Towards Developing Robust Multimodal Databases for Emotion Analysis

Maria Alejandra Quiros-Ramirez*, Senya Polikovsky†, Yoshinari Kameda†, Takehisa Onisawa*
*Humanistic Systems Laboratory, † Computer Vision and Image Laboratory
Graduate School of Systems and Information Engineering
University of Tsukuba
1-1-1 Tennodai, Tsukuba, 305-8573 Japan
alejandra@fhuman.esys.tsukuba.ac.jp

***Abstract*—Understanding emotions can make the difference between succeeding and failing during communication. Several systems have been developed in the field of Affective Computing in order to understand emotions. Recently these systems focus into multimodal emotion recognition. The basis of each of these systems is *emotion databases*. Even though a lot of attention has been placed in capturing spontaneous emotion expressions, building an emotion database is a task with several challenges that are commonly neglected, namely: quality of the recordings, ground truth, multiple device recording, data labeling and context. In this paper we present a new spontaneous emotion database, with human-computer and human to human interactions. This database is composed by eight different synchronized signals, in four interaction tasks. Strategies on how to deal with emotion database construction challenges are explained in detail.**

***Index Terms*—emotion recognition; microexpressions; facial expression; gestures; infrared image; spontaneous emotions; multimodal synchronization.**

I. Introduction

A sly smile, a frown or a sigh can tell us a lot about the internal feeling of the people surrounding us, either if we are in direct communication with them or not. Emotions are essential in every type of daily life interactions. Automatic recognition of emotions has been actively studied in the last decade [1]. The application of affect and emotion recognition systems is clearly broad and it spans fields such as HCI, marketing, health. Studies on detecting and decoding emotions have one point in common: the need of robust and high-quality *emotion databases*.

Due to the growth and advances in the field of *Affective Computing*, work on emotion databases has also evolved. Still, most of the existing databases are focused on the emotional content, neglecting technical aspects such as quality of the data and management of huge amounts of information. An important point that has not been explored deeply so far is the synchronization between multiple sensors for recording multimodal interaction, to find the relation between signals in time.

A. Previous Emotion Database Efforts

Emotion recognition development is strongly coupled with emotion databases that are required to train and test emotion recognition in systems. The first available databases were focused mainly in single cued emotion recognition. Facial emotion recognition had been the main focus modality through these years, therefore most of the databases were focused on the face, facial expressions and FACS [2]. Another characteristic of the first emotion databases is the use of actors or posed emotions.

Recently, several of the studies in emotion recognition are focused on analysing emotions from multiple modalities, as well as natural interaction instead of the acted emotions studied in the past [3]. Following these requirements, it has become necessary to design and develop new databases that present natural expressions with dual or multiple cue support [4], [5], [6]. These databases are mainly focused on visual and auditory cues in order to study facial expressions, gestures, speech and prosody.

Most of the current databases are focused in audiovisual expressions, yet there are some emotional databases which provide other varieties of cues as complement to audiovisual modalities. For example, the group of Wang et al. [7] has presented a natural visible and infrared facial expression database, providing visible images as complement through thermal images for emotion recognition.

Another important feature not quite studied at the time of data base creation, is the synchronization among recording devices. [8].

B. Challenges in Emotion Databases

Building an emotion database is a time consuming and hard task. Yet, the database is an indispensable basic step towards the construction of emotion recognition and emotion expression systems. There are several issues at the time of developing or choosing a free distribution emotion database:

- *Quality of the recordings*. While going through the process of building an emotion database, a lot of effort is usually invested in obtaining rich emotional display, which is necessary in a quality database with emotion analysis goals. Yet, emotional quality is not the only determining factor in a successful emotion database. It is required to invest the same amount of effort in the quality of the recording settings in order to be able to extract expressions and emotions automatically. Currently, several available databases present a lot of rich emotional contents, but the quality of the recordings and experi-

TABLE I
DATA CAPTURING SYSTEM DEVICES (INDEXED FOLLOWING FIGURE 1)

#	Signal	Device	Sampling Rate	Frame Size	Purpose
1	Motion Capture	Optitrack V120:Trio	120FPS	640x480(3)	Subject position ground truth
2	Pressure Sensors(2)	N/A	20000Hz	N/A	Feet pressure
3	Microphone(2)	Sony	22050Hz	N/A	Speech & Prosody
4	High Definition	PointGrey Flea2 CCD Camera	30FPS	1024x768	Waist-up body movement and gestures
5	High Speed	PointGrey Grasshoper CCD Camera	180FPS	640x480	Facial Microexpressions
6	Infrared	FLIR SC300	60FPS	320x240	Position ground truth & Temperature changes
7	High Definition	PointGrey Flea2 CCD Camera	30FPS	1024x768	Head and face movement and gestures
8	Photosensor	N/A	20000Hz	N/A	Stimulus onset recording

mental environment set-ups makes it difficult to perform automatic emotion analysis.

- *Ground truth*. As for ground truth, it is necessary to clear up the difference between emotion ground truth and feature ground truth. Emotion ground truth refers to the real emotion displayed during the recordings. Feature ground truth refers to the real feature location. It is necessary to provide support for both types of ground truth in order to assess the eventual result of an automatic emotion decoding or encoding system.
- *Multiple recording devices*. Multimodal emotion analysis presents in many cases the need of combining several devices to capture the subject's interaction in detail. Having several recording devices brings up the challenge of synchronization in order to properly support multimodal automatic analysis of emotions. Very few has been done so far to support high level synchronization between different signals. Synchronization is a key point in order to analyze emotional changes in time and the relation between different cues in representing emotion.
- *Large amounts of data labeling*. Databases are usually constituted of several hours of recordings. Adding several devices to an emotion recording session makes the size of the database even harder to handle. Two main points stand out: how to speed the process of segmenting the sessions into the significant material and how to label the sessions both in the emotional and feature dimensions for ground truth.
- *Context*. It is necessary to consider different factors as the whole contextual background while recording a database and utilize it for emotion recognition. Such factors include specifics about the interaction, for example, what is the subject's task or what is he or she interacting with. It also includes specifics of the subject like age group and ethnicity.

C. Scope of this study

A new synchronized multimodal database is presented. Each of the steps of the construction of this database has been designed to deal with the problems cited in the previous section. This database can be used by different types of professionals, with the aim of understanding, analyzing and recognizing emotions in different human interaction scenarios. The database has been recorded through a robust and highly scalable capturing system, which assures high synchronization between the different sensors. In this paper, the following topics are discussed: the basics of the capturing system and experimental setup, the experiments carried out in order to elicit natural emotions in the laboratory, the current database and strategies to deal with issues that arise while constructing such databases.

II. DATA CAPTURING SYSTEM

It is necessary to dedicate the same attention to quality of the recording system as the effort given to design the emotional display experiments. The goal of this database is to provide an emotional corpus for researchers to explore the internal emotional state of the subjects. In this case, with the purpose of being as least invasive as possible, devices for recording bio-electrical signals are not considered.

For this database, the aim is to explore emotional internal state mainly through visual and auditory cues. Devices like cameras and microphones emulate human senses of vision and audition. Yet, as a complementary exploring mechanism thermal and pressure devices have been added to the recording process. Although these devices do not represent any type of human sense, they are useful to collect information that is possible for people to observe or listen, but hard to extract from cameras or microphones. Examples of them are a blushing face that can be obtained from heat information, or a tapping foot that can be obtained by pressure sensors under the feet. Thus, the data capturing system includes cameras, microphones, pressure sensors and illumination. Figure 1 depicts the details of the layout of the system in the experimental room.

In order to support our experiments, an experimental environment has been designed such that participants would be able to interact with a computer and also with an interviewer. The experimental room was selected in order to provide privacy to the subjects and avoid any interference or contact with entities outside of the experiment. The size of the room was 3.3mx3mx2.5m (length, width, height) with a single door and no windows. Having no windows allowed the complete control of the illumination settings.

An arrange of six computers is used to provide dedicated device control and data recording, one of which performs as a server or main computer. The whole system starts and stops

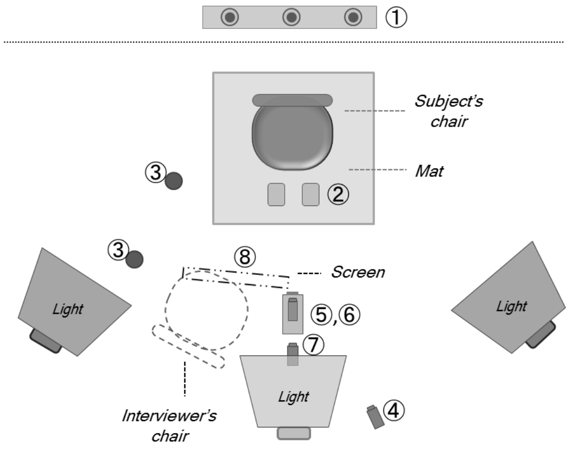

Fig. 1. Experimental Room and Data Capturing System Layout. The data capturing system consists of an ensemble of high speed and high definition video cameras (4,6,7), infrared camera(5), directional microphones(3), motion capturing system(1) and pressure sensors for the feet of the subject (2). Besides, a photosensor has been attached to the experimental screen in order to ease the ground truth labeling process(8). A room with no windows was chosen in order to control the lightning conditions and assure no interaction between the subject and agents outside the experimental environment.

by sending signals from this main computer to shorten time and effort. An acquisition card has been set up in the main computer to control the timing of the signals and generate synchronization between different devices.

Table I describes the different devices in our recording ensemble indexed according to the numbers in Fig. 1. The device arrange is explained in more detail below.

A. Video signals

Four cameras are used to record the videos. Each of the cameras records a specific desired cue. Two PointGrey Flea2 CCD cameras are set up to record in high definition. Both cameras record at 30 frames per second (FPS) with a resolution of 1024x768. One of them (index number 7), is set up to have a closer view of the subject's face. For this purpose the camera's lense is set with a distance of 1.7m in front of the subject at a height of 1.1m from the ground. The other high definition camera (index number 4) is positioned in order to record a full body view of the subject at a distance of 2.2m in an angular position from the subject at a height of 0.8m. Both high definition cameras were separated by a distance of 0.7m.

A high speed camera PointGrey Grasshoper CCD (index number 5) was positioned on the same line of our first high definition camera. This camera captures at 180FPS with a resolution of 640x480 with main focus to the face of the subject. The camera is positioned 1.2m from the ground, separated 1.2m from the subject. The purpuse of the high speed camera is to record the slightest movements in the facial muscles [9].

Finally, a FLIR SC300 infrared camera (index number 6) is also positioned in front of the subject in order to record shoulders and face at 1.2m distance from the subject and 1m height from the ground. This camera records at 60FPS with a resolution of 320x240.

The recordings of four cameras were not compressed, thus, recording every frame in RAW format. Both Flea2 cameras, Grasshoper and FLIR were plugged to three different workstations respectively, in order to provide dedicated control of recording process and data saving and storage process.

B. Audio signals

Two directional microphones (index number 3) are used in order to capture the vocal interaction of the subject as well as the interviewer's interaction in the third experiment. The sampling rate of each microphone is 22050Hz. The subject's microphone is placed 0.5m away from the subject, while the interviewer's microphone is placed 0.6m away from the interviewer's place. Both microphones are separated by a distance of 0.6m.

Both microphones are connected to a dedicated workstation. The recorded data is stored as binary files.

C. Complementary signals

The capturing system includes two pressure sensors for the feet and a photosensor (index 2 and 8, respectively). Both sensors sampling rate is 20000Hz.

As depicted in Figure 1, both pressure sensors are placed beneath a mat to cover them from the view of the subject. On the other hand, the photosensor is attached to the bottom right corner of the screen where the stimulus is displayed during the experiments. The displayed stimulus were prepared with a small square on the bottom right corner, of white and black color. When the stimulus is being displayed, the square would be white, otherwise black. This change in the monitor brightness generates a signal to tag automatically the onset of the stimulus, therefore, to tag the point of interest of the interaction.

A motion capture camera is placed behind the subject's seat (index 1). The camera is set up horizontally at a height of 2.45m, 0.9m behind the subject. The camera's sampling rate is 120 FPS with a frame size of 640x480 for each of the three lenses. A grey curtain is set up as a background in order to cover the pole holding the Optitrack. This curtain is placed 0.6m behind the subject's seat. The Optitrack is used only during the fourth task of the experiment.

D. Illumination

Three different sources of light are placed, one in front of the subject and two in the sides. The side lights are placed 1.8m away from the subject seat and 1.2m from the ground, facing towards the subject. The frontal light was place 3m away, 2.40m height from the ground.

E. Other items

In order to support our experiments, a screen to display pictures and videos is used, and a chair for the interviewer. The screen is positioned 0.85m away from the subject at a height of 0.83m from the ground, on the left under the arrange of cameras. While not in use, it is removed. The chair is placed

TABLE II
SUBJECTS CONTEXT INFORMATION

Region	**Female**	**Male**	**Total**
Africa	2	2	4
America	5	3	8
Asia	9	6	15
Europe	2	5	7
Oceania	1	1	2
Total	19	17	36

in the same distance for the interview section, removed as well while not in use.

III. EXPERIMENTAL SETUP

Data collection procedures vary depending on the purpose of the emotion database. In this case, our goal is to obtain spontaneous emotional displays from participants with different cultural backgrounds. The experiment consists of four phases: pictures display, videos display, interview scenario and a task of simple body motions for motion capture detection. The data acquisition process is explained below.

A. Subjects

A total of 36 naive people from different nationalities participated voluntarily in the experiment. Their ages range from 21 and 35, 14 of the participants were female and 19 male. Each of them had different educational backgrounds from undergraduates to post-doctoral fellows, from the University of Tsukuba and nearby research centers. English proficiency ranged from intermediate to native. Five of the subjects wore glasses and one had beard. The breakdown of participants by region and gender can be observed in Table II.

B. Task 1: Pictures display

For the first stage of the experiment, the subject was asked to observe images displayed in the screen. A total of 20 emotionally loaded images were selected from the GAPED affective picture database [10]. For each of the images in the database two dimensional valence/arousal labeling is provided. A total of 20 images were selected, 8 described as positive, 8 described as negative and 4 described as neutral. Each subject took about 2 minutes and a half to complete this task.

During the experiment, a grey screen is displayed for 3 seconds before each picture. The picture itself is displayed for 5 seconds. The subject was asked to look carefully at it. After each picture, the subject was instructed to evaluate aloud his or her feeling about the picture with a five point scale from -2 to 2 representing negative to positive feeling. This scale was displayed in the screen after each picture.

The order of the pictures was randomized for each participant.

C. Task 2: Videos display

For this task a subset of the videos presented in [11] was selected. These videos were originally labeled in the categories of Amusement, Disgust, Fear, Sadness or Neutral. Table III presents the list of the videos selected per emotion and the length of each video.

TABLE III
SUBSET OF VIDEO EXCERPTS OBTAINED FROM HEWIG'S STUDY [11]

Emotion	**Video excerpt**	**Length (s)**
Neutral	Hanna and her Sisters	102
Amusement	When Harry met Sally	95
	An Officer and a Gentleman	124
Disgust	Pink Flamingos	107
	The Godfather	168
Fear	Halloween	89
	Silence of the Lambs	113
Sadness	The Champ	122
	Officer and a Gentleman	109

During the experiment the subject is presented with five videos to rate: one for each of the four emotions and a neutral video. The presented videos and their order were randomized for each participant.

Before each video a grey screen is presented for 3 seconds. After this, the video is reproduced automatically. The subject is asked to watch it until the end. After the video, a screen presenting the words neutral, anger, amusement, disgust, fear and sadness is presented. The subject should select the word that is closer to his or her feeling and indicate the strength of the emotion in a scale of zero to five. The assessment is performed aloud.

D. Task 3: Interview

Besides interactions with the computer, an experiment in order to study human to human interaction is designed. An interviewer different from the experimenter is in charge of asking questions to the subject.

E. Task 4: Simple Motions

For the final task, markers are placed in the head and shoulders of the subject in order to perform the tracking. The subject was requested to perform a series of head and shoulder motions, guided by the experimenter.

F. Experimental Protocol

Each subject is received in a registration counter, where he or she was asked to fill a personal identification form, a present mood questionaire and an ethnic questionaire. The experimenter explains the four different tasks and informs the participant that the session would be recorded. At this point, the participant receives a consent form and is asked to read it carefuly. If the subject agrees with it, then he or she will sign the form once before the experiment and once again after the experiment.

After filling the form, the subject receives a numeric identifier that is attached to the subject's shirt in order to be visible during the whole experimental process. Then, the height of the participant is measured in order to set the correct size of the subject's chair before starting the recordings. After this the subject is guided inside the experimental room.

Before starting the tasks, a picture of the participant is taken for database indexing purposes. The experimenter explains again the first task and leaves the room before the task begins. This process is repeated for the second task. For the third task, the experimenter instructs and prepares the subject for the interview task. She makes sure the subject understands the task. After this the interviewer enters the experimental room and the the experimenter leaves the room before the tasks begins. For the last task the experimenter guides the subject in several body motions.

After completing the four tasks, the subject is guided again to the reception desk, where he or she would signs a second time the consent form. Three pieces of chocolate are given as a gratitude sign for participating in the experiment.

Fig. 2. Four views of the subject during one of the experiments: (a) high definition face and head, (b) high speed face and head, (c) wide angle high definition full body, (d) infrared head and shoulders.

IV. MULTIMODAL EMOTION DATABASE

Setting up a more detailed recording system symplifies the post-processing of the recorded data. The array of devices used for collecting the data of this database was selected in order to obtain the most of each of the interactions induced by the four tasks of the experimental protocol. The purpose of the construction of this database is to follow a series of steps that will provide the best conditions to record, collect and manage multimodal emotional data.

Next, the strategies used to deal with challenges found in emotion database construction are explained.

A. Ground Truth

Two different strategies are used in order to tag emotional ground truth and feature ground truth automatically.

Emotional ground truth is obtained in tasks 1 and 2 of the experiment through the emotional information implicit in the stimulus plus the verbal self-assessment provided by the subject.

In the case of task 1, there is a valence/arousal coordinate linked with each presented picture [10]. The subject describes his or her own feeling as negative, neutral or positive and indicating the strength of this feeling in a three point scale. For task 2, subjects are asked to label their own emotion first by choosing one of the six emotional words explained in the previous section. Then, rate this emotion's strength in a five point scale.

Regarding feature ground truth, the fourth task of the experiment is dedicated to motions in order to set a baseline for the movements and positioning of each subject. The motion caption system provides exact coordinates of the position of the subjects head and shoulders. This feature ground truth provides an easy comparison support for automatic feature extraction algorithms of both head and shoulders in order to verify and assess their correctness.

B. Multidevice Recordings

There are two main challenges when recording emotional data from several devices: it is necessary to develop a protocol in order to assure automatical operation of all the devices, as well as a paradigm in order to be able to locate in time the correct sample from each of the devices. This latter point is critically important for the understanding and analysis of multimodal interactions.

In our case, the computers network has each terminal controlling one single device and centralizing the operation through a main computer. The signals from the main computer controls the operation of all the devices. To assure synchronization, an acquisition card is set, which sends a start and stop bit to mark the begining and end of the recording for each device.

C. Data Labeling

The next challenge consists in finding a scheme for labelling the data collected during the experiments. The capturing system provides two main advantages: the start and stop bit of the synchronization and the photosensor that detects the stimulus presentation on screen.

The start and stop bit allows removing pieces of recording that do not contain information of interest both before and after the experimental interaction. Tasks one and two of the experiment have been designed to be segmented automatically obtaining the subjects stimulus exposure segments from the information recorded with the photosensor. The beginning and end of each interaction per stimulus are defined by each variance in the contrast of the stimuli square and the data from the 8 devices can be segmented due to the synchronization of the signals. The photosensor was used as well for the fourth task of the experiment signaling manually the start and stop points of each motion performed by the subject.

The interview task needs to be treated different. The interactions during the interview are segmented by hand using the recordings of both microphones to signal the start and stop of each question.

D. Data Base Contents

The present database consists of recordings from 36 subjects with different regional backgrounds. The contents of the database are separated in 4 different tasks, images, video, interview and motion baseline, in order to differentiate between human to human interaction and human-computer interaction.

Image task contains 660 interactions in total corresponding to 55 minutes of emotional reactions to images presented to the subjects. Video task consist of 165 interactions in total corresponding to 4 hours 52 minutes. Interview task consists of 20 interactions of about 5 minutes each. Motion baseline task contains 627 posed motion interactions of about 1s each.

An example of images from the database is presented in fig. 2. In this figure four types of signals are presented: high definition (a), high speed (b), wide angle high definition (c) and infrared (d). Each entry of the database contains information of these four signals, plus feet pressure and voice. The samples are labeled with emotional ground truth and context information is included (type of interaction, nationality and gender). Using this information it is possible to train and test emotion recognition and expression systems considering context of the interaction and subjects.

A cross-cultural emotion recognition system built based on this database can be observed in [12]

V. Discussion and Conclusions

The present study introduces a new human interaction database for emotion and communication analysis goals. The database offers four types of interaction, three tasks to record spontaneous emotions and one task to record posed motions from 36 subjects with different cultural backgrounds. Several strategies were employed in order to deal with the issues that emerge in the process of construction of an emotion database.

The main focus in most of the emotion databases available nowadays corresponds to emotional display of emotions. Although it is clear that emotional display is a major point inside of an emotion database, poor management of the database construction could lead to collection of information that is not easy to be analysed or to be managed.

Several problems have been found in usual emotion databases which compromise the quality of the data available. A group of solutions were presented in order to improve these issues. In this paper, two major dimensions that have been disregarded at the time of building an usual database are considered: data-capturing setup and data labeling.

About the data-capturing setup, a recording system array of different types of signals is prepared: video, audio, infrared and feet pressure with the purpose of recording in detail the interaction of the participants.The system provides strong synchronization between the eight devices. A lack of synchronization between the devices would not allow analysis of the concurrence or consequence of the display of emotion at some point in time, which is a very important characteristic of emotional expressions. Most of the databases currently available neglect this point. Synchronization also simplifies the task of segmenting the interactions in samples of interest, since all the signals can be segmented concurrently.

Although the amount of data after recording is usually considerably big, our system provides a photosensor that tags the regions of interest during the recording time and reduces the initial amount of data to label. The labelling process is supported by providing initial ground truth of images and videos presented to the subject. This initial ground truth gives an estimate of the emotion that the subject is expected to feel. Besides, the self assessment of internal emotion state by each subject provides emotional ground truth of the *real* emotion that the subject felt. Feature ground truth is supported as well with the task of the experiment. This ground truth provides a feature baseline for each subject. The importance of ground truth is the capacity of comparison the extracted feature with the real feature and the recognized emotion with the real emotion. Without this information, it is impossible to evaluate the efficacy of the emotional system.

These characteristics provide a robust multimodal emotion database, not only in the emotional display dimension, but also from the recording and labeling point of view.

It is necessary to build databases that are purpose-driven and context-driven instead of general-purpose databases, in order to improve the quality of the interactions and therefore, the quality of the emotion recognition system. In the future, it is necessary to consider the context of the recordings in deeper detail.

References

[1] R. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *Affective Computing, IEEE Transactions on*, vol. 1, no. 1, pp. 18–37, 2010.

[2] H. Gunes and M. Piccardi, "Creating and annotating affect databases from face and body display: A contemporary survey," in *Proceedings of the International Conference on Systems, Man and Cybernetics IEEE SMC'06*, vol. 3, 2006, pp. 2426–2433.

[3] H. Gunes, B. Schuller, and M. Pantic, "Emotion representation, analysis and synthesis in continuous space: A survey," *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pp. 827–834, 2011.

[4] A. Abbasi, T. Uno, M. Dailey, and N. Afzulpurkar, "Towards knowledge-based affective interaction: situational interpretation of affect," *Affective Computing and Intelligent Interaction*, pp. 452–463, 2007.

[5] E. Douglas-Cowie, R. Cowie, I. Sneddon, C. Cox, O. Lowry, M. Mcrorie, J. Martin, L. Devillers, S. Abrilian, A. Batliner, and Others, "The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data," *Affective computing and intelligent interaction*, pp. 488–500, 2007.

[6] S. Afzal and P. Robinson, "Natural Affect Data - Collection & Annotation in a Learning Context," in *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*. IEEE, 2009, pp. 1–7.

[7] S. Wang, Z. Liu, S. Lv, Y. Lv, and G. Wu, "A natural visible and infrared facial expression database for expression recognition and emotion inference," *Multimedia, IEEE*, vol. 12, no. 7, pp. 682–691, 2010.

[8] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A Multi-Modal Affective Database for Affect Recognition and Implicit Tagging," *Affective Computing,*, pp. 1–14, 2011.

[9] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor," in *3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009)*. Iet, 2009.

[10] E. S. Dan-Glauser and K. R. Scherer, "The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance." *Behavior Research Methods*, vol. 43, no. 2, pp. 468–477, 2011.

[11] J. Hewig, D. Hagemann, J. Seifert, M. Gollwitzer, E. Naumann, and D. Bartussek, "A revised film set for the induction of basic emotions," *Cognition & Emotion*, vol. 19, no. 7, pp. 1095–1109, Nov. 2005.

[12] M. A. Quiros-Ramirez and T. Onisawa, "Assessing Emotions in a Cross-Cultural Context," in *Proceedings of the International Conference on Systems, Man and Cybernetics IEEE SMC'12*, 2012.