社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report

Detection and Measurement of Facial Micro-Expression Characteristics for Psychological Analysis

Senya Polikovsky, Yoshinari Kameda and Yuichi Ohta

Department of Intelligent Interaction Technologies, Graduate School of Systems and Information Engineering, University of Tsukuba.

E-mail: senya@image.esys.tsukuba.ac.jp, kameda@iit.tsukuba.ac.jp, ohta@iit.tsukuba.ac.jp

Abstract Detection of facial micro-expression is considered to be useful for recognition of human's intention, especially for hostile intent and danger demeanour. In order to promote the psychological analysis on facial micro-expression, we propose a new computer vision method of measuring the facial micro-expression. By combining the proposed method with high speed camera of 200fps, we succeeded to estimate specific characteristics of facial micro-expression that could help a further progress of its psychological analysis. We evaluated the proposed classification method over 10 subjects and confirmed its performance. In addition, this paper describes a new experimental setup, that will estimate the exact location and orientation of the face at 200 fps time resolution and a new GUI interface, which is useful to acquire efficient ground truth tagging of micro-expressions from the recorded videos.

Keyword Hostile intent detection, Facial motion analysis , 3D Gradient Descriptor, High speed camera

1. Introduction

Recent extreme actions such as terrorist attacks around the world have motivated us to search for new technological solutions that can help to detect and prevent such actions. Computer vision combined with psychology has potential for developing such a solution.

After thirty yeas of research by Ekman, Frank and O'Sulliva [1] in addition to in depended group of Portet [2] micro-expressions were found to be an essential behavioural source for detecting deception and can be used for danger demeanour detection as well [1].

Facial micro-expression is a brief, involuntary expression shown on the face of humans when they are trying to conceal or repress an emotion. Micro-expressions usually occur in high-stakes situations, where people have something to lose or gain [3].

From the technical point of view the detection of facial micro-expressions is not an easy task using the traditional approaches. Their duration is 1/3 to 1/25 seconds, and they appear with low muscle intensity. The need for analyzing such momentary expressions requires a use of a high-speed camera.

These factors call for a use of new 3D gradient descriptor on predefined face regions. The location of facial regions was defined by following facial action coding system [4] introduced by Ekman.

This paper presents a new approach for facial micro-expression recognition by using high-speed camera. Based on the physiological characteristics of the facial muscles 200fps camera was selected to insure that the fastest facial motion appears in at least 10 frames. This is the sufficient time resolution for detecting of the micro-expression.

For initial testing we prepared a simple database containing paused micro-expressions of 10 participants. In this paper we shall describe new experimental setup that will allow creation of more sufficient video testing datasets. This procedure is based on psychology interrogation protocol that speeds up the process of creating more realistic and sufficient datasets. For increasing the accuracy of ground truth extraction, we developed a new GUI for FACS tagging.

The structure of the paper is as follows. In Section 2 we discuss related work. Suggested algorithm is described in Section 3. Section 4 presents our video dataset that was used for the experiments. A new experimental setup for facial micro-expressions collection is presented in section 5. Experimental results are discussed in Section 6. We conclude the paper in Section 7.

2. Related work

Before discussing the detection and measurements of facial micro-expression we review the theory and recent progress of expression analysis.

The process of automatic facial expression analysis consists of three main phases: (1) face acquisition, (2) facial data extraction and representation, and (3) facial expression recognition.

Face acquisition targets an automatic face detection and face tracking in the video. Extraction of 3D localization of the face is also a part of this phase.

Geometric feature-based and appearance-based are two main approaches that are used for facial data registration.

With geometric based method, facial components and facial feature points such as lips corners, centre of the eyes , eyebrows edges, tips of the nose and etc. are extracted by computer vision techniques. The coordinates of there points form a feature vector that represents the face geometry [5].

Recently, superior research results were reported on Active Appearance Model (AAM) by Kanade [6] group.
Two problems in using of AMM are: One is the need for extensive training dataset with large amount of manually tagged facial feature points. The other is the degradation in tracking results on the faces that were not included in the training set.

Another approach is based on direct tracking of 20 facial feature points (e.g. eye and mouth corner, eyebrow edges) by particle filter [7]. In more recent work [8] a relationship model between micro-expression was added to track them. This approach delivers good results for some facial motions, but fails in detecting subtle motions, which can be detected only by observing skin surface. The performance there approaches strongly rely on the accuracy of the facial feature points tracking. In practice, facial feature points tracking algorithm can not deliver the suffician recognition accuracy for large number of micro-expression.

In appearance-based methods, variable-intensity templates are exploited to describe the changes in the intensity of multiple pairs-of-points in the vicinity of predefined facial parts such as eyebrows, eyes, nose and mouth during facial expressions [9]. The use of this approach for micro-expressions detection invokes the problem of defining the "right" distance between the pairs-of-points.

Another approach exploits on Gabor wavelets filter, that is applied to either the entire face region in the image or specific regions in the face, creating vector of filter response as descriptor of facial state. This method was applied for spontaneous facial motion analysis and considered to be the state of the art approach [10] However, this method estimates the face state frame by frame, without taking into account the correlation and transition between frames.

This paper focuses on effective extraction, recognition, measurements and analysis of facial micro-expression. In this work we skip face detection and face tracking step however the proposed algorithm can be easily combined to the state of the art of face tracking algorithms . For now we assume that examined frontal faces are located relatively in the same place. More detailed description of data capture will be presented in dataset section.

The 3D gradient oriented histogram descriptor was chosen for facial motion detection due to its ability to capture the correlation between the frames [20].

3D gradient descriptors were proved to be effective approach for classifying motions in video signal [11]. In Dollar work [12] different local descriptors such as normalized pixel values, brightness gradients, and windowed optical flow were evaluated and compared for action recognition. Those descriptors, however, were computed by simple concatenation of all gradient vectors in a region.

Scovannere [13] proposed more advanced descriptor by extending the SIFT descriptor form 2D to 3D vectors. If a video cube is defined as v(x,y,t), spatio-temporal gradients of each pixel are computed as $\delta v_x(x,y,t)$, $\delta v_y(x,y,t)$ and $\delta v_t(x,y,t)$ respectively. Then by using (1) vectors magnitude and angles are calculated.

$$\begin{aligned}
m_{3D}(x,y,t) &= \sqrt{\delta v_x^{\,2} + \delta y_y^{\,2} + \delta y_t^{\,2}} \\
\theta_{3D}(x,y,t) &= \tan^{-1}\left(\delta y_y^{\,2} / \delta y_t^{\,2}\right) \\
\phi_{3D}(x,y,t) &= \tan^{-1}\left(\delta y_t^{\,2} / \sqrt{\delta v_x^{\,2} + \delta y_y^{\,2}}\right)
\end{aligned} \tag{1}$$

All pixels vote into the spherical grid of histograms of oriented gradients and a polar coordinates are used for computing the orientation quantization. This leads to the problems of singularities at the poles since bins get progressively smaller in the top and bottom part of the sphere.

Polyhedrons can be used as solution for the singularity problem. Histogram computation is done by projecting gradient vectors onto the axes running through the centre of the polyhedron face centres [11].

The methods of [11] and [13] can be considered as general descriptor though they miss the physical connection between the values of the descriptor and the actual facial movements.

Our proposing descriptor can be viewed as more specifically adapted for facial movements and it allows us to examine the facial movements through observing gradient's histogram.

3. Detection of Facial Micro-Expressions

The algorithm of our micro-expressions recognition detection is presented in this section. In order to capture videos of micro-expressions we use high speed camera .

The algorithm has three steps: First, division and extraction of twelve facial video cubes (section 3.1). Second, computation of the 3D gradient orientation histogram descriptor for each video cube (section 3.2). In the last step all the descriptors are classified for indentifying micro-expressions (section 3.3).

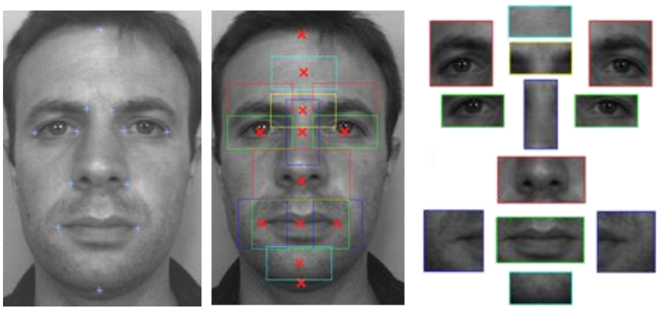

Figure 1: a) Manually selected point on the first frame. b) Calculated region centers, c) Selected facial regions

3.1. Facial features regions

Structure and movements of facial muscles are complicated. Therefore we divide the face to 12 regions. On following the facial action coding system (FACS) [4] that decomposes facial expressions in terms of 46 component movements, we select the most representative regions in term of facial micro-expressions. (FACS is the most widely used expression coding system in the behavioral sciences.)

In order to simplify the classification step each facial region is carefully set in a way that its appearance would be influence only by a limited number of muscles.

Hence during the classification step, every region will hold a limited number of classes to distinguish. The regions are set manually using the following procedure: 12 points are marked on the first frame of the input video manually (Figure 1: a). The positions of the facial points on the face were defined in [6] and [7] work. The selected point are used to locate all the regions of interest on the face. Then the centers of the regions (Figure 1: b) and the average size of the eyes are calculated . Eye size has an important proportional value for face, as size of facial features and their location are regularized in proportion to it. The size of the regions are defined little bit bigger than the actual region to make sure that in spite of small face movements and rotations the important features will stay in side the region.

Next 3D facial cube is extracted for every region. The dimensions of the cubes are X ,Y and t as it presented in Figure 2. In other words facial cubes can be seen as cropped video of facial parts such as eyes, nose and etc.

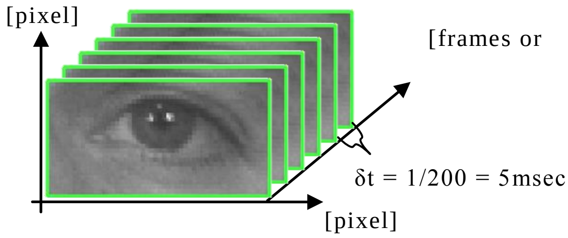

Figure 2: Facial cube captured by 200fps camera.

3.2. 3D Orientation Gradients Histogram

In the case of 2D, I(x,y) stands for image and $\delta I_x(x,y)$, $\delta I_y(x,y)$ stand for image partial derivative. The gradient magnitude and their orientation of each pixel is defined as follows:

$$m_{2D}(x,y) = \sqrt{\delta I_x(x,y)^2 + \delta I_y(x,y)^2}$$
$$\theta_{2D}(x,y) = \tan^{-1}\left(\delta I_y(x,y)^2 / \delta I_x(x,y)^2\right) \qquad (2)$$

In 3D case of (x,y,t), there are number of ways to represent a spatio-temporal gradient vector as discussed in section 2. Our descriptor was specifically adapted for facial movement. Facial movement is examine by observing gradients histogram.

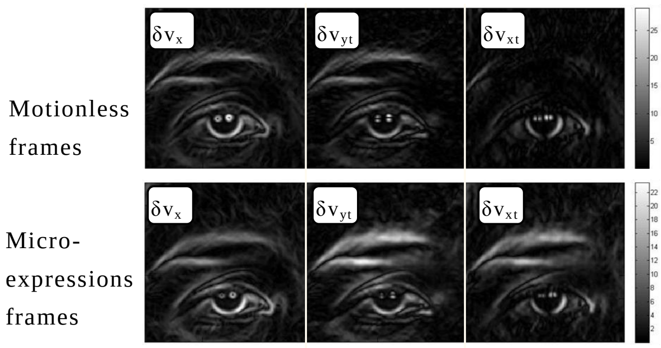

Figure 3: Magnitude values of vectors in δv_{xy}, δv_{yt}, δv_{xt}. The upper row presents combined gradient vectors magnitude computed from the from motionless section of the video. The bottom row presents the combined gradient vectors magnitude computed from the frames during micro-expression

For 3D descriptor calculation given a video cube v(x,y,t) first its partial derivatives along x,y, and t are denoted by $\delta v_x(x,y,t)$, $\delta_y(x,y,t)$, and $\delta v_t(x,y,t)$ respectively extracted. Then, for each couple of partial derivatives $(\delta v_x,\delta v_y)$, $(\delta v_y,\delta v_t)$, and $(\delta v_x,\delta v_t)$ magnitude $m_{xy}(x,y,t)$, $m_{yt}(x,y,t)$, $m_{xt}(x,y,t)$ and orientation $\theta_{xy}(x,y,t)$, $\theta_{yt}(x,y,t)$, $\theta_{xt}(x,y,t)$ are computed using equation (3).

We interpretation each couple of partial derivatives as follow: $\delta v_{xy} = (\delta v_x,\delta v_x)$ represents surface shape, $\delta v_{yt} = (\delta v_y,\delta v_t)$ represents vertical motion and $\delta v_{xt} = (\delta v_x,\delta v_t)$ represents horizontal motion from frame to frame.

$$
\begin{aligned}
m_{xy}(x,y,t) &= \sqrt{\delta v_x(\mathrm{x,\,y,\,t})^2 + \delta y_y(\mathrm{x,\,y,\,t})^2} \\
\theta_{xy}(x,y,t) &= \tan^{-1}\left(\frac{\delta v_x(\mathrm{x,\,y,\,t})^2}{\delta v_y(\mathrm{x,\,y,\,t})^2}\right) \\
m_{yt}(x,y,t) &= \sqrt{\delta v_y(\mathrm{x,\,y,\,t})^2 + \delta v_t(\mathrm{x,\,y,\,t})^2} \\
\theta_{yt}(x,y,t) &= \tan^{-1}\left(\frac{\delta v_y(\mathrm{x,\,y,\,t})^2}{\delta v_t(\mathrm{x,\,y,\,t})^2}\right) \\
m_{xt}(x,y,t) &= \sqrt{\delta v_x(\mathrm{x,\,y,\,t})^2 + \delta v_t(\mathrm{x,\,y,\,t})^2} \\
\theta_{xt}(x,y,t) &= \tan^{-1}\left(\frac{\delta v_x(\mathrm{x,\,y,\,t})^2}{\delta v_t(\mathrm{x,\,y,\,t})^2}\right)
\end{aligned}
\tag{3}
$$

Figure 3 presents the magnitude values of the δv_{xy}, δv_{yt}, δv_{xt} during motionless frames (upper row) and micro-expressions frames (bottom row) of the video.

Next, gradient orientation histograms are computed for every frame for δv_{xy} , δv_{yt} and δv_{yt} cubes. The δv_{xy} (surface shape) gradient orientation histogram contains 8 bins as shown in Figure 4.a . Figure 4.b exemplify the change of the δv_{xy} gradients magnitudes and orientation during the micro-expression. The left image presents frame with neutral expression, right image presents a frame with eyebrow lowered expression (FACS Action Unit 4 (AU)) at the next moment. Blue vectors on both images represent the gradients magnitudes of the surface. The changes of the surface can be observed by the changes in the gradients density, which is represented by blue vectors at the central part of the images.

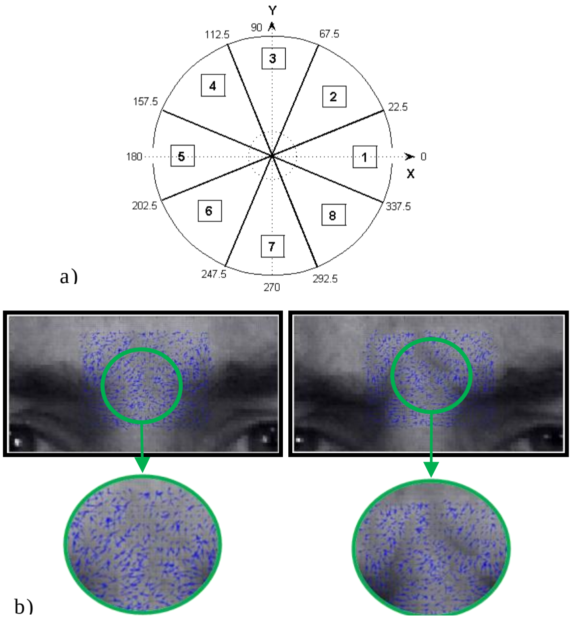

Figure 4: a) δv_{xy} (surface shape) gradient orientation histogram bins. b) Eyebrow before and during eyebrow lowered expression (Action Unit 4)

The gradient orientation histograms for δv_{yt} (vertical motion) and δv_{xt} (horizontal motion) contains 12 bins as shown in Figure 5.1. The figure illustrates the spread of the bins borders in the angle space between the Y and t axes, the spread between X and t axes is identical so all the following explanations concern both histograms .

The angle space of δv_{yt} and δv_{xt} is split for 14 regions; 1 to 12 regions that are used for descriptor calculation and two NaN regions that are excluded from the descriptor. In addition, 12 regions divided to three subgroups (a) , (b) and (c). The differences between the subgroups are NaN described below.

NaN region: contains all the gradients vectors whose change rate in direction t is small, and they indicate no change between the frames in the corresponding pixel. The gradients in to x and y direction are already included in the δv_{xy} histogram, so we do not include them in the δv_{yt} and δv_{xt} histograms.

(a.) region: contains gradients vectors which have small changes in x or y direction but have a significant change in t direction. It means that the corresponding pixels have a significant change only in their intensity between the frames.

(b.) region: contains gradients vectors with similar change rate in X and t (or Y and t) directions.

(c.) region: contains gradients vectors that indicate high rate change in Y direction and relatively small in t.

(Figure
5.2.)

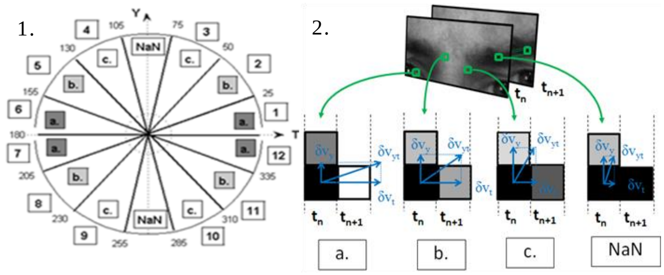

Figure 5: 1) δv_{yt} and δv_{xt} orientations histogram are split to 14 bins and grouped in 4 subgroups. (only12bin are actually used in descriptor).
2) Illustrates 4 vectors in Y-T surface that are associated to subgroups a ,b ,c and Nan, respectively from left to right.

Then, with these group classifications the facial expression changes could be assigned to these regions. NaN region indicates no or little motions in the face between the frames. Region (a.) indicates big change in pixel intensity between the frames and can represent motions such as blinking and eyebrow movements. Regions (b.) and (c.) indicate different motions, appearances or disappearances of the skin folds on the face surface. During the histogram computation small vectors in all directions are excluded.

3.3. Descriptor

First, histograms δvxy , δvyt and δvyt are calculated. Then all the histograms from the same frame are concatenated. At last, feature vector are normalized. This way the motion between every frame in the video v(x,y,t) is represented by normalized 32 bins vector. (Figure 6)

This time k-mean clustering method was for descriptor classification in every facial region seperatly. Results will be presented in the experiments section.

4. Datasets

Establishment of facial expressions analysis resorch suffer from lack of extensive dataset [17].

The only publicly know database of facial spontaneous expressions is RU-FACS [14]. The database was create consists by using a 'false opinion' paradigm and contain 100 subjects.

In this paradigm, first the subjects inquire regarding some issue and then asked to take the opposite stand from what they had reported before. In the next step they are interviewer by retired FBI agent and the subjects were asked to convince him that they are telling the truth. This paradigm has been shown to elicit a wide range of emotional expressions as well as speech-related facial expressions [15]. The subjects' faces were captured by four synchronized Dragonfly cameras by Point Grey, whose maximum frame rates with 640x480 resolution are 30fps. Only 33 subjects had been FACS-coded .

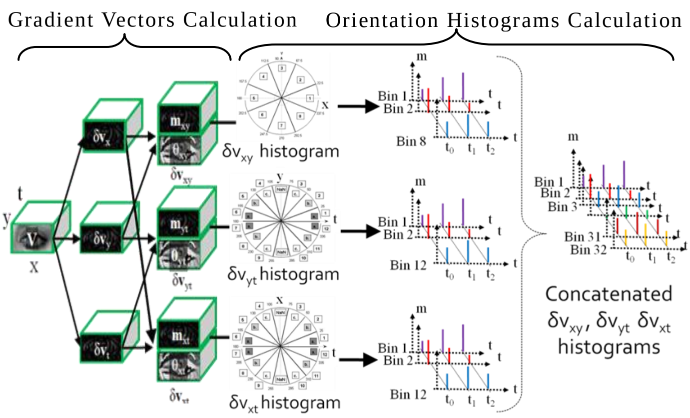

Figure 6: Illustration of the descriptor calculation diagram for a facial cube procedure.

Other two well known FACS-coded datasets are: Cohn and Kanade's DFAT-504 [16] that contain 100 university students and Ekman and Hager dataset with 24 subjects. In both datasets the subjects were instructed to perform different expressions, while the emphasis was put on regular facial expressions (not micro-expressions). The facial expressions in those datasets were FACS-coded by two certified FACS coders. The lack of micro-expression dataset captured by high speed camera brought the need to create one.

Two basic approaches can be used for setting the ground truth for facial expressions in the videos: One is by using descriptive tags such as FACS as it was done the in previous datasets. Second, is by measuring the displacement of the predefined feature points on the face and reconstruction of the 3D shape of the face [17] . Concerning our main goal in this resorch FACS is seems as a more suitable approaches for us as it can be easily liked to emotional state afterward.

We separated this creating dataset task into two stages. First stage contained videos of posed facial expressions and will allow first evaluation of our algorithm. Second stage include more realistic videos as it will be explained in the next section.

This paper focuses on the results of the first stage dataset that contains posed micro-expressions. For video capturing Grasshopper camera by Point Grey was used. Camera settings: 480x640 resolution, 200fps, RAW8 mode (in this mode minimum internal signal processing executed by camera allows 200pfs).

McCabe [16] recommendations for mugshot and facial image filming was used as a guidelines. Three lights were used for shadow cancelation, left, right and upper point lights with diffusion sheets to minimise hot spots on the facial image. Uniform background approximately 18% gray was used and the camera was rotated 90 degrees (640X480) to maximise the amount of pixel on the face region. Dataset contains 10 university student subjects (5 Asian, 4 Caucasian, 1 Indian) (See Figure 7).

Figure 7: Subjects from our new dataset

The participant were instructed to perform 7 basic emotions with low facial muscles intensity and to go back to the neutral face expression as fast as possible, simulating the micro-expression motion. Video facial cubes were extracted from all the faces in the dataset.

Then, all the frames in every cube was FACS-coded. We extended the FACS-coding by adding three time-tags to each AU as follows: i) “Onset” - from a neutral expression towards one of the AU's (contraction of the muscles), ii) “Apex” - expression it self (holding the muscles contracted) and iii) “Offset” - from AUs towards neutral expression (release of the constructed muscles).

5. Experimental Setup

For proceeding with the second stage of more realistic video dataset collection we develop a new experimental setup that simplifies and allow to speed up all the process. This tiem we would like to include videos that resemble hostile and dangerous behaviour. Until know for collecting a micro-expressions videos the subject participated in a ’false opinion’ paradigm as it was explained before. First, following this paradigm we face the need to have an access to well trained investigator, like retired FBI of police officers. Second, for collecting the ground truth information of micro-expressions from the recorded videos the FACS experts were requested to tag the micro-expressions in frame by frame manner carefully analyzing the conversation between subject and the interviewer. This is very time consuming and complicated process. In addition to that, for extensive tests of our algorithm, the accurate extraction of the 3D face location in the video are required. Simple solution for 3D face location extraction is tracing of the physical marker that are placed on the head in advance. However, the face to face questioning process will be influenced by the appearance of the markers on head.

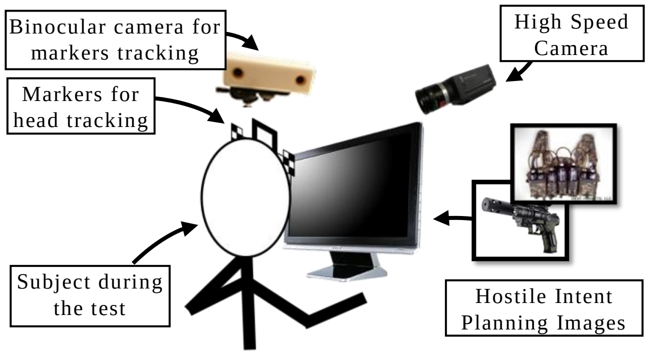

Figure 8: Illustration of experimental setup, where the subject is question by computer about his hostile action planning using special images. For recording subject facial motions high speed camera is used. Binocular camera is used for extracting the 3D location of the face by tracing special markers that are attached to subject head.

Together with Dr. Burgoon we designed a new experimental setup the suspect will sit in front of the computer, alone and the interaction with the suspect will be done using a computer program. In this way no special human interview is requested. The usage of the computer program will automatically extracted the critical time in the video for future analysis of FACS experts. In addition, attachment of the special markers to the subject won't interferer the experimental procedure and will minimize the influents on the emotion state. Figure 8 illustrates out new experimental setup. For speedup the frame by frame FACS tagging procedure a new GUI program was develop as well. Compare to the program that was used in RU-FACS [14] it allow more complicated combinations of FACS tags to be processed and adds the ability to identify the different phases of AUs such as “Onset”, “Apex” and “Offset”. Testing results of our algorithm ageist videos that was collected by using new experimental setup will be presented in near future.

6. Experiments

6.1. Micro-expressions classification

Current implementation of the algorithm done in Matlab using "Piotr's Image&Video Toolbox" [18] without special emphasis on performance. However, parallel structure of the algorithm can get benefit from implementation on GPU environment.

Tree parameters was tuned during the experimental procedure. First, 3D Gaussian smooth filter values were defined separately to each group of facial cubes. The other two parameters are magnitude cut-off threshold (section 3.2). and size of the "NaN" bins in histogram computation process.

The facial cubes from our dataset were divided to 8 groups: (a) forehead, (b) left and right eyebrows, (c)left and right eyes, (d) between the eyes, (e) lower nose, (f) mouth, (g) left and right mouth corner, (h) chin.

We used k-mean cluster methods. The number of clusters depends on the number of different motions that should be detected in each cube, it varies as 4 in (a), (e), (d),(e),(h) cubes; 7 in (b) cubes and 13 in (c) and (f) cubes.

Facial Cubes	Neutral	FACS AU	Onset	Apex	Offset
a) Forehead	0.95	AU2	0.93	0.95	0.91
b) Eyebrows	0.93	AU4	0.84	0.83	0.9
	-	AU5	0.86	0.83	0.85
c) Eyes	0.92	AU4	0.84	0.83	0.84
	-	AU7	0.86	0.8	0.81
	-	AU43	0.85	0.85	0.84
d) Between the Eyes	0.9	AU4	0.93	0.9	0.84
e) Lower Nose	0.94	AU10	0.95	0.93	0.95
f) Mouth	0.88	AU12	0.81	0.79	0.85
	-	AU24	0.81	0.67	0.79
	-	AU26	0.83	0.77	0.8
g) Mouth corners	0.83	AU13	0.85	0.81	0.89
h) Chin	0.89	AU17	0.83	0.83	0.84

Table 1: Classification results of micro-expression with 10 subjects.

In table 1 we report a classification precision for each class.

The results indicate good classification precision in groups (a) forehead, (d) between the eyes, (e) lower nose are high. It is due to the small number of classes and they differ greatly from each other. The group (h) showed lower rates because of the beards on two faces in the dataset. Mouth group (f) shows the worst results. In many works on face expression recognition mouth movements were mentioned as the most challenging for classification.

By analysing the results in table 1 it can be seen that "Onset" and "Offset" phases of almost all AU have higher classification precision then the "Apex" phase. This indicates that the proposed descriptor is more suited for motion recognition and segmentation than for classification of the frames with "Apex" tags.

6.2. Time measurements of micro-expressions

The ability to extract and measure the duration of the tree micro-expressions phases such as “Onset”, “Offset” and “Apex” is important for future classification methods. Moreover also it is know that the timing of the micro-expression is an important characteristic in psychology analysis, in privies works the psychology researches did not had an easy access to that information.

We believe, this time-measuring ability will be useful for the implementation of micro–expressions for hostile intent and danger demeanour detection.

Figure 9 presents the values of (1,5,6,8) bins in δv_{xy} histogram, the values of (2,5,8,11) in δv_{xt} and (1,6,7,12) bins in δv_{yt} histogram during the AU4 micro-expression. Note that (1,5,6,8) in δv_{xy} is considered to be a representation of constructed muscles, (2,5,8,11)in δv_{yt} and (1,6,7,12) in δv_{yt} are represent the motions of the muscles. Three main phases of motion can be easily observed: “Onset” - constriction of the muscles from 52 to 58 frames that lasted approximately 0.03s, “Apex” - holding the muscles constructed from 58 to 63 frames, the expression it self that continued 0.25s and “Offset” - the release of the muscles from 63 to 68 frames that lasted approximately 0.03s

7. Conclusion and Future Work

For detection and predicting danger demeanor micro-expressions are found to be important behavioral source of informatio.. In this paper, we presented a novel approach for facial micro-expressions recognition using 200fps high speed camera of for capturing the face motion. In order to initiate the research of detecting micro-expressions using high speed cameras, a new dataset, experimental setup and programs for ground truth extraction was presented. We showed the recognition results of 13 different micro- expressions. The unique feature of our approach is the ability to measure the duration of the three phases of micro-expressions
(“Onset” - constriction of the muscles, “Apex” - muscle construction and the “Offser” - release of the muscles).

Future extensions include face tracking and testing of the system with a wider variety of faces. More sophisticated generic classification algorithm will be applied in near future.

In conclusion, in the academic sense, the use of micro-expressions for hostile intent detection is still unknown.

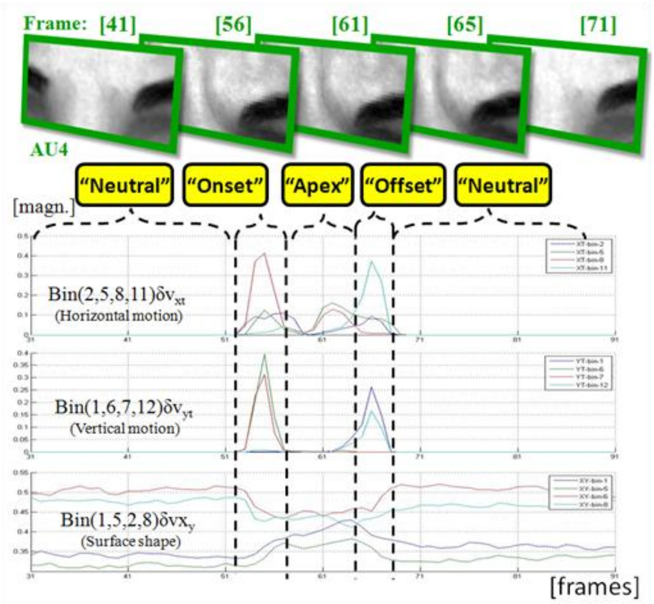

Figure 9: Analysis on AU4 values of (2,5,8,11) bins in δv_{xt} , (1,6,7,12) in δv_{yt} and (1,5,2,8) bins in δv_{xy} histograms.

In practice, the manual analyses of micro-expressions are widely used by secret services and homeland security departments around the world. However, due to the nature of micro-expressions it is difficult to detect or investigate them manually, this leading to the need for an automatic system.

This work is a step towards creating such a system.

Acknowledgements

The authors wish to thanks Piotr Dollar for making his very useful Matlab toolbox open to the public.

References

[1] P. Ekman, "Telling Lies" 2nd Edition, Norton, (2009).

[2] S. Porter and L. Brinke, "Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions", *Psychological Science*, vol. 19, pp. 508-514, (2008).

[3] P. Ekman, "Facial Expressions of Emotion: an Old Controversy and New Findings", *Philosophical Transactions of the Royal Society*, vol. B335, pp. 63-69, (1992).

[4] P. Ekman, "Facial Action Coding System", *Consulting Psychologists Palo Alto*, (1978).

[5] Z. Li, *Handbook of Face Recogntion*. Springer, (2005).

[6] S. Lucey, "Investigating Spontaneous Facial Action Recognition through AAM Representations of the Face", *Face Recognition Book*. Pro Literatur Verlag, (2007).

[7] M. Pantic, "Detecting Facial Actions and their Temporal Segments in Nearly Frontal-View Face Image Sequences", *Systems, Man and Cybernetics*, vol. 4, pp. 3358-3362, (2005).

[8] Y. Tong , J. Chen, and Q. Ji, "A Unified Probabilistic Framework for Spontaneous Facial Action Modeling and Understanding," *Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 258 - 273, (2010).

[9] S. Kumano, K. Otsuka, J. Yamato, E. Maeda, and Y. Sato, "Pose-Invariant Facial Expression Recognition using Variable-Intensity Templates," *International Journal of Computer Vision*, vol. 83, no. 2, pp. 178-194, (2009).

[10] M. Bartlett, "Automatic Recognition of Facial Actions in Spontaneous Expressions", *Journal of Multimenia*, vol. 1, no. 6, pp. 22-35, (2006).

[11] M. Marszalek, "A spatio-temporal descriptor based on 3D-gradients", *British Machine Vision Conference*, (2008).

[12] P. Dollar, "Behavior recognition via sparse spatiotemporal features", *Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp. 65-72, (2005).

[13] P. Scovanner, "A 3-dimensional sift descriptor and its application to action recognition", *International Multimedia Conference*, pp.357 - 360, (2007).

[14] M. Frank. RU-FACS-1 Database. [Online], (2009) http://mplab.ucsd.edu/databases/databases.html.

[15] P. Ekman, "Appearing truthful generalizes across different deception situations", *Journal of ersonality and social psychology,*, vol. 86, p. 486–495, (2004).

[16] T. Kanade, "Comprehensive database for facial expression analysis", *Automatic face and gesture recognition conference*, pp. 46–53, (2000)

[17] Javier R. Movellan, "Tracking Motion, Deformation, and Texture Using Conditionally Gaussian Processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. pp. 348–363, (2010).

[18] V. Rabaud. Piotr's Image & Video Toolbox for Matlab. [Online], (2009) http://vision.ucsd.edu/~pdollar/toolbox/doc/

[19] M. McCabe. Best practice recommendation for the capture of mugshots. [online], (2009) http://www.itl.nist.gov/iaui/894.03/face/bpr_mug3.htm

[20] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial Micro-Expressions Recognition Using High Speed Camera and 3D-Gradients Descriptor", *Conference on Imaging for Crime Detection and Prevention* , pp. 6, (2009).