Int. J. Mach. Learn. & Cyber.
DOI 10.1007/s13042-013-0192-2

ORIGINAL ARTICLE

Considering cross-cultural context in the automatic recognition of emotions

Maria Alejandra Quiros-Ramirez · Takehisa Onisawa

Received: 15 January 2013 / Accepted: 9 August 2013

AQ1 **Abstract** Automatic recognition of emotions remains an ongoing challenge and much effort is being invested towards developing a system to solve this problem. Although several systems have been proposed, there is still none that considers the *cultural context* for emotion recognition. It remains unclear whether emotions are universal or culturally specific. A study on how culture influences the recognition of emotions is presented. For this purpose, a multicultural corpus for cross-cultural emotion analysis is constructed. Subjects from three different cultures—American, Asian and European—are recruited. The corpus is segmented and annotated. To avoid language artifacts, the emotion recognition model considers facial expressions, head movements, body motions and dimensional emotions. Three training and testing paradigms are carried out to compare cultural effects: intra-cultural, cross-cultural and multicultural emotion recognition. Intra-cultural and multicultural emotion recognition paradigms raised the best recognition results; cross-cultural emotion recognition rates were lower. These results suggest that emotion expression varies by culture, representing a hint of emotion specificity. AQ2

Keywords Affect · Culture · Universality · Specificity · Emotional corpus

M. A. Quiros-Ramirez (✉) · T. Onisawa
Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8573, Japan
e-mail: aqira.rei@gmail.com
URL: http://fhuman.esys.tsukuba.ac.jp

1 Introduction

The desire to create applications and devices that better support human needs is evident in fields of study such as human computer interaction (HCI). Considering the context in which users interact is key to creating a human-like interaction [10]. Including cultural context in the systems that support users yields better results in the interaction between humans and computers; thus, adding cultural context can generate applications with worldwide scope [1, 11, 15].

One way to bridge the breach between human and computers is to provide tools that understand the internal mental state of the users [18]. Understanding the emotional state of a user allows the machine to modify its responses accordingly. This new paradigm introduces the possibility of changing the current character of interaction in which the user is typically expected to adapt to the computer instead of the opposite, ideal way.

In contrast to human to compute interaction, in human-to-human communication one person understands another's emotions through different hints such as facial expressions, body language and vocal pitch [17]. An individual sends these signals unconsciously as part of the communication process and the people he or she is interacting with receives them along with other elements of the communication such as speech. Emotional information is crucial for successful communication. Using the same type of hints, a machine would be able to understand the internal emotional state of its users [18].

In emotion theory, the effect of cultural context is in sending and receiving subtle emotional cues is still an open question [20]. Several psychological studies suggest that culture plays a very important role in the mutual understanding of emotions.

Journal : Large 13042 | Dispatch : 10-8-2013 | Pages : 9
Article No. : 192 | □ LE | □ TYPESET
MS Code : JMLC-D-13-00012 | ✔ CP | ✔ DISK

The field of emotion recognition has been advancing quickly, yet the inclusion of the cultural aspect is still missing from currently available emotion recognition systems [8]. It is our purpose to further study the effect of the cultural aspect in the understanding of emotions, both within and outside of a culture.

1.1 Related emotion recognition research

A prototypical emotion recognition system is developed by training a system with several subjects' emotional reactions. There are single or multiple cue emotion recognition systems. A *single cue* emotion recognition system is trained focusing in one of the physiological hints. For example, a single cue emotion recognition system may be based on facial expressions only. A *multiple cue* emotion recognition system mixes several hints, such as facial expression and voice. Besides audiovisual hints, other physiological signals such as brain waves or skin conductivity can also be used to train the emotion recognition system.

Although the *multimodal* (multiple cue) emotion recognition systems and the use of different emotion theories (e.g. categorical and dimensional [17]) continue to be popular among researchers, very few efforts have been made to include contextual information such as cultural context in interaction systems.

Available emotion recognition systems that include cultural context are based mostly on a single cue, for example, body [16] or voice [14]. Up to now, these systems are focused on categorical classification of emotions, e.g., happy, sad, angry.

One study on culture dependence in body motions presented in the work of Kleinsmith and colleagues [16] compares three cultures: Japanese, Sri Lankan and Latin American. The corpus used for this study consists in posed body motions from 13 different subjects. Another study based on speech analysis [14] compares three cultures: European, American and Asian. Kamaruddin and colleagues use utterances obtained from TV shows in three different languages, each one representing one of the cultures being compared. These studies presented evidence of cultural specificity for emotion classification for the cultures studied.

The design of a new database for cross-cultural emotion recognition studies has been published recently [2]. It includes audiovisual information and its main focus is gesture analysis. The interactions in this database are acted and the three groups compared are all European.

1.2 Cross-cultural emotion recognition issues

Universality or specificity of emotions has been debated since the times of [5]. *Universality* of emotions suggests that emotions can be recognized regardless of the cultural backgrounds of the sender and receiver. This means that even if two people belong to different cultures, they would each be able to understand what emotion is being transmitted by the other based on visual and auditory cues. Ekman's [7] multicultural studies lead to the idea that there are six basic universal emotions.

On the other hand, *specificity* of emotions suggests that emotions are expressed and interpreted differently across cultures. Russell's work presents strong evidence disproving Ekman's theory of cultural universality [22]. Recent evidence in emotion perception questions the universality of facial emotions [13].

Although the question of cultural universality or specificity of emotion has been a hot topic for several decades, today it remains without a definite answer. Most of the work done to disentangle this question focuses on facial cues and utilizes discrete categorization of emotions.

The first issue for cross-cultural emotion studies is the lack of a common corpus that can be used for analysis, modeling, training and testing. There are very few open emotion databases [8] and none of these are constructed for the purpose of cross-cultural comparisons. Developing a cross-cultural corpus poses its own challenges: from basic and important points such as ways of gathering subjects from different cultures to complex points such as designing tasks simultaneously suitable for different cultures.

As mentioned in the previous section, most of the research done in this topic is based on single cue analysis. Scherer [20] describes expressions of emotion as a mix of psychobiological, sociocultural and epochal factors. His study presents evidence on the ongoing debate regarding cultural universality and specificity of emotions. His findings suggest that emotion encoding and decoding depend on the context of the interaction and suggests *multimodality* to more deeply study the question of cross-cultural emotions.

Working with multiple cultures might imply working with several languages as well. This is another unresolved issue at the time of the interaction and analysis. Haid and colleagues describe emotion words as *poor anchors* for cross-cultural comparisons [9]. Looking beyond the six most common emotions is suggested for comparing different cultures.

1.3 Scope of this research

Cultural universality or specificity of emotion remains an undecided issue. Several HCI studies have shown that considering the cultural context of computer user leads to better interaction results. Our purpose in this paper is to analyze and compare emotion expression and recognition

Journal : Large 13042 | Dispatch : 10-8-2013 | Pages : 9
Article No. : 192 | LE | TYPESET
MS Code : JMLC-D-13-00012 | CP | DISK

between different cultures and provide hints towards universality or specificity of emotions.

As a foundation for our study, we have prepared an emotion dataset of subjects from multiple cultures. This corpus is labeled at a feature and emotion level. Three cultures–American, European and Asian–considered. Our models analyze multiple cues in a dimensional categorization of emotions. We perform different combinations in training and testing models to further understand the effect of culture in the recognition of emotions.

The paper is organized as follows. Section 2 explains the proposed steps in developing an emotion recognition model and to compare the effect of culture. In Sect. 3, the construction of the emotion corpus is explained. Section 4 discusses the annotation of the corpus of features and emotions. The development of our emotion recognition experiments with different cultural combinations for training and testing is presented in Sect. 5. Finally a discussion and conclusions are presented in Sect. 6.

2 Proposed methodology

Based on the issues described in Sect. 1.2 there are five general requirements for a model that considers the cross-cultural context.

(a) *Cross-cultural corpus*: In order to perform emotion recognition analysis considering cultural differences, it is necessary to design an emotion recognition experiment using an emotional corpus built with a focus on culture. Due to the lack of databases with the features required for this study, it is necessary to construct a corpus with emotional expressions and interactions to use as training and testing material for the experiment. This corpus needs to include people from different cultural backgrounds, and the interactions need to be natural. Conventional studies work with posed or acted interactions. Previous research have shown that when a person acts or poses an emotion, the result tends to differ from natural emotions in at least two points: the timing and synchronization between features and motions tend to be wrong, and the expressions exaggerated and, based on stereotypes of how the posed expression should look [12, 25]. For our current purpose, we want to avoid these issues. Thus, we prepared an emotional corpus with subjects from different nationalities, interacting in situations that elicit emotional reactions in the participants. The expressions that appeared during such emotional interactions were recorded.

(b) *Multimodality:* The model requires analysis of several cues to further study the effect of culture on different audiovisual expressions. Facial expressions, head motions and body movements were considered as three different types of cues.

(c) *Theories of emotion:* Even though most emotion research revolves around the six basic emotions proposed by [7], this study utilizes dimensional categorization of emotions [23]. Two dimensions are considered: *valence*, which means how positive or negative an emotion is; and *arousal*, which represents the intensity of this emotion.

(d) *Language:* The assessment of emotions forces subjects to assign linguistic symbols to their feelings. This is a point of bias in a cross-cultural context. The use of dimensional categorization of emotions diminishes the effect of linguistics. Besides assessment, stimuli that require deep understanding of language could bias the interaction as well. To avoid such bias, pictures are used as stimuli for the experimental interactions to record the emotion corpus.

(e) *Cultural comparisons:* The final goal of the corpus construction and experiment is to compare the emotional expressions among different cultures. For this purpose, a model for each culture is prepared. Next, the recognition is tested first within each culture and then across cultures (e.g. American model tested with European population). Finally, a global model is prepared and tested. Having three different models, differences and similitudes can be observed.

3 Emotional corpus construction

The corpus was constructed following the guidelines of multiple cultures, natural emotions (through emotion elicitation), and language independence. The details are explained below.

3.1 Participants and cultural classification in this study

Eight participants with different cultural backgrounds were recruited from the University of Tsukuba. The participants' countries of origin are Jamaica, France, Costa Rica, India, Spain, Brazil, Japan and Hungary.

Some psychological studies in emotion classify cultures into two groups: Western cultures and Oriental cultures. However no standard is defined for this group division. In this study our cultures are separated heuristically by geographical regions into three main groups: America, Europe and Asia.

Journal : Large 13042 Dispatch : 10-8-2013 Pages : 9
Article No. : 192 ☐ LE ☐ TYPESET
MS Code : JMLC-D-13-00012 ☑ CP ☑ DISK

3.2 Emotion elicitation

The emotional corpus is essential to developing correct emotion recognition models. The ecology of the corpus is important to reflect real behavior, not only for the analysis of cultural context, but also for future use in real life scenarios.

Pictures were used to elicit emotions from the participants. The pictures used were obtained from the GAPED database [6]. The pictures available in this database are emotionally loaded, and they are expected to influence the watchers emotional state. The information about *valence* and *arousal* of each picture is available within the database, which makes them suitable for our study. The contents of the pictures range from pleasant images (for example cute animals or babies) to unpleasant images (like spiders and gross situations).

During the experiment, each participant was invited to enter the experimental room and sit in a chair placed one meter away from the monitor where the images were displayed. Participants were then instructed to watch the pictures that were automatically displayed on the screen.

After observing a picture, the participant was asked to assess his or her own emotional state using a five point scale (from negative to positive, zero being neutral). Two high definition cameras synchronized to each other recorded the whole interaction. The first camera was focused on the face of the participant and the second on the full body.

The pictures were presented in a random order for 5 s each, with a grey screen displayed for 3 s between pictures to let the participant rest. In total each participant observes 20 pictures: 8 positive, 8 negative and 4 neutral.

Each session was recorded continuously. The methods presented in [21] were used to segment the recording and obtain the regions of interest corresponding to the fragments in which the participant was exposed to the picture.

Examples of still shots of the videos recorded with the cameras can be seen in Figs. 1 and 2. Figure 1 presents an expression obtained from the full body camera, when the participant watched a negative image. Figure 2 illustrates, on the top row, responses of the participants while watching negative images, and in the bottom row, responses while watching positive images.

It is important to remark that the information of the pictures is not used further in the study. The only purpose of the pictures is to create some feeling or reaction in the participants. After the participation of the subjects is recorded, there is no relation between the contents of the pictures and the emotional models we construct.

4 Corpus annotation

The corpus consists of 160 segments of video with the emotional responses of the participants as they watched the pictures. Each segment last 5 s and is classified into American, Asian or European. At this stage, it is necessary to annotate features and emotion information (which will be referred to as "emotion label" in this paper) before training the model. An *Emotion label* refers to the real information on emotions inside each segment of interaction.

Fig. 1 Full body high definition camera. Natural response elicited by a European participant (from Hungary) while observing a negative picture

4.1 Feature annotation

Twenty-nine features were labeled for face, head motions and body movements. The features were chosen considering the frequency of movement among all participants. A feature is considered significant if it is observed more than five times in at least two independent participants from any cultural group.

Facial features: Inner eyebrows up, outer eyebrow raiser, eyebrow lowerer, frown, eyelid tightener, eyelids towards each other, multiple blinks, smile, laugh, abnormal breathing, nose wrinkle, jaw drop, lip pressor, lip suck, lip corner puller, lip corner depressor, jaw sideways, swallow, chin raiser.

Head features: move head, move head away, nod, say no, tilt head.

Body features: move finger up and down, move hands, touch or scratch with the hand, press hand, move leg.

Journal : Large 13042 Dispatch : 10-8-2013 Pages : 9
Article No. : 192 □ LE □ TYPESET
MS Code : JMLC-D-13-00012 ☑ CP ☑ DISK

Fig. 2 Face focused high definition camera. On the *top row*, reactions of American (from Costa Rica), Asian (from India) and European (from Hungary) participants, respectively, while they watched negative images. On the *bottom row*, reactions of European (from France), American (from Brazil) and Asian (from India) participants while they watched positive images

4.2 Emotional annotation

Emotional annotation refers to the emotion label assigned to an observed interaction. This label is considered the "real" emotion that the participant in the video segments is feeling. Labels are necessary to train a model and perform associations between interactions and elicited feelings.

There are several techniques to assign emotion labels to the segments. We chose to assign the participant's self-report of emotion as emotion label. Self-report of emotions is considered valid in the cases when subjects report "currently experienced" emotions [17]. In our experiments, the participant's emotion was reported immediately after he or she observed the image, thus this labeling technique is appropriate for our investigation.

Since our interest in this study is to analyze the *expression* of emotions in different cultures, not *understanding* of emotions, we do not include the emotional judgment of third parties [20].

5 Emotion recognition experiment

In this section, the different experimental paradigms are explained. In each set up, the collected data was arranged according to the cultural comparison purpose.

5.1 Training vectors and chosen classifier

Each group of training vectors is composed of the annotated features (presented in Sect. 4.1). This information is matched to the emotional label. For this study two emotion labels representing valence are used: positive valence and negative valence. The following expression describes the type of vector used for training each model:

$$E_{ij} = (f_{ij1}, f_{ij2}, \ldots, f_{ij19}, h_{ij1}, h_{ij2}, \ldots, h_{ij5}, b_{ij1}, b_{ij2}, \ldots, b_{ij5})$$

where i represents the participant's ID, j the number of picture the participant observed, E is the reported emotion, f_{ijk} refers to each labeled facial feature ($k = 1,2,\ldots,19$), h_{ijl} indicates the head motions ($l = 1,2,\ldots,5$), b_{ijm} represents the body movements ($m = 1,2,\ldots,5$). Vectors are chosen for training and testing the models based on participant i's culture.

There are several common classifiers used for emotion recognition tasks [8]. We chose *Support Vector Machines* (SVM) to train each model. An implementation of SVM from SVM-KM Toolbox [3] with *Gaussian* kernel was employed for training and testing. A leave one out cross-validation (LOOCV) procedure was selected in order to use all the vectors for training and testing each model, using each vector as an independent test exactly once.

LOOCV consists of training a model with n-1 vectors and testing it with the remaining one, where n represents the total amount of vectors. The training is performed n times, testing a different vector each time. LOOCV has been chosen instead of data partitioning to avoid biasing the model towards specific participants.

Three emotion recognition experiments were carried out, testing recognition: within a culture, across two cultures and mixing the three cultures together.

5.2 Intra-cultural emotion recognition

Intra-cultural emotion recognition refers to emotion recognition inside a single culture. That is, the model is trained and tested within the same culture. An intra-cultural emotion recognition experiment was performed for each of the three cultures in this study. It is necessary to examine the recognition results within a culture before proceeding to analyze cross-cultural scenarios. LOOCV was utilized in this case. Figure 3 shows the confusion matrices for each culture: American, Asian and European.

Journal : Large 13042 | Dispatch : 10-8-2013 | Pages : 9
Article No. : 192 | LE | TYPESET
MS Code : JMLC-D-13-00012 | CP | DISK

Fig. 3 Recognition results for intra-cultural emotion recognition of the three cultures: American, Asian and European. *Vertical axis* represents the original emotion label. The *horizontal axis* represents the predicted emotion

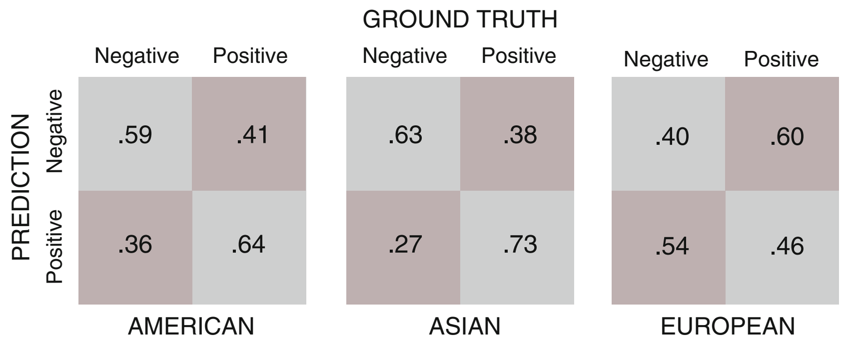

Table 1 Recognition results for positive and negative valence and general accuracy of emotion recognition per culture

Culture	Positive	Negative	Accuracy
American	0.64	0.59	0.62
Asian	0.73	0.63	0.68
European	0.46	0.40	0.43

Table 1 presents a summary of the recognition rates and recognition accuracy per culture in the intra-cultural emotion recognition paradigm. While participants from American and Asian cultures achieved a reasonable accuracy rate, participants from European cultures achieved a very low accuracy rate. In all three cultures it was easier for the model to recognize positive expressions of emotion than negative ones. Participant from Asian cultures achieved the best recognition accuracy among the three culture groups.

5.3 Cross-cultural emotion recognition

This recognition model was trained using one culture and tested within a different culture, for example, train with American data and test with Asian data. The resulting confusion matrices can be observed in Fig. 4.

A summary of the recognition accuracy rates for each combination of cultures is presented in Table 2. The accuracy of emotion recognition decreased in most cases. For the Asian model, the overall recognition accuracy decreased as well in comparison with the initial intra-cultural recognition results. European model elicited rather good recognition rate.

5.4 Multi-cultural emotion recognition

Finally, a multicultural emotion recognition model (combining all cultures) was performed. This training served to improve understanding of partial cultural dependencies and to analyze the effect of mixing several cultures in a single model. Figure 5 shows the recognition results.

In comparison with the cross-cultural set up, the multicultural model raised better recognition results. The cross-cultural set up represented a model that has a single specific cultural background (for example, American model). On the other hand, the multicultural set up had a combined culture background; therefore it is not limited to expressions from a single culture as the cross-cultural model.

The multicultural model recognizes emotions from different participants despite their cultural background. This model has the expressivity knowledge from three different cultures, thus it is possible to match common expressions regardless culture. Such knowledge is not possible to obtain from a model trained based in a single culture. The recognition rate of this model suggests that there are common non-contradicting expressions between the three cultures.

6 Discussion and conclusions

A study on the influence of cultural context in emotion recognition was presented in this paper. To understand the influence of culture in emotion recognition, a multi-cultural corpus was constructed, segmented and labeled considering facial expressions, head movements and body motions.

Three cultures: American, Asian and European were considered. To avoid language artifacts the study was based in dimensional emotions, using emotional *valence* for the experiments. Participants of the data collection experiment were asked to rate their own emotions while watching emotionally loaded images. This information was used as *emotion labels*. The experimental interactions were recorded by two synchronized high definition cameras, one focused in body and the second one focused on the face.

6.1 Intra-cultural vs. cross-cultural emotion recognition

In the intra-cultural setting, the same cultural corpus was used to train and test an emotion recognition model. This set up is used to understand the capability of the model to

Journal : Large 13042 | Dispatch : 10-8-2013 | Pages : 9
Article No. : 192 | LE | TYPESET
MS Code : JMLC-D-13-00012 | CP | DISK

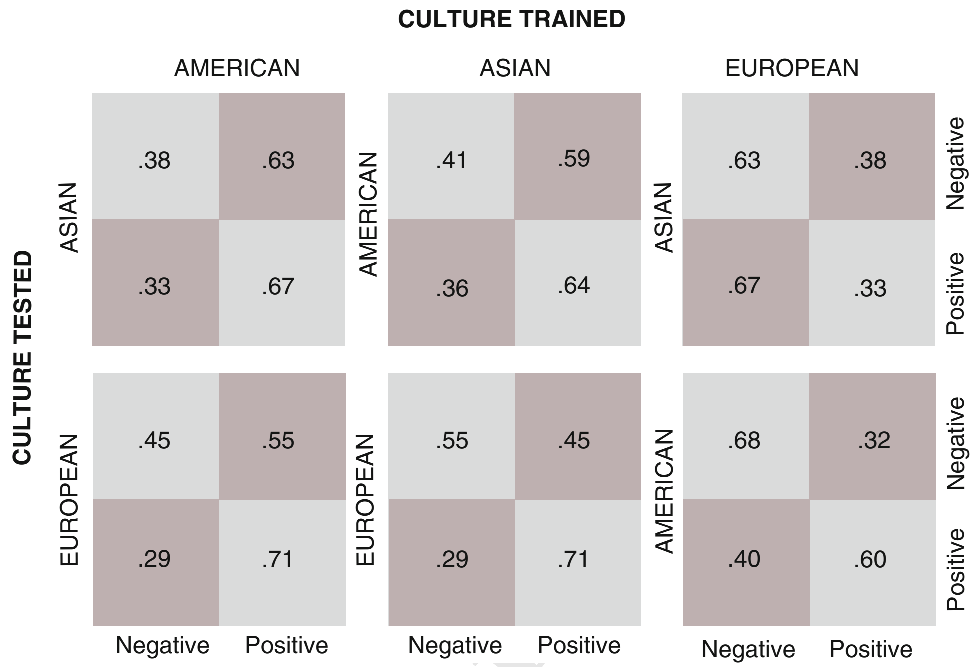

Fig. 4 Recognition results for cross-cultural emotion recognition among cultures. The *first column* was trained with the American corpus, the second one with the Asian and the third one with the European corpus. Each trained model was tested with the two remaining cultures. The *vertical axis* represents the original emotion label while the *horizontal axis* represents predicted emotion

Table 2 Accuracy rates per training/testing trial in the cross-cultural recognition paradigm

Tested culture	Trained culture		
	American	Asian	European
American		0.46	0.48
Asian	0.52		0.64
European	0.58	0.63	

The *columns* represent the corpus used for training and the *rows* represent the corpus used for testing

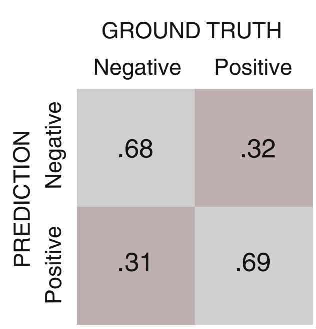

Fig. 5 Recognition results for multi-cultural emotion recognition. The vectors from the three cultures are used for both training and testing, investigating the effect of mixing the three corpuses without considering the cultural background

recognize positive and negative emotions within a single culture.

The results in Table 1 indicate that the model is able to recognize the emotions for the American and Asian cultures with good accuracy. In the case of the European model, the recognition accuracy is low. Further analyzing this issue, variability inside the different cultures that are represented in our corpus by the label "Europe" was recognized as a possible reason for the low recognition accuracy of the European model. The corpus is based on subjects from Spain, France and Hungary. Although the countries belong to the same continent, their cultural backgrounds are quite different; behavioral expressivity seems to be different as well. From this issue, we can understand that it is necessary to define a more sensitive cultural filter. Continental grouping seems to be too broad to reflect the nuances of the different cultural populations.

Overall, the cross-cultural emotion recognition model had lower accuracy results compared with the intra-cultural models. Such decrease in the result indicates that *a model trained to understand emotional expressions from a specific culture fails to recognize with the same accuracy emotional expressions from subjects of a different culture*. This finding suggests *cultural specificity* of expression of emotions. Within a *universal* context, a model trained with subjects from a single culture should not suffer a recognition rate decrease when new subjects are tested, despite their cultural background.

Based on Table 2, it is possible to understand that expressions of participants in the European corpus had low similarity to those participants in the American corpus. On the other hand, there seems to be closer emotion expressivity between the European and Asian. Nevertheless, it is still necessary to refine both corpuses before reaching a conclusion regarding the relative similarity amongst the three cultural groups.

Journal : Large 13042 Dispatch : 10-8-2013 Pages : 9
Article No. : 192 LE TYPESET
MS Code : JMLC-D-13-00012 CP DISK

6.2 Cultural-context consideration vs. cultural blind emotion recognition

The third emotion recognition paradigm consisted on a multi-cultural emotion recognition experiment. The multi-cultural set up represents an emotion recognition model that does not consider culture, thus is culturally blind. The purpose of this set up is to understand the effect of emotion recognition models with full knowledge of the expressivity of the different cultures. In this experiment there is no direct knowledge of the cultural background of each participant.

The multi-cultural emotion recognition model achieved good recognition results in general. This finding shows that even though our results in intra-cultural and cross-cultural experiments indicate *specificity* in the human expressions, when a model is trained with a variety of cultures good results can be obtained and similarities between the expressions of the different cultures can be found.

Despite obtaining good results by mixing the three cultures, based on the results of Sects. 5.3 and 5.4, it is recommended to train emotion recognition models including the cultural background of the subjects. Failing to add cultural context will decrease the model's recognition rate when a new subject from a different culture is presented to the model. This situation can be established observing the recognition accuracy decrease from using Table 1 (intra-cultural test) and Table 2 (cross-cultural test, i.e. new culture tested in the model).

6.3 Final remarks

The results obtained from the comparison of different emotion recognition models that consider cultural context suggest that culture influences the expression and recognition of emotions. This finding is a hint towards demonstrating the cultural *specificity* of emotions.

The comparison of models that consider subject's cultural background and models which do not consider it resulted in similar recognition rates. However, further analysis on a real type scenario of cultural mixing is required before supporting universality of emotional expressions since, as shown in the results of Table 2, testing a model with subjects from a new culture will decrease the accuracy rate.

Since cross-cultural recognition results indicate emotion recognition has cultural dependency, it is predicted that a cultural inclusive recognition model should yield better results than a model that do not consider culture. Further and more detailed experiments will be carried on to study this effect.

The results of the intra-cultural scenarios are quite similar to the results reported in previous studies on multimodality of emotions [24]. Based on our results, positive emotions seem to be easier to recognize than negative emotions. This result is consistent with findings of other groups [2, 4].

A key point to an emotion cross-cultural study is the scope of the meaning *culture* and the choice of filter to separate participants into cultural groups. Our current filter is based on the continental geographical separation. For further studies, it is necessary to consider not only a geographical background but also a sociocultural filter should be included in the participant grouping. It is also necessary to investigate the effect that cultural exchange has on the subjects. Pursuing this question might help in the understanding of cultural bias causes, and what is the effect of sociological and physical traits in the cross-cultural emotion understanding.

Future work includes the construction of a more specific corpus with subjects from closer cultural backgrounds to improve recognition results and boost comparison capabilities. Even though the corpus introduced in this study is robust enough to provide a cultural comparison basis, our results show weakness on the European sub-corpus. It is necessary to extend further the corpus to allow a better cultural filtering to avoid the issues encountered in the intra-European emotion recognition model. Other improvements of the corpus should include different varieties of emotional labels.

In conclusion, the expression and understanding of emotion seems to be influenced by the cultural backgrounds of the people interacting. Understanding the effect of culture in the multimodal recognition of emotions can improve the emotion recognition systems and also enhance interaction between humans and computers in the future.

References

1. Anacleto JC, Pinatti de Carvalho AF (2008) Improving human-computer interaction by developing culture-sensitive applications based on common sense knowledge. In: Pinder S (ed) Advances in Human-Computer Interaction. I-Tech, Vienna, pp 1–30
2. Caridakis G, Wagner J, Raouzaiou A, Lingenfelser F, Karpouzis K, Andre E (2012) A cross-cultural, multimodal, affective corpus for gesture expressivity analysis. J Multimodal User Interfaces. doi:10.1007/s12193-012-0112-x
3. Canu S, Grandvalet Y, Guigue V, Rakotomamonjy A (2005) SVM and kernel methods matlab toolbox. Percept Syst Inf 2:2
4. Cho J, Kato S, Itoh H (2009) Comparison of sensibilities of Japanese and Koreans in recognizing emotions from speech by using Bayesian networks. International Conference on System Man and Cybernetics, SMC 2009, pp 2866–2871
5. Darwin C (1872) The expression of the emotions in man and animals. John Murray, London
6. Dan-Glauser ES, Scherer KR (2011) The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. Behav Res Methods 43:468–477

Journal : Large 13042 | Dispatch : 10-8-2013 | Pages : 9
Article No. : 192 | ☐ LE | ☐ TYPESET
MS Code : JMLC-D-13-00012 | ☑ CP | ☑ DISK

7. Ekman P (1994) Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. Psychol Bull 115:268–287
8. Gunes H, Schuller B (2013) Categorical and dimensional affect analysis in continuous input: current trends and future directions. Imag Vis Comput 31(2):120–136
9. Haid J, Keltner D (1999) Culture and facial expression: open-ended methods find more expressions and a gradient recognition. Cogn Emot 13:225–266
10. Harper R, Rodden T, Rogers Y, Sellen A (2008) Being human: human-computer interaction in the year 2020. Microsoft Research Ltda, England
11. Heimgartner R, Kindermann H (2012) Revealing Cultural Influences in Human Computer Interaction by Analyzing Big Data in Interactions. Lecture Notes in Computer Science pp 572–583
12. Hoque M, Picard R (2011) Acted vs. natural frustration and delight: many people smile in natural frustration. IEEE International Conference on Automatic Face and Gesture Recognition doi:10.1109/FG.2011.5771425
13. Jack RE, Blais C, Scheepers C, Schyns PG, Caldara R (2009) Cultural confusions show that facial expressions are not universal. Curr Biol 19:1543–1548
14. Kamaruddin N, Wahab A, Quek C (2012) Cultural dependency analysis for understanding speech emotion. Expert Syst Appl 39:5115–5133
15. Kamppuri M, Bednarik R, Tukiainen M (2006) The expanding focus of HCI: case culture. Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles pp 405–408
16. Kleinsmith A, De Silva P, Bianchi-Berthouze N (2006) Cross-cultural differences in recognizing affect from body posture. Interact Comput 18:1371–1389
17. Mauss IB, Robinson MD (2009) Measures of emotion: a review. Cogn Emot 23:209–237
18. Picard RW (1997) Affective Computing. The MIT Press, Massachusetts
19. Scherer KR (2005) What are emotions? And how can they be measured? J Soc Sci Inf 44:695–729
20. Scherer KR, Clark-Polner E, Mortillaro M (2011) In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. Int J Psychol 46:401–435
21. Quiros-Ramirez MA, Onisawa T (2012) Assessing emotions in a cross-cultural context. System Man and Machine pp 2967–2972
22. Russell JA (1994) Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. Psychol Bull 115:102–141
23. Russell JA (1980) A circumplex model of affect. J Pers Soc Psychol 36:1161–1178
24. Soleymani M, Lichtenauer M, Pun J, Pantic M (2011) A multimodal affective database for affect recognition and implicit tagging. Int J Affect Comput pp 1–14
25. Wilting J, Krahmer E, Swerts M (2006) Real vs. acted emotional speech. Ninth International Conference on Spoken Language Processing. pp 17–21

Journal : Large 13042	Dispatch : 10-8-2013	Pages : 9
Article No. : 192	☐ LE	☐ TYPESET
MS Code : JMLC-D-13-00012	☑ CP	☑ DISK