International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Facial Expression Recognition Using Facial Movement Features

Ruchi Chaurasia

Abstract: ***Facial expressions give important information about emotions of a person. Understanding facial expressions accurately is one of the challenging tasks for interpersonal relationships. Automatic emotion detection using facial expressions recognition is now a main area of interest within various fields such as computer science, medicine, and psychology and To improve the human-computer interaction (HCI) to be as good as human-human interaction, building an efficient approach for human emotion recognition is required. These emotions could be fused from several modalities such as facial expression, hand gesture, acoustic data, and biophysiological data.This paper proposes an approach to solve this limitation using 'salient' distance features, which are obtained by extracting patch-based 3D Gabor features, selecting the 'salient' patches, and performing patch matching operations. The experimental results demonstrate high correct recognition rate (CRR), significant performance improvements due to the consideration of facial element and muscle movements, promising results under face registration errors, and fast processing time. The comparison with the state-of-the-art performance confirms that the proposed approach achieves the highest CRR on the JAFFE database and is among the top performers on the Cohn-Kanade (CK) database.***

Keywords: Facial expression analysis, feature evaluation and selection, computer vision, Gabor filter, Ada boost

1. Introduction

Face plays an important role in human communication. Facial expressions and gestures incorporate nonverbal information which contributes to human communication. By recognizing the facial expressions from facial images, a number of applications in the field of human computer interaction can be facilitated. Last two decades, the developments, as well as the prospects in the field of multimedia signal processing have attracted the attention of many computer vision researchers to concentrate in the problems of the facial expression recognition. The pioneering studies of Ekman in late 70s have given evidence to the classification of the basic facial expressions. According to these studies, the basic facial expressions are those representing happiness, sadness, anger, fear, surprise, disgust and neutral. Facial Action Coding System (FACS) was developed by Ekman and Friesen to code facial expressions in which the movements on the face are described by action units. This work inspired many researchers to analyze facial expressions in 2D by means of image andvideo processing, where by tracking of facial features and measuring the amount of facial movements, they attempt to classify different facial expressions. Recent work on facial expression analysis and recognition has used these seven basic expressions as their basis for the introduced systems. Almost all of the methods developed use 2D distribution of facial features as inputs into a classification system, and the outcome is one of the facial expression classes. They differ mainly in the facial features selected and the classifiers used to distinguish among the different facial expressions. Information extracted from 3D face models are rarely used inthe analysis of the facial expression recognition. This chapter considers the techniques using the information extracted from 3D space for the analysis of facial images for the recognition of facial expressions. The first part of the chapter introduces the methods of extracting information from 3Dmodels for facial expression recognition. The 3D distributions of the facial feature points and the estimation of characteristic distances in order to represent the facial expressions are explained by using a rich collection of illustrations including graphs, charts and face images. The second part of the chapter introduces 3D distance-vector based facial expression recognition. The architecture of the system is explained by the block diagrams and flowcharts. Finally 3D distance-vector based facial expression recognition is compared with the conventional methods available in the literature. Facial movement features, which include feature position and shape changes, are generally caused by the movements of facial elements and muscles during the course of emotional expression. The facial elements, especially key elements, will constantly change their positions when subjects are expressing emotions. As a consequence, the same feature in different images usually has different positions, as shown in Fig. 1a. In some cases, the shape of the feature may also be distorted due to the subtle facial muscle movements. For example, the mouth in the first two images in Fig. 1b presents different shapes from that in the third image. Therefore, for any feature representing a certain emotion, the geometry-based position and appearance based shape normally change from one image to another image in image databases, as well as in videos. This kind of movement features represents a rich pool of both static and dynamic characteristics of expressions, which play a critical role for FER. The vast majority of the past work on FER does not take.

The dynamics of facial movement features into account [1]. Some efforts have been made in capturing and utilizing facial movement features, and almost all of them are video based. These efforts try to adopt either geometric features of the tracked facial points (e.g., shape vectors [2], facial animation parameters [3], distance and angular [4], and trajectories [5]), or appearance difference between holistic facial regions in consequent frames (e.g., optical flow [6], and differential-AAM [7]), or texture and motion changes in local facial regions (e.g., surface deformation [8], motion units [9], spatiotemporal descriptors [10], animation units [11], and pixel differences [12]). Although they achieved promising results, these approaches often require accurate location and tracking of facial points, which remains problematic [11]. In addition, it is still an open question as to how to learn the grammars in defining dynamic features

Licensed Under Creative Commons Attribution CC BY

and handle ambiguities in the input data [12]. On the other hand, image-based FER techniques provide an alternative way to recognize emotions based on appearance-based features in a single image, and are important for the situation where only several images are available for training and testing. However, to the best of our knowledge, no research has been reported on image-based FER that considers facial movement features. In this paper, we aim for improving the performance of FER by automatically capturing facial movement features in static images based on distance features. The distances are obtained by extracting "salient" patch-based Gabor features and then performing patch matching operations. Patch based Gabor features have shown excellent performance in overcoming position, scale, and orientation changes [11],
[6], [7], as well as extracting spatial, frequency, and orientation information [8]. They also show a great advantage over the commonly used fiducially point-based

Gabor [9] graph-based Gabor [4], and discrete Fourier transform [5] features in capturing regional information. Although other appearance-based features, such as local binary patterns (LBP) [2] Haar [12], and histograms of oriented gradients (HOG) [9], have shown good performance in FER, they lack the capacity of capturing facial movement features with high accuracy. This is due to the fact that these appearance-based features are based on statistic values (e.g., histogram similarity) extracted from sub regions; therefore, they produce similar results even when facial features move a bit from the original position. On the other hand, Gabor features have the capacity to accurately capture movement information and have been proven to be robust, even in the case of face misalignment [3].

Figure 1: Facial movement features. (a) Feature position (left mouth corner) changes. (b) Feature shape (mouth) changes. Facial regions are manually cropped from two subjects, "KA" and "KL," on the JAFFE database.

2. Proposed Framework

Fig. 2 illustrates the proposed framework, which is composed of preprocessing, training, and test stages. At the preprocessing stage, by taking the nose as the center and keeping main facial components inclusive, facial regions are manually cropped from database images and scaled to a resolution of 41*41 pixels. No more processing is conducted to imitate the results of real face detectors. Then, multi resolution Gabor images are attained by convolving eight scale, four-orientation Gabor filters with the scaled facial regions. During the training stage, a whole set of patches is extracted by moving a series of patches with different sizes across the training Gabor images. Then, a patch matching operation is proposed to convert the extracted patches to distance features. To capture facial movement features, the matching area and matching scale are defined to increase the matching space, whereas the minimum rule is used to find the best matching feature in this space. Based on the converted distance features, a set of "salient" patches is selected by Adaboost. At the test stage, the same patch matching operation is performed on a new image using the "salient" patches. The resulting distance features are fed into a multiclass support vector machine (SVM) to recognize six basic emotions, including anger (AN), disgust (DI), fear (FE), happiness (HA), sadness (SA), and surprise (SU).

The rest of this section gives an introduction of Gabor filters and SVM. The details of building distance features and feature selection are explained in Sections 4 and 5, respectively. In this paper, 2D Gabor filter [31] is adopted and it can be mathematically expressed as:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right)$$

$$x' = x \cos\theta + y \sin\theta$$

$$y' = -x \sin\theta + y \cos\theta$$

Instead of the widely used five scales, eight scales (5:2:19 pixels) are adopted here to test the results using a larger number of scales. The values of the rest of the parameters are set based on [15] due to the high reported performance. As a result, four orientations (————————————45, 90, 45, 0 degrees) are used. SVM [32] is one of the most widely used machine learning algorithms for classification problems. This paperdirectly uses the LIBSVM [33] implementation of SVMs with four different kernels, including linear, polynomial, radial basis function (RBF), and sigmoid. The six emotion class problem is solved by the one-against-the-rest strategy.

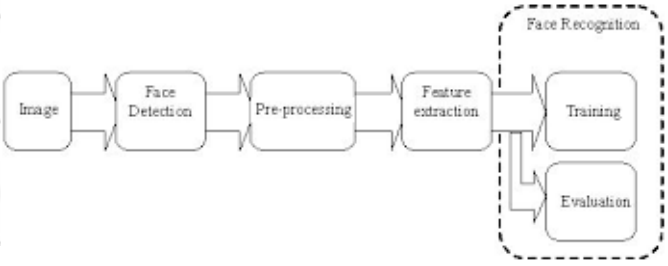

Figure 2: Some techniques for facial expression recognition

3. Module Description

3.1 Skin Color Segmentation

For skin color segmentation, first we contrast the image. Then we perform skin color segmentation. Then, we have to find the largest connected region. Then we have to check the probability to become a face of the largest connected region. If the largest connected region has the probability to become a face, then it will open a new form with the largest connected region. If the largest connected regions height & width is larger or equal than 50 and the ratio of height/width is between 1 to 2, then it may be face.

Licensed Under Creative Commons Attribution CC BY

3.2 Face Detection

For face detection, first we convert binary image from RGB image. For converting binary image, we calculate the average value of RGB for each pixel and if the average value is below than 110, we replace it by black pixel and otherwise we replace it by white pixel. By this method, we get a binary image from RGB image.Then, we try to find the forehead from the binary image. We start scan from the middle of the image, then want to find a continuous white pixels after a continuous black pixel. Then we want to find the maximum width of the white pixel by searching vertical both left and right site. Then, if the new width is smaller half of the previous maximum width, then we break the scan because if we reach the eyebrow then this situation will arise. Then we cut the face from the starting position of the forehead and its high will be 1.5 multiply of its width. Then we will have an image which will contain only eyes, nose and lip. Then we will cut the RGB image according to the binary image.

3.3 Eyes Detection

For eyes detection, we convert the RGB face to the binary face. Now, we consider the face width by W. We scan from the W/4 to (W-W/4) to find the middle position of the two eyes. The highest white continuous pixel along the height between the ranges is the middle position of the two eyes.

Then we find the starting high or upper position of the two eyebrows by searching vertical. For left eye, we search w/8 to mid and for right eye we search mid to w – w/8. Here w is the width of the image and mid is the middle position of the two eyes. There may be some white pixels between the eyebrow and the eye. To make the eyebrow and eye connected, we place some continuous black pixels vertically from eyebrow to the eye. For left eye, the vertical black pixel-lines are placed in between mid/2 to mid/4 and for right eye the lines are in between mid-+(w-mid)/ 4 to mid-+3*(w-mid)/ 4 and height of the black pixel-lines are from the eyebrow starting height to (h- eyebrow starting position)/4. Here w is the width of the image and mid is the middle position of the two eyes and h is the height of the image. Then we find the lower position of the two eyes by searching black pixel vertically. For left eye, we search from the mid/4 to mid - mid/4 width. And for right eye, we search mid + (w-mid)/ 4 to mid-+3*(w- mid)/ 4 width from image lower end to starting position of the eyebrow. Then we find the right side of the left eye by searching black pixel horizontally from the mid position to the starting position of black pixels in between the upper position and lower position of the left eye. And left side for right eye we search mid to the starting position of black pixels in between the upper position and lower position of right eye. The left side of the left eye is the starting width of the image and the right side of the right eye is the ending width of the image. Then we cut the upper position, lower position, left side and the right side of the two eyes from the RGB image.

3.4 Lip Detection

For lip detection, we determine the lip box. And we consider that lip must be inside the lip box. So, first we determine the distance between the forehead and eyes. Then we add the distance with the lower height of the eye to determine the upper height of the box which will contain the lip. Now, the starting point of the box will be the ¼ position of the left eye box and ending point will be the ¾ position of the right eye box. And the ending height of the box will be the lower end of the face image. So, this box will contain only lip and may some part of the nose. Then we will cut the RGB image according the box. So, for detection eyes and lip, we only need to convert binary image from RGB image and some searching among the binary image.

4. Building Distance Features

Fig. 3 depicts the two main processes to build distance features: feature extraction and patch matching operation. Feature extraction aims to collect a set of discriminating 3D. Patches for all emotions, whereas the patch matching operation converts these patches to distance features which can capture facial movement features.

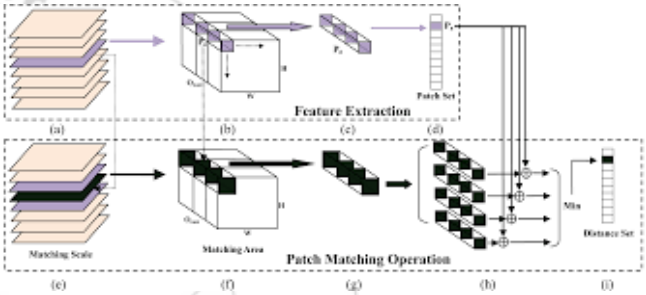

Figure 3: Building distance features. (a) One scale is selected from eight-scale Gabor images. (b) Patches are extracted across all rows and columns in the selected scale image. (c) One extracted patch Pa. (d) Extracted patch set. (e) Defined matching scale. (f) Defined matching area. (g) One matching area. (h) Distance calculation. (i) Distance feature set.

4.1Patch-Based Feature Extraction

The amount of digital documents increases daily, and with it the need to organize this torrent of data in order to retrieve something again. Especially for digital images no ideal solution has been found yet. The manual annotation of images is very labor intensive, so the vast majority of images will remain annotated. Techniques for content based image retrieval (CBIR) are able to find similar images based on pixel content only; however, usually the definition of similarity is on a color and texture level, not on a semantic level. Most users do not want to find things with just the same texture and color, but want to find semantic entities, images with particular objects like cows, sheep or cars. This is why the main focus of research is now drawn to the recognition of visual object classes rather than the already widely researched area of traditional CBIR.

4.2Patch Matching Operation

As shown in the lower part in Fig. 3, the patch matching operation comprises of four steps for each patch and each training image: First, the matching area and matching scale are defined to provide a bigger matching space (Figs. 3e and

Licensed Under Creative Commons Attribution CC BY

3f). Second, the distances are obtained by matching this patch with all patches within its matching space in a training image (Fig. 3h). This step takes two patches as inputs and yields one distance value based on a distance metric. Third, the minimum distance is chosen as the distance feature of this patch in the training image (Fig. 3h). Finally, the distance features of all patches are combined into a final set with 148,032 elements (Fig. 3i).

4.2.1Matching Area and Matching Scale Definition

The matching area and matching scale are used to accurately capture the position and scale changes caused by facial feature movements. The idea of them stems from the observation that the position and scale of one feature do not move or change a lot in different facial images once these images are roughly located by a face detector. Thus, the invariance to position and scale changes can be accomplished by defining a bigger area and a larger scale for each patch when performing patch matching. An automatic method of image-to-image registration, which may be applied to register overlapped images of a scene from different views and dates, is presented. The proposed approach is a feature-based matching with constraints of orientation consistency and one-to-one match. Area features of homogeneous regions in gray level are extracted from images as matching entities. The boundaries of area features are matched in the frequency domain, i.e., matching their Fourier descriptors. The spatial meaning of the matching is transforming a feature to fit the other optimally. After the matching process, this scheme provides not only a quantitative evaluation of the remaining lack of fitness as an objective index of shape similarity, but also the solved transformation parameters to represent the relative orientation between features. The evaluation of shape similarity is used as the key information in recognizing conjugate features for overlapped images. Furthermore, the consistency of relative orientation between matched pairs is considered as the principal constraint to dissolve improperly registered features.

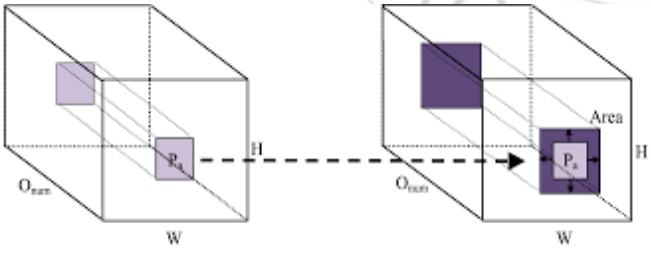

Figure 4:.Matching area. (a) One patch Pa. (b) The corresponding matching area "Area" in a new image.

5. Patch-Based Feature Selection and Analysis

In this section, we use Adaboost for discriminative (called "salient" here) patch selection on the Japanese female facial expression (JAFFE) and CK databases. To give a deeper understanding of the selected "salient" patches and provide useful information on the design of Gabor filters and feature extraction algorithms, we also present a description on their position, number, size, scale, and overlap distributions.

5.1Databases

The JAFFE database [5] contains 213 gray images of seven facial expressions (six basic + neutral) poses of 10 Japanese females. Each image has a resolution of 256*256 pixels. Each object has three or four frontal face images for each expression and their faces are approximately located in the middle of the images. All images have been rated on six emotion adjectives by 60 subjects

The Cohn-Kanade AU coded facial expression (CK) database [36] is one of the most comprehensive benchmarksfor facial expression tests. The released portion of this database includes 2,105 digitized image sequences from 182 subjects ranged in age from 18 to 30 years. Sixtyfive percent are female; 15 percent are African-American and 3 percent Asian or Latino. Six basic expressions were based on descriptions of prototypic emotions. Image sequences from neutral to target display were digitized into 640,480 or 490 pixel arrays with eight-bit precision for gray scale values.

In this paper, all the images of six basic expressions from the JAFFE database are used. For the CK database, 1,184 images that represent one of the six expressions are selected, four images for each expression of 92 subjects. The images are chosen from the last image (peak) of each sequence, then one every two images. The images of 10 subjects in the JAFFE database are classified into 10 sets, each of which includes images of one subject. Similarly, all images in the CK database are classified into 10 similar sets and all images of one subject are included in the same set.

5.2 "Salient" Patch Selections

The feature extraction step produces a feature set containing 148,032 patches. To reduce the feature dimension and the redundant information, it is necessary to select a subset of "salient" patches. In this paper, the widely used and efficiency proven boosting algorithm—Adaboost [7]—is used for "salient" patch selection. Since Adaboost was designed to solve two-class problems, in this research the one-against-the-rest strategy is used to solve the six-emotion-class problem. The training process stops when the empirical error is below 0.0001 with an initial error of 1. This setting is inspired by the stopping condition in [20] that there is no training error and the generalization error becomes flat. For the training set, the JAFFE database includes all database images, whereas the CK database is only composed of the peak frames.

Licensed Under Creative Commons Attribution CC BY

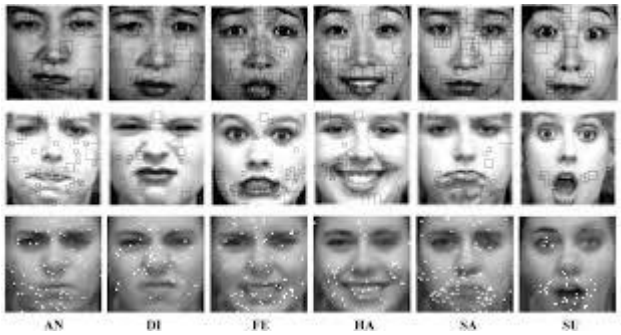

Figure 5: Position distribution of the selected "salient" patches for six emotions. The first and second rows show positions of the selected patches in the proposed approach on the JAFFE and CK databases, respectively, whereas the third row reveals positions of the selected point-based Gabor features in [8] on the CK database.

5.3Position Distribution of "Salient" Patches

The position distribution of the "salient" patches demonstrates the most important facial areas for each emotion. In Fig. 5, the patches are distributed over different Gabor scales, and they are drawn in one scale image for a simple and clear demonstration. Based on this figure, we can see that the positions are distributed differently over six emotions. However, most of these patches for all emotions tend to concentrate on the areas around the mouth and eyes. For sadness and surprise, the "salient" patches on JAFFE focus on the eye areas, while those on CK focus on the mouth area. For the remaining four emotions, they have similar distributions between two databases. As shown in the second and third rows in Fig. 5, the positions of the "salient" patches in our work and those of the point-based "salient" features in [8] for the same emotion tend to focus on the same areas. This suggests that there exist the same "salient" areas for each emotion regardless of use of pointbased or patch-based Gabor features. However, the overall number of the "salient" patches is much less than that of the point-based "salient" features (177 versus 538).

5.4 Number and Size Distributions of "Salient" Patches

The number and size distributions can provide useful hints on the number of patches for different emotions and howto choose suitable patch sizes during feature extraction. As seen in Fig. 6, two databases have a similar overall number of the "salient" patches. Among six emotions, fear and sadness need the largest numbers of patches to achieve the preset recognition accuracy, whereas surprise requires the least number. Within four patch sizes, the size 4* 4 takes a significant proportion of the overall number of the "salient" patches.

On the other hand, there are also some differences between two databases. The number for anger on JAFFE is much less than that on CK, while the number for disgust on JAFFE is much bigger than that on CK. Moreover, four patch sizes are evenly distributed among six emotions on JAFFE, but the patch size 2* 2 takes a significant proportion of the overall number of the "salient" patches on CK. This reflects that emotions in JAFFE images need big sizes of patches to represent useful information, whereas those in CK images only require small sizes of patches. The reason may be that the emotions in CK are more distinct than those in JAFFE.

5.5Scale Distribution of "Salient" Patches

The scale distribution is a very important factor for determining the scale number of Gabor filters. As observed in Fig. 7, the "salient" patches of two databases are unevenly distributed across eight scales. JAFFE emphasizes the eighth scale and CK focuses on the fourth scale. For both databases, the higher scales (fourth to eighth) contain more patches than the lower scales (first to third). Therefore, the emotional information is distributed across all scales with an emphasis on the higher scales, which confirms Littlemore's argument that a wider range of spatial frequencies, particularly high frequencies, could potentially improve performance [20].

5.6 Overlap Distribution of "Salient" Patches

Table 1 demonstrates the characteristics of number, size, scale, emotion pair of the overlapping patches, which are selected as "salient" patches more than one time. As can be seen, the JAFFE database has a larger number of the overlapping patches than the CK database (eight versus three). As for the patch size, 4 4 dominates the overlapping patches on JAFFE, while 2 2 takes most of these patches on CK. With respect to the patch scale, the patches of JAFFE tend to distribute on the sixth, seventh, and eighth scales, whereas those of CK are all included by the fourth scale. The patch size and scale distributions again reveal that JAFFE needs larger patches than CK. For the emotion pair, the majority of the overlapping patches are shared by the same emotion. As shown in Fig. 9, the overlapping patches are mainly distributed over disgust and anger.

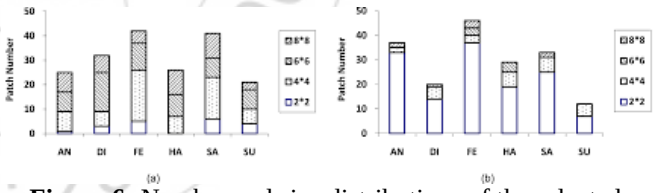

Figure 6:.Number and size distributions of the selected "salient" patches for six emotions on (a) the JAFFE and (b) the CK databases. Note that Onum is four for all patch sizes of Pj*Pj* Onum and it is not shown.

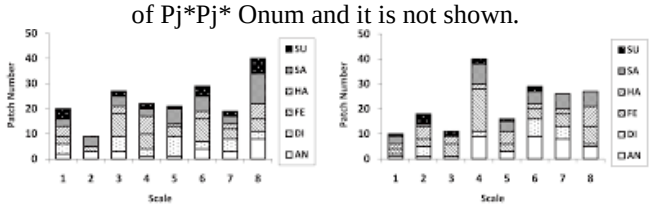

Figure 7: Scale distribution of the selected "salient" patches for six emotions on (a) the JAFFE and (b) the CK databases.

6. Recognition Performance

6.1JAFFE Database

The performance results are obtained by averaging the correct recognition rate (CRR) of all sets in 10 leave-one set

Licensed Under Creative Commons Attribution CC BY

out cross validations. Table 2 shows the results obtained using four SVMs and four distances. From this table, we can see that the proposed approach performs the best with a CRR of 92.93 percent using DL2 and linear SVM. Regarding the performance of distances,DL2 achieves higher CRRs than the other three distances for all SVMs.WhenL1 is used, sparse distances outperform dense distances for linear, RBF, and sigmoid SVMs. On the contrary, when L2 is used, dense distances outperform sparse distances for all SVMs (note that the CRR of DL2 and sigmoid SVM is not shown). For both sparse and dense distances, L2 performs better than L1 for all SVMs. Among four SVMs, linear and RBF outperform polynomial and sigmoid for all distances. More exactly, the best performance is obtained by linear, which is followed by RBF, whereas sigmoid ranks the lowest. Table 3 demonstrates the confusion matrix of six emotions using DL2 and linear SVM. Observed from this table, disgust and surprise belong to the most difficult facial expressions to be correctly recognized with the same CRR of 90.00 percent, whereas anger is the easiest one with a CRR of 96.67 percent. Regarding the misrecognition rate, anger contributes the most; as a result, it has a major negative impact on the overall performance. The emotion that follows in misrecognition rate is fear.

Table 1: Overlapping Patches on the JAFFE and CK Databases

	JAFFE	CK
Patch Size	5(4*4); 1(6*6); 2(8*8)	2(2*2); 1(8*8)
Patch Scale	1(3^{rd}); 2(6^{th}); 2(7^{th}); 3(8^{th})	3(4^{th})
Emotion Pair	3(AN-AN); 2(DI-DI); 1(FE-FE); 1(FE-SA); 1(SA-SA)	2(AN-AN); 1(AN-DI)
Total Number	8	3

Table 2: CRRs of Six Emotions on the JAFFE Database

	DL_1	DL_2	SL_1	SL_2
Linear	81.52%	**92.93%**	87.50%	88.59%
Polynomial	54.89%	64.13%	45.65%	60.33%
RBF	76.63%	89.67%	82.07%	87.50%
Sigmoid	25.00%	-	26.09%	29.35%

The CRR of DL_2 and sigmoid SVM is not shown.

Table 3: Confusion Matrix of Six Emotions on the JAFFE Database

	AN	DI	FE	HA	SA	SU	Overall
AN	29	1	0	0	0	0	**96.67%**
DI	2	27	0	0	0	1	**90.00%**
FE	2	0	30	0	0	0	93.75%
HA	1	0	1	29	2	0	93.55%
SA	0	0	1	1	29	0	93.55%
SU	0	1	1	1	0	27	**90.00%**

6.2CK Database

The CRRs using four SVMs and four distance metrics are shown in Table 4, in which the proposed approach obtains the highest CRR of 94.48 percent using DL2 and RBF SVM. Regarding the performance of distances, DL2 keeps the highest CRRs for all SVMs (note that the CRR of DL2 and sigmoid SVM is not shown). Moreover, dense distances have a higher overall performance than sparse distances.

This reflects that emotional information in the CK images is distributed over all orientations rather than the dominant orientation of Gabor features. As for SVMs, RBF performs the best for dense distances, while linear performs the best for sparse distances. This confirms with the results in that RBF and linear perform better than polynomial on the CK database.

Table 5 shows the confusion matrix of six emotions using DL2 and RBF SVM. As can be seen, surprise performs the best, with a CRR of 100 percent, the following one is happiness, with a CRR of 98.07 percent. On the other hand, anger is the most difficult facial expression to be correctly recognized, with a CRR of only 87.10 percent. The reason probably is that surprise images on CK are often characterized as an exaggerated "open mouth," while those on JAFFE are normally with a "close or slightly open mouth. "It can be seen from Fig. 6 that the selected patches for CK focus on the mouth region, but those for JAFFE are mainly distributed around the eyes regions. Similarly, anger images are better expressed by the patches selected in the eye regions on JAFFE than those selected in the whole face on CK. Among six emotions, anger and sadness contribute most to the misrecognition rate.

Table 4: CRRs of Six Emotions on the CK Database

	DL_1	DL_2	SL_1	SL_2
Linear	90.20%	93.36%	83.67%	86.71%
Polynomial	87.73%	91.22%	65.43%	80.97%
RBF	92.34%	**94.48%**	80.07%	86.26%
Sigmoid	26.46%	-	37.39%	66.67%

Table 5: Confusion Matrix of Six Emotions on the CK Database

	AN	DI	FE	HA	SA	SU	Overall
AN	81	1	2	0	9	0	**87.10%**
DI	4	92	2	2	1	1	90.20%
FE	0	4	138	7	0	1	92.00%
HA	0	2	2	203	0	0	98.07%
SA	6	0	2	0	118	3	91.47%
SU	0	0	0	0	0	207	**100%**

Table 6: Error Thresholds Used to Control the Number of Patches

Index	Error thresholds
1^{st} -10^{th}	(10, 9, 8, 7, 6, 5, 4, 3, 2, 1) * 0.1
11^{th} -19^{th}	(9, 8, 7, 6, 5, 4, 3, 2, 1) * 0.01
20^{th} -28^{th}	(9, 8, 7, 6, 5, 4, 3, 2, 1) * 0.001
29^{th} -38^{th}	(9, 8, 7, 6, 5, 4, 3, 2, 1, 0) * 0.0001

Note that '1' rejects all features, whereas '0' accepts all features.

Licensed Under Creative Commons Attribution CC BY

Figure 13: Sample images with errors simulated by uniform random noises ranged (a) [–3 percent, 3 percent] and (b) [–6 percent, 6 percent] of face width.

7. Conclusion and Future Work

This paper explores the issue of facial expressionrecognition using facial movement features. The effectiveness of the proposed approach is testified by the recognition performance, computational time, and comparison with the state-of-the-art performance.

The experimental results also demonstrate significant performance improvements due to the consideration of facial movement features and promising performance under face registration errors. The results indicate that patch-based Gabor features show a better performance over point-based Gabor features in terms of extracting regional features, keeping the position information, achieving a better recognition performance, and requiring a less number. Different emotions have different "salient" areas; however, the majority of these areas are distributed around mouth and eyes. In addition, these "salient" areas for each emotion seem to not be influenced by the choice of using point-based or using patch-basedfeatures. The "salient" patches are distributed across all scales with an emphasis on the higher scales. For both the JAFFE and CK databases, DL2 performs the best among four distances. As for emotion, anger contributes most to the misrecognition. The JAFFE database requires larger sizes of patches than the CK database to keep useful information. The proposed approach can be potentially applied into many applications, such as patient state detection, driver fatigue monitoring, and intelligent tutoring system. In our future work, we will extend our approach to a video-based FER system by combining patch-based Gabor features with motion information in multiframes. Recent progress on action recognition [7] and face recognition [8] has laid a foundation for using both appearance and motion feature.

References

[1] http://www.asprs.org/a/publications/pers/97journal/august/1997_aug_975-983.pdf
[2] http://lmb.informatik.unifreiburg.de/papers/download/te_dagm06.pdf
[3] http://freeproject.co.in/source/Facial-Expression-RecognitionUsing-Facial-Movement-features2011.aspx?pf=.Net&t=ieee
[4] http://www.denniscodd.com/dotnetieee/Facial/Expression/RecognitionUsing/Facial.pdf
[5] http://www.academia.edu/2696024/Facial_expression_recognition_using_3D_facial_feature_distances
[6] http://www.slideshare.net/Projucti/facial-expression-recognition-using-facial-movement-features-28159320
[7] http://www.ijarcsse.com/docs/papers/Volume_4/4_April2014/V4I4-0235.pdf
[8] http://ibug.doc.ic.ac.uk/media/uploads/documents/EncycBiometrics-Pantic-FacExpRec-PROOF.pdf
[9] http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/CHIBELUSHI1/CCC_FB_FacExprRecCVonline.pdf
[10] http://research.microsoft.com/enus/um/people/zhang/papers/ijprai.pdf
[11] http://www.eecs.qmul.ac.uk/~sgg/papers/ShanGongMcOwan_IVC09.pdf
[12] http://arxiv.org/ftp/arxiv/papers/1203/1203.6722.pdf

Licensed Under Creative Commons Attribution CC BY