Songklanakarin J. Sci. Technol.
35 (4), 461-468, Jul. - Aug. 2013

SJST SONGKLANAKARIN JOURNAL OF SCIENCE AND TECHNOLOGY
http://www.sjst.psu.ac.th

Original Article

Use of septum as reference point in a neurophysiologic approach to facial expression recognition

Igor Stankovic* and Montri Karnjanadecha

Department of Computer Engineering, Faculty of Engineering,
Prince of Songkla University, Hat Yai, Songkhla, 90112 Thailand.

Received 29 March 2012; Accepted 7 May 2013

Abstract

Studies of basic facial expression recognition have always shown different recognition rates for different emotional expressions. Happiness and surprise detection easily exceed 90% recognition rates, while other basic emotions (i.e. sadness, anger, fear, and disgust) produce much lower rates. In this paper we present a simple approach for reducing this gap, increasing the recognition rates for the other four basic emotions, based on more closely analyzing the displacements of extracted facial landmarks. With the use of a reference point in all the image sequences that represent an emotional expression, calculations become more resistant to head movement errors, thereby reducing recognition errors. Also, emotions are expressed differently across time so, besides first and peak frames, we also analyze the frames one third and two thirds along each image sequence. Our results show great improvements in recognition, yielding total accuracies of 96.8%, and lowering the recognition gap between facial expressions.

Keywords: facial expression recognition, emotion recognition, support vector machines, the extended Cohn-Kanade database

1. Introduction

For several decades, the field of facial expression recognition has been an important research area, especially in human-computer interaction. Ekman and Friesen (1971) discussed six emotions: happiness, sadness, surprise, anger, fear, and disgust, which became the “basic” emotions, used in much related research since. In one of their later studies, Ekman and Friesen (1977) defined facial action coding system (FACS) by closely examining facial movements. They concluded that every emotion facial expression is a combination of the movements of several facial muscles. Each basic facial movement is coded as an action unit (AU), so that every facial expression can be represented by a group of several AUs.

Facial expression recognition research can be divided roughly into three parts: 1) facial feature extraction, 2) the examination of the changes of those extracted features, and 3) the classification of the gathered information.

Tian *et al*. (2001) used permanent features, such as optical flow, Gabor wavelets, and multi-state models, together with canny edge detection as transient features. Dornaika and Davoine (2008) chose a candidate face model to track features, while Lucey *et al*. (2010) presented their baseline results in facial feature extraction by utilizing active appearance models (AAMs). Facial landmarks were extracted by Michel (2003) by employing an *Eyematic FaceTracker* application. Expressions were classified by calculating displacement vectors for each landmark between the first and peak frames in every expression sequences. In his paper, and in the study by Lucey *et al*. (2010), support vector machines (SVM) were used as classifiers, giving excellent results. Cohen *et al*. (2003) proposed a new multilevel architecture of hidden Markov models (HMM) for automatic segmentation and recognition of human facial expressions from video

* Corresponding author.
Email address: bizmut@neobee.net

sequences. They conducted their research with several classifiers, such as naive Bayes (NB), tree-augmented naive Bayes (TAN), HMM, multilevel HMM, and stochastic structure search (SSS), reaching a maximum accuracy of 83.3% with the TAN classifier and the Cohn-Kanade database. Sebe *et al*. (2007) utilized Bayesian nets, SVM, and decision trees for classification, and reached an accuracy of 93.6%.

There are several facial expression databases, including the MMI facial expression database (Pantic *et al*., 2005), the Japanese female facial expression (JAFFE) database, the Cohn-Kanade database (also known as the CMU Pittsburg database) (Kanade *et al*., 2000), and the improved Cohn-Kanade (CK+) database (Lucey *et al*., 2010). Many studies, such as Lucey *et al.* (2010), show excellent results for recognizing happiness and surprise, while the other four basic emotions (sadness, anger, fear, and disgust) have much lower results. This is probably due to the more extreme facial deformation and movements used to express happiness and surprise, making them easier to recognize. This recognition gap was stressed by many researchers, such as Bettadapura (2009), who call for more work towards recognizing all expressions with equal or similar accuracy.

Our research focuses on increasing the recognition rate by tracking facial landmarks extracted by Lucey *et al.* (2010). Our system calculates the displacement of each facial landmark to a frame's base-point in the first and peak frames. We chose the septum, the skin that separates the two nostrils, as the base-point. Our result is that our system is more resistant to head movement errors, thereby increasing the recognition accuracy. Emotions are expressed differently over time (Batty and Taylor, 2003) so, our system also utilizes the frames one third and two thirds along each time sequence. We employed a linear SVM classifier, utilizing leave-one-subject-out cross-validation approach, which proved to be both simple and effective for classifying facial expressions. Our experiments show excellent results with around 90% recognition accuracy for all expressions, yielding a total accuracy of 96.8%, thereby closing the recognition gap between the basic emotions. In Section 2, we present our methodology, and explain our experiments in Section 3. Our results are shown in Section 4, and conclusions made in Section 5.

2. Methodology

After the Cohn-Kanade (Kanade *et al*., 2000) database was released in 2000, it soon became one of the most frequently used databases for face recognition algorithm development and evaluation. To address a few concerns, the authors released the CK+ database in 2010 (Lucey *et al*., 2010). The number of sequences was increased by 22% and the number of subjects by 27%. The database was tested using AAMs and a linear SVM classifier (using a leave-one-subject-out cross-validation method) for both AU and emotion detection. The resulting emotion and AU labels, together with the extended image data and tracked facial landmarks, were made publicly available. Figure 1 shows the peak frames for the six basic facial expressions from the CK+ database. Using AAMs, 68 facial landmarks were extracted (Lucey *et al*., 2010), and utilized as the starting point for our experiments.

Our calculations were based on tracking the displacements of landmarks from the first neutral frame to the last peak frame, in an image sequence representing one emotional expression (see Figure 2). This method has been utilized in several related studies, such as Michel (2003) (see Figure 3), giving good results, but with high recognition accuracy differences between the basic facial expressions. Our approach tracks the displacements more closely, increasing the accuracy.

2.1 Head movements correction

Our first assumption was that part of the recognition error in tracking landmark displacements comes from head movements that occur between the neutral and the peak frames. Out of the 68 extracted facial landmarks (Lucey *et al*., 2010) we chose one landmark to present the base-point in every frame. This base-point should be fixed during emotion expressions, so it can be employed to detect head movement, without additional facial movements. Also, it should be

Figure 1. Peak frames of the six basic facial expressions (happiness, sadness, surprise, anger, fear, and disgust) from the CK+ database.

Figure 2. Example of an image sequence from CK+.

Figure 3. Natural and the peak frame producing vectors of displacements.

located at the middle of face, so that the calculations are equally sensitive to movements by all the other landmarks. In the six basic facial expressions, the nose region seems to move the least during facial expressions. Therefore, by being in the middle of the facial region, the point between the two nostrils (called the septum), was chosen as our base-point in each frame (see Figure 4). It corresponds to the 34[th] landmark in the standard group of 68 extracted landmarks. The displacements of other landmarks were calculated with reference to this base-point, using Euclidian distance:

$$d_i = \sqrt{(x_{i,P} - x_{34,P})^2 + (y_{i,P} - y_{34,P})^2} - \sqrt{(x_{i,N} - x_{34,N})^2 + (y_{i,N} - y_{34,N})^2} \quad (1)$$

d_i is the subtraction of two Euclidian distances for a landmark i. The first distance is from the base-point (the 34[th] landmark) to the landmark in the peak (P) frame, while the second distance subtracts the base-point from the landmark in the neutral (N) frame.

By introducing the base-point in each frame, our calculations became more resistant to head movements that may appear between the first and the peak frame. Figure 5 shows an example of feature vectors from previous method and from our proposed method. The two feature vectors at the top of the Figure 5, which represent anger and fear, were obtained by the previous method. They were classified in-

Figure 4. Landmarks with the base-point (the septum).

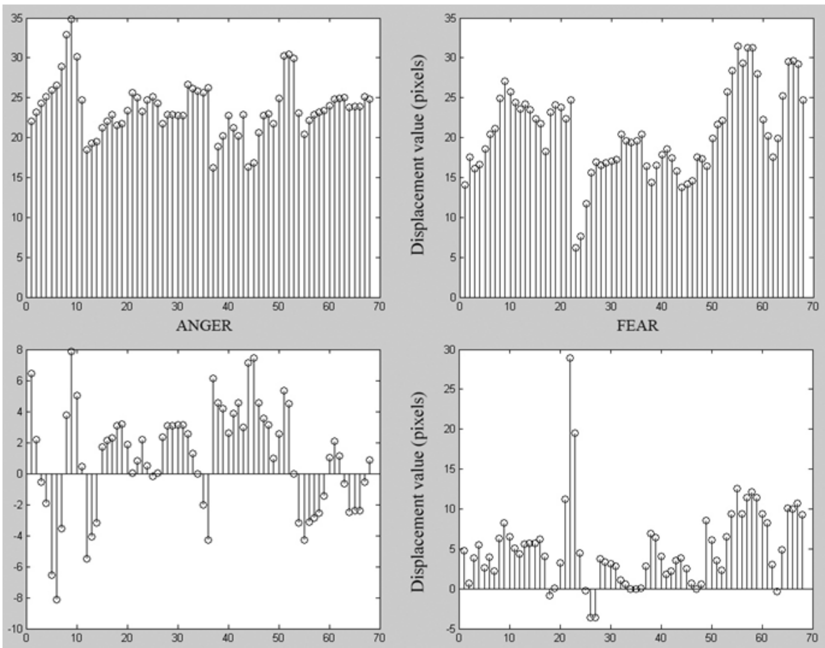

Figure 5. Two feature vectors for anger and fear using the previous method (top) have similar shape, leading to an incorrect classification. The same feature vectors, calculated using our proposed method (bottom) are classified correctly due to their much different shape.

correctly, because their feature patterns have a similar shape, which is confusing for a classifier. The same feature vectors, obtained using our proposed method, are displayed at the bottom of the Figure 5. After removing the feature calculation "noise", caused by head movements, the differences in these two vectors are much more obvious. Using our method, these two vectors were classified correctly.

2.2 Neurophysiologic approach

The important aspect of the speed with which facial emotions are processed and expressed has only recently been investigated in neurophysiology (Batty and Taylor, 2003). It seems that different emotions use different brain regions, so the time to process and express emotions differs. This difference is much more obvious when examining positive emotions (i.e. happiness, surprise) to negative ones (i.e. sadness, anger, fear), but it also differs from emotion to emotion. Since emotions are expressed over different durations, the time (the number of frames) needed for expressing an emotion in a time sequence was added to our system.

Furthermore, the movement of facial muscles is different for each emotion. For example, as shown here, http://face-and-emotion.com/dataface/emotion/x_happy.html, happiness is universally and easily recognized, and is interpreted as enjoyment, pleasure, and friendliness (Ekman *et al*., 2002). Happy expressions are frequently produced by people on demand in the absence of any real emotion, or to hide other emotions, or to deceive or manipulate. On the other hand, many cultures contain a strong censure against public displays of negative emotions, such as sadness and anger. Also, some emotions, such as fear, are not often seen in societies where personal security is typical.

As a consequence of neurophysiologic brain structure, social influences, and differences in usage frequency, different emotions trigger facial muscle movements in different ways over time. This observation is why our system examines time sequence frames at the one third and two thirds points of every sequence (see Figure 6).

Lucey *et al*. (2010) only examines the first (neutral) and the last (peak) expression frames, and does not take into account the facial changes that lead to the peak expression. However, middle frames were used during a visual inspection of the clip, to determine whether the expression is a good representation of an emotion. In our work, two frames, one third and two thirds along the image time sequence, proved an excellent way to distinguish precisely between emotions, and reduce the overall error (see Section 4). Similarly to the 68-feature vectors described early, the distances from the first middle frame (M1) and the neutral frame, as well as the second middle frame (M2) and M1 were calculated using the Euclidian distance formulas:

$$d1_i = \sqrt{(x_{i,M1} - x_{34,M1})^2 + (y_{i,M1} - y_{34,M1})^2} - \sqrt{(x_{i,N} - x_{34,N})^2 + (y_{i,N} - y_{34,N})^2} \quad (2)$$

$$d2_i = \sqrt{(x_{i,M2} - x_{34,M2})^2 + (y_{i,M2} - y_{34,M2})^2} - \sqrt{(x_{i,M1} - x_{34,M1})^2 + (y_{i,M1} - y_{34,M1})^2} \quad (3)$$

In Equation 2, $d1_i$ subtracts the Euclidian distance of each i landmark from the base-point (the 34[th] landmark) in M1 from the distance of the landmark from N. Equation 3 for $d2_i$ is similar but subtracts the distance from M2 and M1.

2.3 Movements in X– and Y–axes

After making our system more resistant to head movement errors and adding middle frames, it showed great improvements for recognizing all emotions, but errors were still present, especially when detecting sadness. Our data shows that happiness and sadness yields similar distances (from the neutral to the peak frames) for the points at the edge of the mouth. Unfortunately, those points are the most important for recognizing happiness, and so sadness can be confused with happiness. While smiling, edge points drastically change along the x-axis, with almost no changes in the y-axis. However, sadness shows small changes on both axes, which made the Euclidian distances for happiness and sadness at the edge points almost the same. To address this problem, we factored the landmark changes on the x- and y-axes into our calculations, using:

$$dx_i = x_{P,i} - x_{N,i} - (x_{P,34} - x_{N,34}) \quad (4)$$

$$dy_i = y_{P,i} - y_{N,i} - (y_{P,34} - y_{N,34}) \quad (5)$$

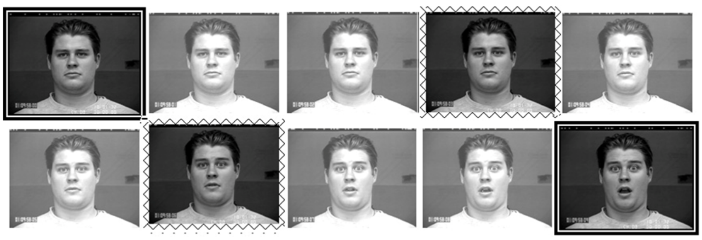

Figure 6. Example image sequence for expressing surprise taken from the CK+ database (straight line framed – the first and the peak frames; curved line framed – additional middle frames).

where dx_i and dy_i represent the subtract movements in the x- and y-axes of landmark i from the base-point (the 34th landmark) in the peak (P) frame and the neutral (N) frame. These feature vectors improve the results, giving perfect results for detecting happiness and sadness, and improving the recognition of the other emotions.

3. Experiments

The CK+ database (Lucey *et al*., 2010) includes 593 image sequences using 123 subjects. The sequences vary in duration (from 6 to 71 frames) and each one presents a subject's face from the first (neutral) frame to the peak formation of the given facial expression. An image sequence for a surprised expression is displayed in Figure 2.

Lucey *et al*. (2010) decided to evaluate the sequences by studying the middle frames in each image sequence. They concluded that only 327 of those 593 sequences represent a natural emotion expression. The other sequences, which failed their criterion, were discarded from their experiments. The final inventory of their selection process is given in Table 1.

Table 1. Frequency of the emotions in the CK+ database.

Emotion	Abbreviation	Number of sequences
Happiness	Hap.	69
Sadness	Sad.	28
Surprise	Sur.	83
Anger	Ang.	45
Fear	Fear	25
Disgust	Dis.	59
Contempt	Con.	18
Total	-	327

Our work focuses only on the six basic facial expressions, so contempt sequences are not included, resulting in 327–18=309 samples for our experiments. Our feature vector for each image sequence comprises 342 features comprised from the following:

- 68 features for the displacements (for each of 68 landmarks) between the neutral frame (N) and the peak frame (P).
- 68 features for the displacements between the frame at the one third point of the sequence (M1), and N.
- 68 features for the displacements between the frame at the two thirds point of the sequence (M2), and M1.
- 2 x 68 features for movements in the x- and y-axes from N to P.
- 1 feature for the number of frames in the sequence.
- 1 feature for the presence/absence of nose wrinkles (Lucey *et al*., 2010).

Figure 7 presents examples of the feature vectors for each basic emotion. Our feature dataset is represented as a 309×342 matrix, relating each of the sequences to their features. Finally, the process of collecting this feature matrix is explained in a pseudo-code (Figure 8).

In several other studies, such as Lucey *et al*. (2010) and Michel (2003), linear SVM has produced good results, and proved to be simple and effective for classifying facial expressions. Motivated by those studies, we tested our 309×

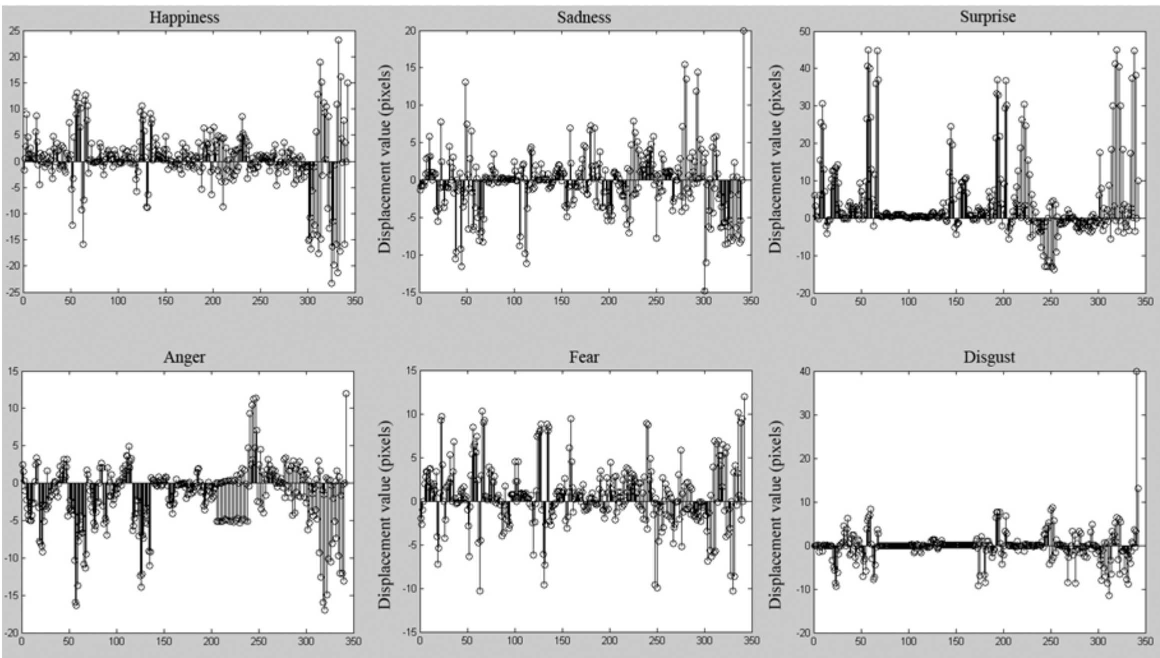

Figure 7. Example feature vectors for each basic expression.

```
FOR i = 1 TO No_of_image_sequences_in_CK+

      SEQ= image_sequence(i)          //i-th image seq from CK+
      FR = length(SEQ)                //no of frames in the SEQ
      N  = SEQ(1)                     //first (neutral) frame
      P  = SEQ(FR)                    //last (peak) frame
      M1 = SEQ(round(FR/3))           //frame at 1/3 of seq time
      M2 = SEQ(round(2*FR/3))         //frame at 2/3 of seq time

      FOR j = 1 TO No_of_facial_landmarks_in_frames //=68

         //diff of Euclidians to the base-point (34th landmark):
            //1. from N to P
            NtoP(j)    = Euclidian_distance(N(j),N(34))-
                         -Euclidian_distance(P(j),P(34))
            //2. from M1 to N
            M1toN(j)   = Euclidian_distance(M1(j),M1(34))-
                         -Euclidian_distance(N(j),N(34))
            //3. from M2 to M1
            M2toM1(j)  = Euclidian_distance(M2(j),M2(34))-
                         -Euclidian_distance(M1(j),M1(34))

         //movements in x- and y-axes from N to P
            X_NtoP(j)  = X(N(j))-X(P(j))  -  (X(N(34))-X(P(34)))
            Y_NtoP(j)  = Y(N(j))-Y(P(j))  -  (Y(N(34))-Y(P(34)))
      END FOR

      //wrinkle detector [as calculated in Lucey et al. (2010)]
      wrinkle = Wrinkle_detector(i)

      //342-feature-vector
      Feature_vector(i) =
           [ NtoP, M1toN, M2toM1, X_NtoP, Y_NtoP, FR, wrinkle ]
END FOR
```

Figure 8. Process of collecting our feature matrix.

342 dataset with a linear one-versus-all (i.e. anger versus not anger, happiness versus not happiness) multi-class SVM classifier, utilizing the leave-one-subject-out cross-validations method. Matlab's *libsvm* toolbox (Chang and Lin, 2011) was used in our experiments.

4. Results

Our main focus was improving recognition results for the six basic facial expressions: happiness, sadness, surprise, anger, fear, and disgust. Lucey *et al*. (2010), however, added contempt as a new emotion, and used 118 different training and test sets for emotion detection. Removing contempt from our experiments raised the recognition rate slightly, but our usage of bigger training and test sets, made the rates drop, producing results similar to those of Lucey *et al*. (2010).

The use of frame base-points made our system more resistant to head movement errors. Also, adding middle frames (at the one third and two thirds point of every sequence) increased the distinction between emotions. Furthermore, our observation of movements in the x- and y-axes reduced the confusion between some emotions. Finally, additional features that represent a sequence's duration (using the number of frames) for an emotion, and a nose wrinkle detector, as calculated by Lucey *et al*. (2010), improved our results by making a bigger difference between positive and negative emotions. A summary of our accuracy results are displayed in Table 2, with the rows and columns: happiness, sadness, surprise, anger, fear, and disgust, respectably in both directions. The main diagonal represents correctly classified samples (happiness classified as happiness, sadness classified as sadness, etc.), while the other fields represent the system error.

4.1 Result comparison

Table 3 compares the accuracy results obtained with our method and with those from several related papers. Visutsak (2005) uses the displacements of only eight points from the lower part of the face for classifying basic expressions. His results are good for detecting happiness and surprise, but his 8-feature vectors are not informative enough for the other four emotions, yielding lower results, and a total accuracy of 74.5%. Michel (2003) employed *Eyematic FaceTracker* application to extract 22 facial landmarks, but outperforms our results only for the recognition of surprise, due to his smaller training and test sets (20 samples per emotion). His final recognition rate is 86.3%. Sebe *et al*. (2007) present an emotions database composed from spontaneous reactions caught using hidden cameras. They utilized

Table 2. Summarized recognition accuracy results. For abbreviations see Table 1.

Accuracy[%]	Hap.	Sad.	Sur.	Ang.	Fear	Dis.
Hap.	**100**	0	0	0	0	0
Sad.	0	**100**	0	0	0	0
Sur.	0	1.2	**97.6**	0	1.2	0
Ang.	0	0	0	**93.3**	0	6.7
Fear	8.0	4.0	0	0	**88.0**	0
Dis.	0	1.7	0	1.7	0	**96.6**

Table 3. Result comparison. For abbreviations see Table 1.

Paper	Database	Accuracy [%]						
		Hap.	Sad.	Sur.	Ang.	Fear	Dis.	Total
Proposed method	CK+	**100**	**100**	97.6	**93.3**	88.0	**96.6**	**96.8**
Lucey *et al.* (2010)	CK+	**100**	68.0	96.0	75.0	65.2	94.7	88.6
Sebe *et al.* (2007)	Their own	95.7	92.0	88.7	91.2	**94.7**	85.6	93.6
Michel (2003)	CMU	95.3	85,6	**98.8**	78.4	76.2	83.9	86.3
Visutsak (2005)	JAFFE	91.5	61.0	97.5	67.7	66.7	62.3	74.5

Bayesian networks, k-nearest neighbor (kNN), and SVM classifiers, producing an accuracy average of 93.6%. Their results outperform ours when recognizing fear (see Table 3), probably due to the usage of spontaneous expressions in their experiments. Fear is particularly difficult to express on demand, so their approach captures facial movements that are absent in acted databases such as CK+. However, our total accuracy, using a simple method, outperforms their results. Lucey *et al*. (2010) extracted 68 landmarks with AAMs, and produced excellent results for detecting happiness and surprise. Their nose wrinkle detector, which was also utilized in our system, meant that the recognition of disgust had high accuracy. However, sadness, anger, and fear have much lower recognition rates than in our work. They reached final recognition rate of 88.6%. Our results show that our system produces excellent recognition rates for all the six basic emotions. In particular, the recognition of sadness, anger, and fear are drastically improved with our approach. Our overall accuracy outperforms results from other work, with total recognition rate of 96.8%.

5. Conclusion

Existing facial expression recognition systems are good at detecting happiness and surprise, but perform poorly for the other basic emotions. Our method presents an approach which reduces this performance gap, while being both simple and effective. Looking at previous studies, part of the recognition errors for sadness, anger, fear, and disgust are caused by head movements. Since our system examines the relative displacements of facial landmarks with respect to a frame's reference point, it is more resistant to head movement errors, and so yields better results.

Emotions are expressed differently over time, so our approach utilizes additional frames at the one third and two thirds time points of each sequence. Other error-reducing elements include considering movement along the x- and y-axes, expression duration, and nose wrinkle detection. The proposed error-reducing elements are simple and intuitive, without being too time-consuming. For example, with a few additional Euclidian distances calculated, utilizing two middle frames in a sequence, we gain a new time-dimension that keeps a better track of the facial changes that lead from the neutral to the peak emotion expression. The human eyes observe and analyze facial expressions in a similar fashion, which makes our approach more intuitive and natural, and we proved in our experiments that these additional facial changes are very informative for the classifier, reducing the overall system error. With accuracies of around 90%, we have successfully bridged the recognition gap between the six basic facial expressions. Our final overall recognition rate reaches 96.8%.

References

Batty, M. and Taylor, M.J. 2003. Early processing of the six basic facial emotional expressions. Cognitive Brain Research. 17, 613–620.

Bettadapura, V. 2009. Face expression recognition and analysis: the state of the art. Columbia University, Computer and Information Science, U.S.A.

Chang, C.-C. and Lin, C.-J. 2011. LIBSVM: a library for support vector machines. Association for Computing Machinery (ACM) Transactions on Intelligent Systems and Technology. 2(3), Article No. 27.

Cohen, I., Sebe, N., Garg, A., Chen, L. S., and Huang, T. S. 2003. Facial expression recognition from video sequences: temporal and static modeling. Computer Vision and Image Understanding, 91, 160-187.

Dornaika, F. and Davoine, F. 2008. Simultaneous facial action tracking and expression recognition in the presence of head motion. International Journal of Computer Vision. 76(3), 257–281.

Ekman, P. and Friesen, W.V.1971. Constants across cultures in the face and emotions. Journal of Personality and Social Psychology. 17(2), 124-129.

Ekman, P. and Friesen, W.V.1977. Manual for the Facial Action Coding System, Consulting Psychologists Press, Palo Alto, U.S.A.

Ekman, P., Friesen, W.V., and Hager, J.C. 2002. Facial Action Coding System: A Human Face, Salt Lake City, Utah, U.S.A.

Kanade, T., Cohn, J.F., and Tian, Y. 2000. Comprehensive database for facial expression analysis. Proceedings of the Fourth Institute of Electrical and Electronics Engineers (IEEE) International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, March 2000, 46-53.

Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. 2010. The extended Cohn-Kanade dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. Proceedings of the Third International Workshop on Computer Vision and Pattern Recognition (CVPR) for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, U.S.A., June 18, 2010, 94-101.

Michel, P. 2003. Support vector machines in automated emotion classification. Computer Science Tripos Part II Project Dissertation, University of Cambridge, U.K.

Pantic, M., Valstar, M.F., Rademaker, R., and Maat, L. 2005. Web-based database for facial expression analysis. Proceeding of the 13th Annual Association for Computing Machinery (ACM) International Conference on Multimedia (Multimedia'05), Singapore, November 6-11, 2005, 317-321.

Sebe, N., Lew, M. S., Sun, Y., Cohen, I., Gevers, T., and Huang, T. S. 2007. Authentic facial expression analysis. Image and Vision Computing. 25, 1856-1863.

Tian, Y., Kanade, T., and Cohn, J. 2001. Recognizing action units for facial expression analysis. Institute of Electrical and Electronics Engineers (IEEE) Transaction on Pattern Analysis and Machine Intelligence.23(2), 97–115.

Visutsak, P. 2005. Emotion recognition through lower facial expressions using support vector machines. Proceeding of the Fifth National Symposium on Graduate Research, Kasetsart University, Bangkok, Thailand, October 10-11, 2005