on Facial Action Coding Systems [6], [7]). Characteristic points, located in specific segments of the human face (mouth, eyebrow, eyes etc.) are tracked and their movement is identified as a certain type of emotion [8], [9], [10]. However in our opinion, emotions represents extremely complicated states of the human's brain and they have rather fuzzy nature, thus tracking changes in the position of only few characteristic points might not be effective. Due to the nature of the observed phenomenon there is no simple method of

distinguishing certain type of emotion among diversity of emotional states. Fig. 1 shows how complicated are representations of emotional states when they are expressed by human beings. Taking into account the movement of the lips or eyebrows we are not able to distinguish in what emotional state the tested person is. In addition, presenting on Fig. 1 images are taken when the emotion is the most intensive (for example the moment when the smile is the widest and brightest). We have decided to create our own database of facial expressions because we would like to investigate the process of creation facial expression from the beginning to the end (when the face comes to neutral state). Data gathered in the new database let us track changes on the face throughout the period of observation.

The caption of the Fig. 1 is a starting point for the further investigations. Humans ability to recognize different emotions is very difficult and complicated process to implement in the machine like computer. Recognizing emotional states is important for human-computer interaction or other form of coexisting in an artificial environment. In the opinion of the research team., dynamics of the facial expressions is the key to find the efficient way of recognizing emotional states. The main goal of presented experiment was to develop a new method of tracking changes in facial expressions, because, in the opinion of research group, it might be a promising direction to develop fully automatic method of recognition emotional states. Proposed method opens a long way to find the solution of the problem how emotions are expressed and how they arise.

Recent developments in imaging cameras has opened a new way of analyzing facial expression of emotions [11]. We would like to present a new method of imaging and processing images of human face in order to quantify the response to the particular stimuli [12]. This article presents a method of processing images of human face recorded by a fast camera Phantom MIRO 310. Method was applied on a collection of 65 persons (30 female students and 35 male students) expressing happiness, sadness, anger, fear and disgust, as well as surprise.
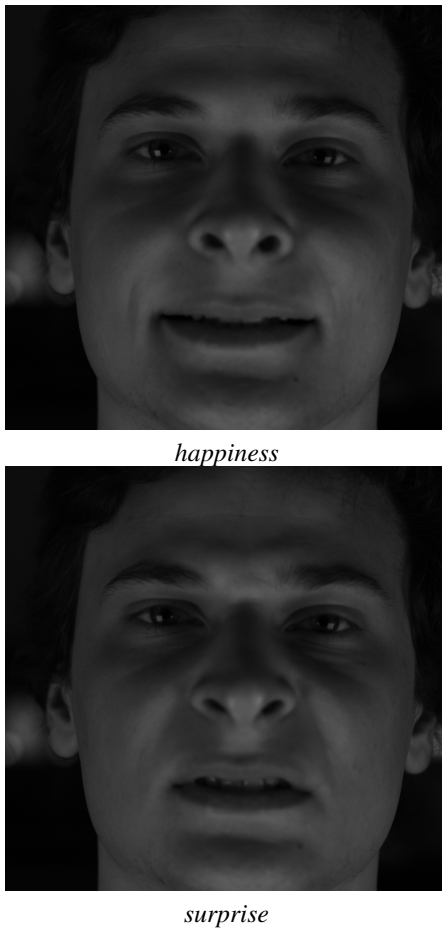
happiness

surprise

Fig. 1. 2 of 6 basic emotions expressed by research participant, it is noticeable that there are no clear limits between emotional states.

Fig. 2. Single object trajectory, $Z - coord$ - represents brightness
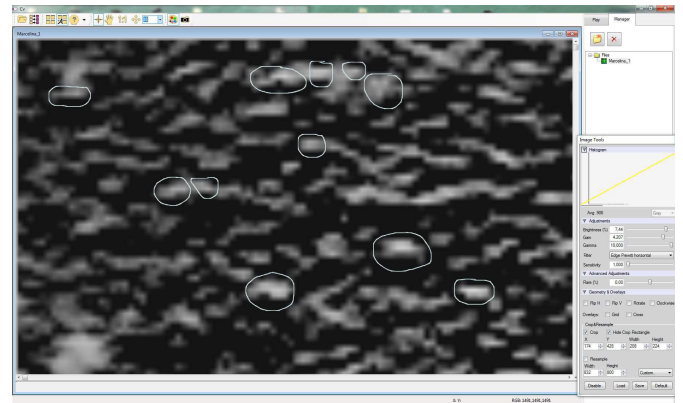
Fig. 3. Objects on the human face, characteristic for the certain structure of the skin

Each emotion was expressed be the students for about 2 seconds. Resulting images represent series of frames recorded at the speed of 1000 frames per second. Presented method is based on the assumptions coming from Particle Image Velocimetry (PIV) [13], [14]. The most important one is that there are some objects on the human face, characteristic to the certain type of skin (see Fig. 4) (as well as certain mechanical and chemical properties of the skin), that are determined and identified while processing series of images. Those objects are of non-spherical shape (as it is presented on the Fig. 3). Working hypothesis of presented research states that the facial expressions are determined by the mechanical and chemical properties of the human's skin and the tension of the muscles, however all those properties might influenced the motion of the identified objects only in very limited range, as it is presented on the Fig. 2.

The following sections are the attempt to the answer the question whether the working hypothesis is true or false.

The main idea of proposed research comes from the methods of studies colloidal particle's trajectories and particle image velocimetry. While both image processing, photographic, study of colloidal suspensions and particle image velocimetry
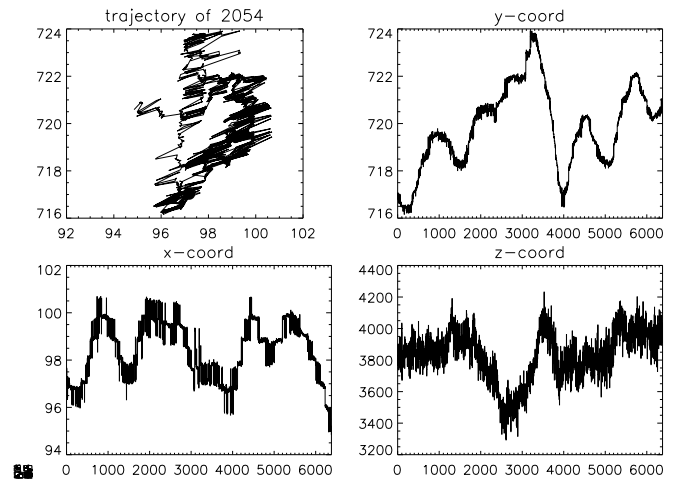
are well developed fields, applying them to the field of affective computing and psychology gives new view on the problem of facial expressions what in consequence might constitute new approach to the problem of automatic recognition of human emotions [15], [2]. The methods described below are generalized to the objects or areas of non-spherical shape that can be identified on human face (see Fig. 3). Shape and position of non-spherical objects located on the face plays very important role in this method. Thanks to the speed of the recording (1000/frames per second) collected positions of the identified objects create trajectories of their movement.

In the following, section typical instrumentation required for collecting digital video images of human face has been described, and some details of the steps required to convert a digital movie into an ensemble of single-object trajectories. We stress those aspects of the analysis which allow us to track particular objects on the human face. High-resolution trajectory data makes possible a wide range of quantitative measurements of the changes occurring on human face.

oryginal image

processed image

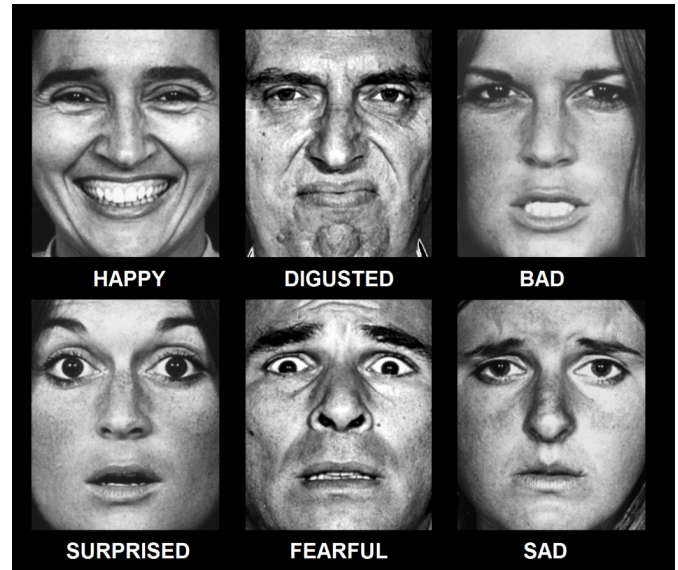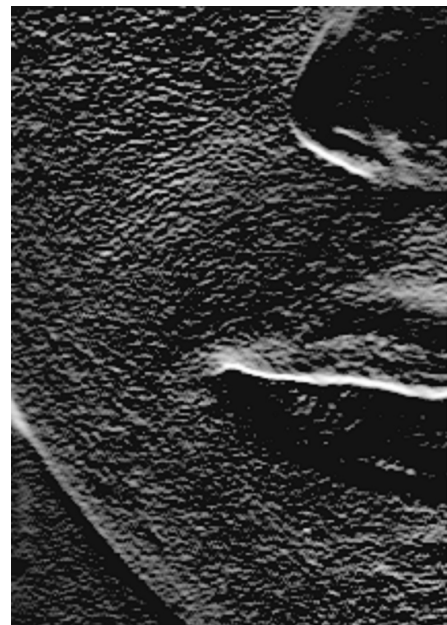Fig. 4.   An example of recorded frames of human face

Fig. 5.   Standard representations of the basic emotions

expensive but also because they tend to have superior noise figures and greater sensitivity to subtle brightness variations. Color information, furthermore, is not used in the techniques we describe below. We use an Phantom Miro 310 camera connected to the Dell m4700 mobile workstation. To apply Particle Images Velocimetry method we have to transform an image of human face in to map of small objects like it is presented on Fig. 6

II. METHODS

Research was conducted in the Gdansk University of Technology, Remote Sensing Lab on 30 female and 35 male, students of the Psychology at the Gdansk University. The study was based on the assumption that students have to reproduce emotion which was showed them on the laptop's screen. Our team has prepared PowerPoint presentation based on the set of facial expression images proposed by Professor Eckman (see Fig. 5). Each emotion was earlier presented to the examined person and then on the mark he or she should reproduced it straight to the camera. It seems necessary to stress that the stimuli - picture of certain emotion wasn't active when examined person was asked to reproduced the image, because the eyes of the research participant were directed straight at the camera, not on the computer screen with the image of emotion pattern. In fact, only the imagination of the image is used to render the emotion.

Standard commercial video cameras produce 30 complete images per second but in order to track changes on the human face we need speed camera with ability to record at least 1000 complete images per second. Monochrome CCD cameras may be preferable to color models not only because they are less

map of small objects

Fig. 6.   Zoom of the ROI from Fig. 4

At this stage of the processing the filtering methods are very helpful, in this particular case Prewitt filter has been used, which was implemented in the Phantom Miro 310 camera software. The following parameters has been used in this experiment:

- brightness of the image set to 7.44,
- gain factor set to 4.207,
- gamma factor set to 10.00.

By using camera software it was possible to collect series of frames from the recorded movie already transformed into a map of small objects, and represents them as a series of tiff images directly from the camera without loosing the time for additional computations and transformations, an example is presented on the Fig. 6. After this operation a series of images, about 2000 frames for each of 6 basic emotions, have been created. Every image represents face as a map of small objects that move while facial expression (typical for the certain type of emotion) come out on the face (see Fig. 6).

Digital video analysis and methods of tracking and identifying objects, enables extracting trajectories of individual micro-objects from a series of images. The time evolution of the distribution of interested objects (or areas) can be calculated according to the following equation

$$\theta(r,t) = \sum_{i=1}^{N} \delta(r - r_i(t)),\tag{1}$$

where:

- $r_i(t)$ is the location of the i-th object in a field of N objects at time t,
- δ - distance.

III. RESULTS.

The process of extracting useful information from a sequence of digital images consists of five logical steps:

1) correcting imperfections in the individual images,
2) locating candidate particle positions,
3) refining these positions,
4) discriminating „false" objects,
5) and finally linking the time-resolved object locations into trajectories.

In this case, the tracking algorithm has to deal with the changing appearance of the interested areas as they move. Furthermore, it has to distinguish marginally focused particles from noise. In practice, the image $A(x,y)$ must be cast from an array of bytes to a higher precision data format, such as a floating point array, before convolution. We have used an object oriented method for identifying candidate object within an image as a local brightness maxima's. The pixel is assigned as a candidate if no other pixel within a certain distance w is brighter. Because only the brightest pixels correspond to the location of interested us objects on the human face, we further require candidates to be in the upper 50th percentile of brightness for the entire image, see Fig. 7.

This operation is called, the gray-scale dilation [16], [17] and provides an implementation of the regional maximum
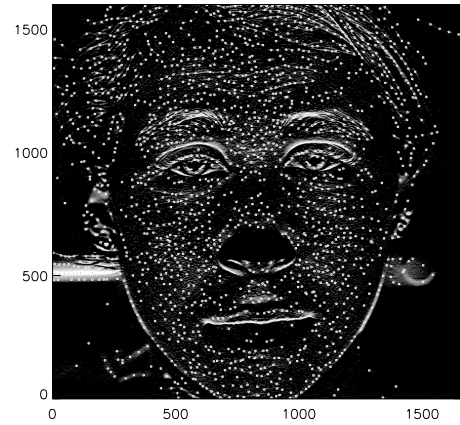
Identified objects
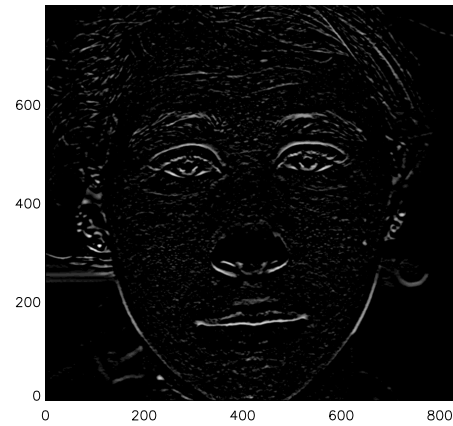
Objects of the same brightness

Fig. 7.   Non-spherical areas (objects or particles) on the human face after processing. Objects represents the same spectral feature - brightness

selection criterion. Gray-scale dilation is an elementary morphological operation which sets the value of pixel $A(x,y)$ to the maximum value within a distance w of coordinates (x,y). A pixel in the original image which has the same value in the dilated image is then a candidate particle location. We use the same value of w as was used in the filtering step.

Having already found a locally brightest pixel at (x,y), which presumably is near the geometric center of an interested us object's at the coordinates x_0, y_0 , we calculate the offset from (x,y) to the brightness-weighted centroid of the pixels in a region around (x,y) according to following equation:

$$\begin{pmatrix} \epsilon_x \\ \epsilon_y \end{pmatrix} = \frac{1}{m_0} \sum_{i^2+j^2 \leq w^2} \begin{pmatrix} i \\ j \end{pmatrix} A(x+i, y+j),\tag{2}$$

where:

- moment $m_0 = \sum_{i^2+j^2 \leq w^2} \binom{i}{j} A(x+i, y+j)$ is the integrated brightness of the sphere's image.

Having located objects or particles in a sequence of images, the next step is to match up locations in each image with corresponding locations in later images to produce the trajectories. This requires determining which particle in a given image most likely corresponds to one in the next image. Tracking more than one object is very difficult. Thus, it is necessary to seek the most probable set of N identifications between N locations in two consecutive images. Linking particle distributions into trajectories is only feasible if the typical single particle displacement in one time step is sufficiently smaller than the typical inter-particle spacing, otherwise, particle positions will become inextricably confused between snapshots. In this case the speed of recording (1000/frames per second) let us discriminate very small displacement between objects, and thus the trajectories are more accurate and precise. This process is repeated for the particle locations in each frame until is completely determined. The examples of calculated trajectories are presented on Fig. 8.

One of the most important aspects of particle dynamics in the fluid flow, which underlies the proposed method, is the mean squared particle displacement while the time of observation evolve. This quantity can be calculated according to the Eq. 3. On the Fig. 9 we would like to present results of calculating msd for the few members of the research group. The $Panel$ A represents MSD $curves$ calculated for 6 basic emotions reproduced by participant according to the images showing on the laptop's screen.

$$MSD(\tau) =< \triangle r(t)^2 >=< [r(t+\tau) - r(t)]^2 > \quad (3)$$

where:
- $r(t)$ particle position at time t,
- τ lag time between two positions,
- operator $<>$ designates a time-average over t and/or an ensemble-average over several trajectories.

Figure 9 $Panel$ B presents the examples of calculating MSD $curves$ for selected number of representatives of the research group. Each curve for each person, is an average of 6 basic emotion presenting on the $Panel$ A. In order to keep the figure as readable as possible only 3 MSD-curves are drawn. The most important conclusion from this figure states, that it is possible to see differences between particular curves (Fig. 9 $Panel$ B). The second conclusion - shape of the MSD is similar for examined persons. However, in our opinion, it depends on the individual features of human skin. As it was mentioned in the Introduction in our opinion, mechanical and chemical properties of the skin plays an important role in this phenomenon. The next conclusion is, that noticeable differences between different emotional states or different person starts from about $t = 100$ of time of observation. While watching video of recorded persons it is noticeable that the facial expression appears after a few moments of inactivity. This behavior has its own mark in the results of calculating MSD. And the last conclusion which in opinion of the research group seems to be very important, especially

in the field of neural science and psychology, there is some activity on the human face even before the facial expression appears (at the Fig. 9 segment from $t = 0$ till $t = 100$). We believe that human brain express the certain emotion before the face is able to showed it. This conclusion may acknowledge recent discoveries of Professor Laeng in the field of imagination and perception.

IV. CONCLUSION

The preceding sections describes image analysis methods we have developed to perform quantitative time-resolved imaging studies of non-spherical objects (or areas) on the human face which we have found useful in the process of tracking facial expressions.

We have shown that the trajectories of the identified areas on the human face are distinctive enough to find differences between persons taking part in the research.

The measurable effect of the presented in the article transformations of the collected series of images is the collection containing the identification number of the identified areas (same size and brightness) on the face and the trajectories of their movement throughout the observation period.

The study of that phenomenon is based on the assumption that there are objects on the human face and they movement depend on the muscles tension, mechanical and chemical properties of the skin. Presented method may open the way to a much larger and broader question of the automatic identification of emotional states. It gives the possibility to find quantitative description of the observed changes of emotional states happening on human face [18], [19] what may have big value for the security systems.

Recent research on the connections between imagination and perception suggest that unconditional reflex narrowing of the pupil and caused by changes in the amount of light reaching them, can be elicit also by only imagining the appropriate situation. Usually people treat ideas as private and subjective experience, which is not accompanied by significant physiological changes. The results achieved by the Professor Laeng team and presented in [20] challenge this view. They suggest that imagination and perception are based on similar sets of neural processes. Presented in this article experiment on tracking facial expressions also confirms their thesis.

REFERENCES

[1] R. W. Picard, S. Member, E. Vyzas, and J. Healey, "Toward Machine Emotional Intelligence : Analysis of Affective Physiological State," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.

[2] D. Ghimire and J. Lee, "Geometric Feature-Based Facial Expression Recognition in Image Sequences Using Multi-Class AdaBoost and Support Vector Machines." *Sensors (Basel, Switzerland)*, vol. 13, no. 6, pp. 7714–34, Jan. 2013. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3715259\&tool=pmcentrez\&rendertype=abstract

[3] G. Guo, R. Guo, and X. Li, "Facial Expression Recognition Influenced by Human Aging," *IEEE Transactions on Affective Computing*, vol. 4, no. 3, pp. 291–298, Jul. 2013. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6517179

[4] D. Heylen, M. Ghijsen, A. Nijholt, and R. D. Akker, "Facial Signs of Affect During Tutoring Sessions," pp. 24–31, 2005.
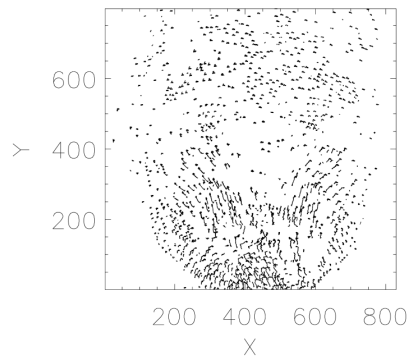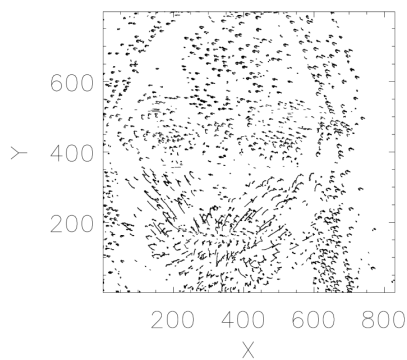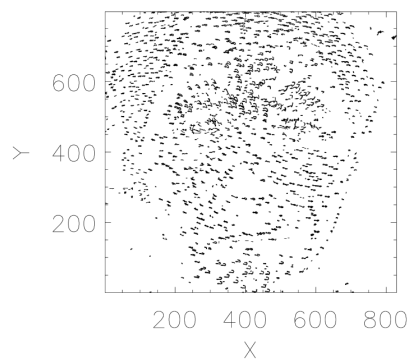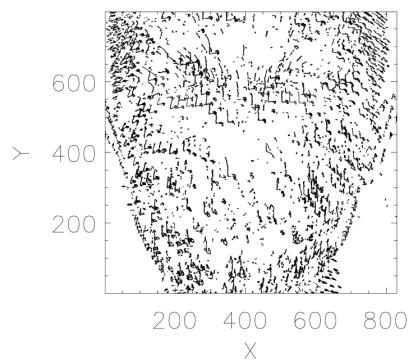
Person 1

Person 2

Person 3

Person 4

Fig. 8. Trajectories of the identified objects on the faces of 4 representatives tested group. Each person reproduced the same emotion - happiness
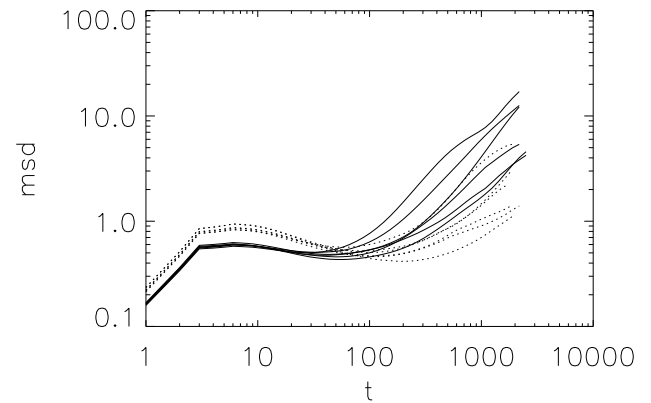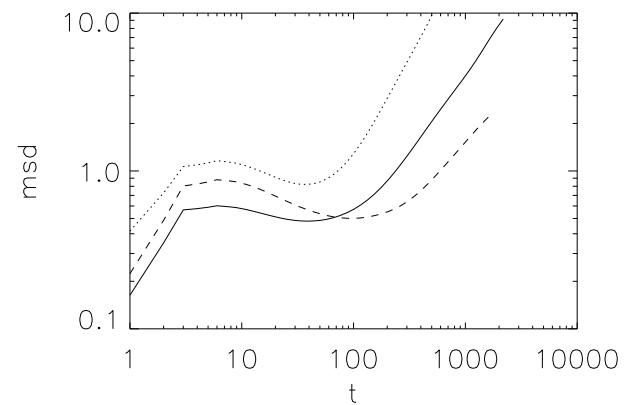

Panel A

Panel B

Fig. 9. Mean squared displacement.

[5] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "ECG Pattern Analysis for Emotion Detection," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 102–115, Jan. 2012. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5999653

[6] P. Ekman and J. C. Hager, "Computer Measurement of Sign Vehicles in Body Movement and Facial Expression." [Online]. Available: http://face-and-emotion.com/dataface/misctext/iwafgr.html

[7] P. Ekman, W. V. Friesen, and J. C. Hager, "Facial Action Coding System - Title Page," 666 Malibu Drive, Salt Lake City UT 84107, Tech. Rep., 2002. [Online]. Available: http://face-and-emotion.com/dataface/facs/manual/TitlePage.html

[8] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, Jan. 2003. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0031320302000523

[9] M. Pantic and M. S. Bartlett, *Machine Analysis of Facial Expressions*, 2007, no. June.

[10] T. Pfister, X. Li, G. Zhao, and M. Pietik, "Recognising Spontaneous Facial Micro-expressions."

[11] R. C. Gur, R. Sara, M. Hagendoorn, O. Marom, P. Hughett, L. Macy, T. Turner, R. Bajcsy, A. Posner, and R. E. Gur, "A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies." *Journal of neuroscience methods*, vol. 115, no. 2, pp. 137–43, Apr. 2002. [Online]. Available:

http://www.ncbi.nlm.nih.gov/pubmed/11992665

[12] T. Balomenos, A. Raouzaiou, S. Ioannou, A. Drosopoulos, K. Karpouzis, and S. Kollias, "Emotion Analysis in Man-Machine Interaction Systems," in *in Proc. MLMI, LNCS 3361*, 2005, pp. 318–328.

[13] M. Raffel, C. E. Willert, and J. Kompenhans, *Particle Image Velocimetry: A Practical Guide ; with 24 Tables*, 1998. [Online]. Available: http://www.google.pl/books?hl=pl\&lr=\&id=enOLTmfYVPQC\&pgis=1

[14] A. Schroeder, C. E. W. Eds, D. P. J. Barz, H. F. Zadeh, and P. Ehrhard, *Particle Image Velocimetry*, topics in ed. Springer - Verlag GmbH, 2008.

[15] Y. Cao, W. Zheng, L. Zhao, and C. Zhou, "LNCS 3784 - Expression Recognition Using Elastic Graph Matching," *Lecture Notes in Computer Science*, vol. 3784, pp. 8–15, 2005.

[16] A. K. Jain, *Fundamentals of digital image processing*, 1989. [Online]. Available: http://books.google.pl/books/about/Fundamentals_of_digital_image_processing.html?id=GANSAAAAMAAJ\&pgis=1

[17] W. K. Pratt, *Digital Image Processing*. New York, USA: John Wiley & Sons, Inc., 2001. [Online]. Available: http://doi.wiley.com/10.1002/0471221325

[18] R. W. Picard, "Emotion Research by the People, for the People," *Emotion Review*, vol. 2, no. 3, pp. 250–254, Jun. 2010. [Online]. Available: http://emr.sagepub.com/cgi/doi/10.1177/1754073910364256

[19] J. Tao, T. Tan, R. Picard, A. Choi, and W. Woo, *Affective Computing and Intelligent Interaction*, ser. Lecture Notes in Computer Science, A. C. R. Paiva, R. Prada, and R. W. Picard, Eds. Springer Berlin Heidelberg, 2007, vol. 4738, no. September. [Online]. Available: http://www.springerlink.com/index/10.1007/978-3-540-74889-2

[20] B. Laeng and U. Sulutvedt, "The eye pupil adjusts to imaginary light." *Psychological science*, vol. 25, no. 1, pp. 188–97, Jan. 2014. [Online]. Available: http://pss.sagepub.com/content/25/1/188