Face and Facial Expression Recognition
Fusion based Non Negative Matrix Factorization

Humayra Binte Ali[1] and David M. W. Powers[1,2]
[1]*School of Computer Science, Engineering, and Mathematics, Flinders University, Adelaide, South Australia*
[2]*Beijing Municipal Lab for Multimedia and Intelligent Software, Beijing University of Technology, Beijing, China*

Keywords: NMF-Non Negative Matrix Factorization, OEPA- Optimal Expression- specific Parts Accumulation, FR-Face Recognition, FER- Facial Expression Recognition.

Abstract: Face and facial expression recognition is a broad research domain in machine learning domain. Non-negative matrix factorization (NMF) is a very recent technique for data decomposition and image analysis. Here we propose face identification system as well as a facial expression recognition, which is a system based on NMF. We get a significant result for face recognition. We test on CK+ and JAFFE dataset and we find the face identification accuracy is nearly 99% and 96.5% respectively. But the facial expression recognition (FER) rate is not as good as it required for the real life implementation. To increase the detection rate for facial expression recognition, our propose fusion based NMF, named as OEPA-NMF, where OEPA means Optimal Expression-specific Parts Accumulation. Our experimental result shows OEPA-NMF outperforms the prevalence NMF for facial expression recognition. As face identification using NMF has a good accuracy rate, so we are not interested to apply OEPA-NMF for face identification.

1 INTRODUCTION

The face has been termed as the most prominent perceptual stimulus in real world for social and person-to-person communication (Frith and Baron-Cohen, 1987). Each face possesses the uniqueness and robustness, which makes is completely distinguishable from other faces. For this reason, from safety enhancement to person authority checking, face recognition is a valuable application area. Face recognition covers both the domain of Face Identification and Face Verification. Face Identification means to find the identity of a given person out of a pool of *N* persons (1 to *N* matching) and this Face Identification is widely used in video surveillance, information retrieval, video games and some other human computer interaction areas. On the other hand, Face Verification establishes the process of confirming or denying the identity claimed by a person (1 to 1 matching). To verify access control into computer or mobile device or building gate, and digital multimedia data access control, Face Verification technique is needed. Facial expression plays a great role in both human to human and human to machine communication. (Charlesworth and Kreutzer, 1973) mentioned infants as young as three months old are able to discern facial emotion. To express emotion, attitude and feelings human communicate through speech, facial expression and also body language. Facial expression has a wide variety of applications, like, pain level measurement for medical purpose, terrorist identification, lie detection etc. Stemming from Darwin's work, the earliest discrete theories of emotion hypothesized the existence of a small number of basic emotions, such as happiness, sadness, fear, anger, surprise and disgust (Ekman, 1994). In this research work our main concern is to increase the detection rate of these basic emotions only. We apply here NMF for data factorization and euclidian distance as classifier, to recognize basic facial expressions. We also compare the system with principal component analysis based FER system.

2 RESEARCH BACKGROUND

Subspace learning algorithms has been successfully used in image analysis, data mining and video compressing areas. Basically these methods are widely used to handle large amount of data as it performs dimension reduction as well as finds the direction along with certain properties. The most prevalence

Binte Ali H. and M. W. Powers D..
Face and Facial Expression Recognition - Fusion based Non Negative Matrix Factorization.
DOI: 10.5220/0005216004260434
In *Proceedings of the International Conference on Agents and Artificial Intelligence* (ICAART-2015), pages 426-434
ISBN: 978-989-758-074-1

subspace learning techniques are Principal Component Analysis (PCA), Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), Non Negative Matrix Factorization (NMF) etc. Compared with others, NMF is a recent technique in machine learning areas. This research work is mainly based on NMF and comparison with PCA following same procedure. PCA based face recognition system has been developed in (J. Buhmann and Malsburg, 1990), where singular value decomposition (SVD), an extension of PCA, has been used. PCA by applying Bayesian method is used for face recognition (Moghaddam and Pentland, 1997). These (M. Lades and Konen, 1993) (C. Hesher and Erlebacher, 2003) are some other successful applications of PCA. Many of these mentioned works are performed based on the early research of PCA (Turk and Pentland, 1991) with some extensions.

NMF is becoming popular in face recognition research areas. for sparse representation of data (Cichocki, 2009). Now we are going to discuss some research works based on NMF. An extension of basic NMF has been implemented on ORL face database (Ensari and Zurada, 2012). (Li and Oussalah, 2010) claims 75% recognition rate for face recognition. (L. Zhao and Xu, 2008) proposes polynomial kernel NMF for both face and facial expression recognition on JAFFE and CK dataset of facial expression. They identified PNMF is superior to KPCA or KICA although NMF has very competitive result with PCA and ICA based on classifier diversity. They also concluded that NMF algorithm retrieves more powerful latent variables for pattern classification, as evidenced by their experimental results (I. Buciu and Pitas, 2007). The work of (Yeasin and Bullot, 2005) compares the performance of linear and non-linear data projection techniques in classifying facial expressions using Principal Component Analysis (PCA), Non-negative Matrix Factorization (NMF), Local Linear Embedding (LLE) and concludes that 90.9%, 88.7% and 92.3% accuracy for PCA, NMF and LLE respectively. Research work of (L. Zhao and Xu, 2008) and (Zilu and Guoyi, 2009) have also done the similar work for NMF based face and expression analysis.

In this section we have discussed some research approach of face and facial expression recognition using NMF and PCA on 2D images. It is really difficult to compare among several subspace learning algorithms as they have been tested on different datasets. Also the normalization and distance measure varies method to method. One method which has high recognition rate and tested on neutral front faced images may not be logically better than the method with low error rate tested on noisy images with varying head poses and vice versa.

3 PRINCIPAL COMPONENT ANALYSIS

Research shows that Principal Component Analysis is a well established method in subspace learning algorithm areas. It has been implemented widely in machine learning, computer vision and data mining areas. Basically PCA is a linear transformation method, which finds the directions that maximizes the variance of datasets. It projects the dataset in a different subspace without the class labels. Mathematically for an mXm matrix P, Data can be decomposed into $P = P \wedge P^T$, here eigenvector is each column of P and eigenvalue is the diagonal matrix $\wedge$. This way of matrix decomposition is called eigen decomposition. Below are the steps of PCA algorithm. Now we want to finally give PCA algorithm.

It is given

$$D = p^1, ..., p^n. \tag{1}$$

First compute

$$\bar{p} = \frac{1}{n}\sum_i p^i \tag{2}$$

and

$$\sum = \frac{1}{n}\sum_{i=1}^{n}(p^i - \bar{p})(p^i - \bar{p})^T. \tag{3}$$

Then find the k eigenvectors of equation(3)with largest eigenvalues:

$$U_1, ..., U_k \tag{4}$$

These are called principal components Project

$$Z^i = ((p^i - \bar{p})^T U_1, ..., (p^i - \bar{p})^T U_k) \tag{5}$$

It is to be noted that only the top eigenvectors need to be calculated, not all of them, which is a lot faster for computation.

4 NON-NEGATIVE MATRIX FACTORIZATION

In the previous section, many machine learning research shows that Non-negative matrix factorization (NMF) is a useful decomposition for multivariate data like face and facial expression recognition. According to research studies (Lee and Seung, 2009) it is clear that NMF can be understood as part based analysis as it decomposes the matrix only into additive parts. This factorization technique of NMF is completely different of Principal Component Analysis (PCA) or Vector Quantization (VQ) in terms of the nature of the decomposed matrix. PCA and VQ works on holistic features where as NMF decomposes a part based representation of matrix (Lee and Seung, 2009). Here we apply NMF and PCA on whole faces and on different facial parts. PCA, ICA, VQ, NMF all these subspace learning techniques reduces the dimension and make a distributed represented in which each facial image can be approximated using a linear combinations of all or selected basis images.

The factorization problem can be written like this,

$$X \approx W.H \tag{6}$$

where $X \in R^{MxN, \geqq 0}$

This is similar to the PCA or ICA initialization. In the above equation, R defines the low-rank dimensionality. Here $[W]$ and $[H]$ are quite unknown; $[X]$ is the known input source. Now we have to estimate the two factors. We have to start with random $[W]$ and $[H]$. Columns of $[W]$ will contain vertical information about $[X]$ and the horizontal information will be extracted in the rows of $[H]$. NMF does additive decompositions and parts make this decomposition. We first have to define the cost functions to solve an approximate representation of the factorization problem of $X \approx W.H$. By using some measure of distance between two non-negative matrices $[P]$ and $[Q]$, such cost functions can be constructed. The square of the Euclidian distance between the matrices $[P]$ and $[Q]$, is one fruitful measure.

$$||P-Q||^2 = \sum_{i,j}(P_{ij} - Q_{ij})^2 \tag{7}$$

The above equation is lower bounded by zero and absolutely vanishes if and only if $[P] = [Q]$. To define the cost function, another useful representation is,

$$D(P \parallel Q) = \sum_{i,j}(P_{ij} log \frac{P_{ij}}{Q_{ij}} - p_{ij} + Q_{ij}) \tag{8}$$

In the above equation, when $\sum_{i,j} P_{ij} = \sum_{i,j} Q_{ij} = 1$, the above Kullback-Leibler or relative entropy reduces. Here $[P]$ and $[Q]$ can be regarded a normalized probability distribution. Now, following the cost function of equation (2), we have to define it for the input matrix $[X]$ and the non-negative decomposed matrix $[W]$ and $[H]$. If we do that, the cost function would be,

$$\parallel V = WH \parallel^2 \tag{9}$$

The main goal is now to reduce the distance $||V - WH||$. To do that, first we have to initialize $[W]$ and $[H]$ matrix. Then we apply the multiplicative update rule, which is described in the paper of (Lee and Seung, 2009). They claim and prove that the multiplicative update rules minimize the Euclidean distance $||P-Q||$ and also the divergence, $D(P||Q)$ is decreasing when multiplicative update rule is applied. In our programming here, we use the Euclidian distance as a cost function and apply the multiplicative update rule to minimize the distance. The rules are defined below,

$$H_{p\beta} \leftarrow H_{p\beta} \frac{(W^T V)_{p\beta}}{(W^T W H)_{p\beta}} \tag{10}$$

$$W_{\alpha p} \leftarrow W_{\alpha p} \frac{(VH^T)_{\alpha p}}{(WHH^T)_{\alpha p}} \tag{11}$$

According to the mathematical analysis, if we use the equation (5) and (6) to decrease the Euclidian distance $||V - WH||$, the distance $||V - WH||$ converges. Our experimental analysis also shows that and we get a significant output on facial expression dataset.

5 DATASETS

For experimental purpose we implemented our algorithm on both Cohn Kanade and JAFFE Facial Expession dataset. Nearly 2000 image sequences from over 200 subjects are in CK+ dataset. All the expression dataset maintain a sequence from neutral to highest expressive grace. We took two highest graced expressive image of each subject. As we took 100 subjects, so the total image becomes 1200. 100 subject x 6 different expression x 2 of each expression. So it becomes 100 x 6 x 2=1400. There is a significant variation of age group, sex and ethnicity.

In the JAFFE datasat, each of the ten subjects posed for 3 or 4 examples of each of the six basic facial expressions (happiness, sadness, surprise, anger, disgust, fear) as well as a neutral face expression. Altogether JAFFE has 219 facial images, and we used all of these in our experiments.In JAFFE set each subject took pictures of herself while looking through a semi-reflective plastic sheet towards the camera. The following figure (Fig. 1) shows a portion of the dataset of our experiment. Fig. 2 is the prepared data to feed in our fusion based prposed mathod which we want to compare against the predominant NMF method. These segmented dataset is prepared by using our algorithm which is described in the corresponding section and our work (Ali and Powers, 2014). In figure3, The first, second, third and fourth rows show mouth, left eye, and right eye and nose respectively.

Figure 1: CK+ and JAFFE dataset.

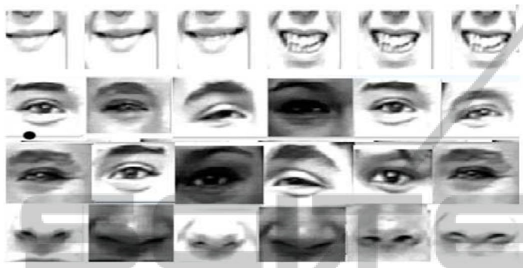

Figure 2: Segmented Four Facial Parts.

6 EXPERIMENTS

6.1 Face and Facial Parts Detection

In CK dataset, the background is large with all the face images. First we apply the Viola-Jones algorithm (Paul and Jones, 2001) to find the faces. For eyes, nose and mouth detection we applied cascaded object detector with region set on already detected frontal faces (Fig. 3). This cascade object detector with proper region set can identify the eyes, nose and mouth. Actually it uses Viola-Jones Algorithm as an underlying system. This object detection algorithm uses a cascade of classifiers to efficiently process image regions for the presence of a target object. Each stage in the cascade applies increasingly more complex binary classifiers, which allows the algorithm to rapidly reject regions that do not contain the target. If the desired object is not found at any stage in the cascade, the detector immediately rejects the region and processing is terminated. This process has been described in our previous work (Ali and Powers, 2014).

We perform face identification and facial parts detection on 1200 images (6 expression X 100 subjects X 2 imagees each images) in CK+ and 219 images on JAFFE dataset. The CK dataset varies greatly in image brightness.For image pre normalization procedure, first we use Contrast Adjustment to enhance the image from very light images. Then to improve the contrast of the very dark images we apply Histogram equalization.

Figure 3: Face and Facial Parts Detection replicated from our previous work (Ali and Powers, 2014).

6.2 Training and Testing Data

For face recognition using NMF we separated the data as 60% for train and 40% for test data. But while using PCA we use 70% of the data as train and 30% of the total data as test images. This is because, PCA gives a good result when the train dataset is large enough than the test portion and this has been seen by our programme. We have given a table in result analysis section in which way the recognition result is depended upon the training sample programmed by PCA and NMF separately.

6.3 Face vs Facial Expression Recognition

We achieved a good recognition result for face recognition using straight PCA and NMF. But result shows NMF has a greater recognition result than PCA. This is because Non-negative matrix factorization (NMF) learns a parts-based representation of faces and part based representation is very suitable for occluded and low intensity or high brightness images.

On the other hand, NMF and also PCA has a poor recognition capability for facial expression compared to face recognition capability. Like some other subspace learning techniques, PCA and NMF tends to find similar faces rather than similar expression. To improve the recognition rate we propose fusion based NMF and PCA algorithms to fuse the parts of faces to recognize the facial expressions and we term it as Optimal Expression-specific Parts Accumulation (OEPA) method. Actually this method has been developed in our previous work (Ali and Powers, 2013). We apply here NMF based OEPA and compare this proposed system with PCA based OEPA.

6.4 Proposed Algorithm

Sometimes a subset of all the four parts of the face is optimal in terms of processing time and accuracy for identifying an expression. In second approach, we adapt similar approach and named it as Optimal Expression-specific Parts Accumulation (OEPA). In case of identifying an expression, if more than one subset of four parts give almost equal accuracy within a threshold value, this algorithm picks the subset of minimal number of parts in order to reduce the processing time. It results in an increased efficiency of the program.

After dividing faces into four facial parts each part has been made of similar size. We make subspace or reduced dimension space for whole face, left eyes, right eyes, nose, mouth. Then we make a fusion of all the different type of combinations of four facial parts. These subspaces have been made for all the six expressions. When a test image needs to classify, we divide it into four facial parts. Then we project it on all the decomposed spaces making different combinations of fusions and take the space with minimum euclidian distance. To compare test data with different face parts and the combination of different parts and whole faces also, the comparison time is more than to compare with only whole face subspaces. But the recognition rate is much much better than the whole face based decomposition and comparison. Pragmatically we can also find the regions of which are face parts are more likely to express a particular expression. We formulate a table which shows the influence of different facial parts for a emoting specific expression.

Also to validate our results, we tested each expression at a time and project it on the whole sets of feature spaces from the whole train dataset which contains a mixture of all six expressions.

The following figure (Fig. 4) is the flow chart of the whole procedure.

6.5 Euclidian Distance Classifier

Here we use euclidian distance to take the minimum distance from the feature subspace. Euclidian distance is the shortest distance between two points on a plane is a straight line and is known as Euclidean distance as shown in the following equation and is a non-parametric classifier. In Euclidean distance metric difference of each feature of query and database image is squared which effectively increases the divergence between them.

$$d_{Euc} = (A,B) = \sqrt{\sum_{k-1}^{m} |A_k - B_k|^2} \qquad (12)$$

In many machine learning data matching areas, euclidian distance classifier (EDC) has been proven a successful classifier. For an example, EDC performs as well as or superior to the sample LDF(inear discriminant function) , even for nonspherical covariance configurations (Marco and Turner., 1987).

Algorithm 1: Pseudocode for Optimal Expression-specific Parts Accumulation (OEPA) approach.

```
procedure OEPA
    Step 1 (Initialization):
    Initialize random population
    Step 2 (Evaluation):
    Let I be the vector [IL, IR, IM,
IN] of an image's subregions (Left
eye, Right eye, Mouth, Nose).

    for i in I do
        Evaluate fitness f(i) where f(i)
is chance-corrected accuracy (kappa)
    Let E be the vector [Hap, Sad,
Disgust, Anger, Fear, Surprise]
representing the six basic emotions.

    for e in E do
        for k = 1 to 4 do
            for P in powerset(I) do
                K(e,k)=argmaxP:|P|=k f(P)
is the best set of k regions for
e.  L(e,k)= maxP:|P|=kf(P) is the
corresponding fitness value.
                K(e)=argmaxk:1-4,P:|P|=k
f(P) is the best set of regions for e.
L(e) = maxk:1-4,P:|P|=k f(P) is the
corresponding fitness.
                K =
argmaxe:E,k:1-4,P:|P|=k f(P) is best
regions and emotion.
    L = max e:E,k:1-4,P:|P|=k f(P) is
the corresponding fitness.
```

7 RESULT ANALYSIS

7.1 Subspace Visualization

NMF works in different approach than PCA. It can be seen through the visual decomposition of both methods. Figure 5 shows a portion of the NMF decomposed faces. The next figure(Fig.6) shows the NMF reduced subspace of several facial parts.

Some Eigenfaces of different expressions (single and mixed expression dataset) and Eigen images of separate face parts are given below in figure 7.

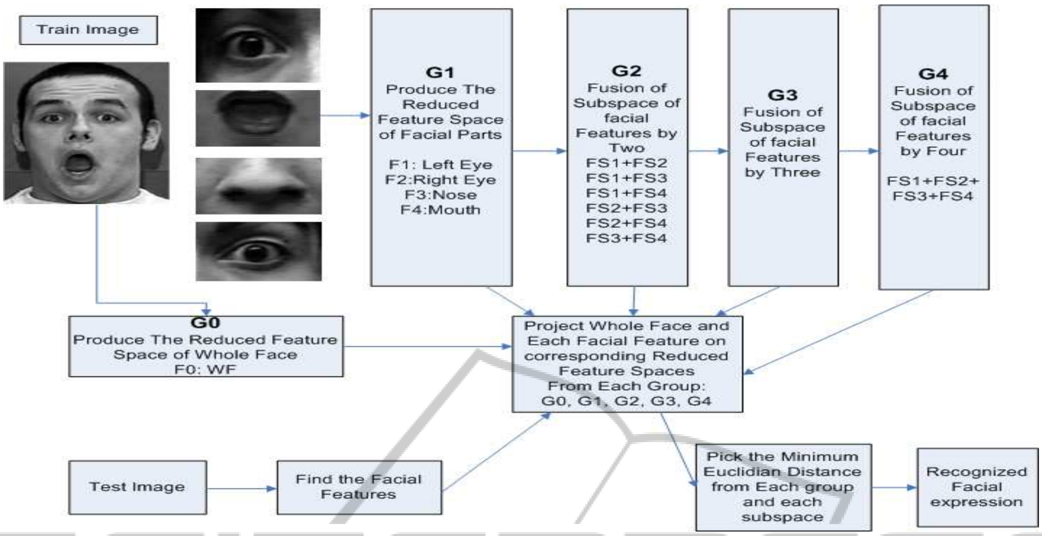

Figure 4: Proposed Method.

Figure 5: A portion of the NMF-decomposed faces from the whole dataset (1200 images).

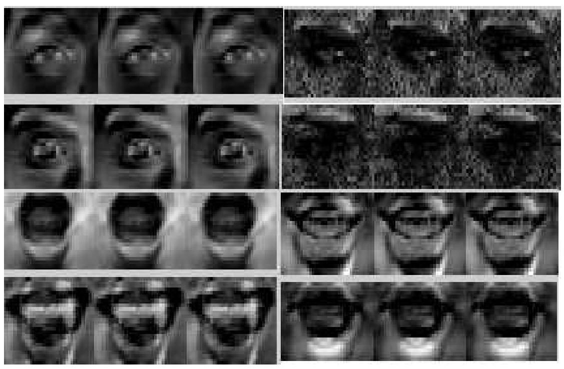

Figure 6: NMF decomposed facial parts.

7.2 Face Recognition

For face recognition, we apply we different variables for segmenting the whole data into train and test folder and do the projection. We use 10% of the as train set and another 90% for testing purpose. Then we increment the train data by 10% each time while deducting the test data by 10%. So it becomes when train dataset is 20%, test dataset is 80%, then 30% train data and 70% for test data and the same thing is happened sequentially.

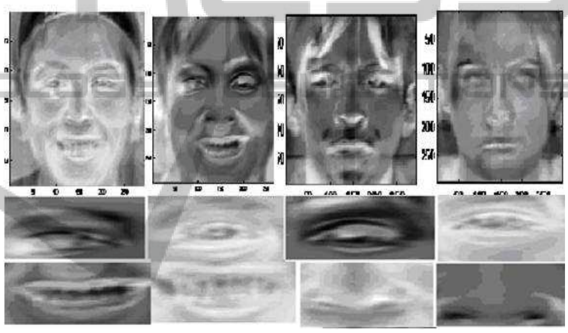

Figure 7: 1st row: first two faces are the 1st and 2nd eigenface of happy faces, 3rd and 4th are the 1st and 2nd eigenface of neutral-angry faces. 2nd row: first two left and right eigen eyes from happy-neutral faces.3rd row: first,2nd two are first eigen mouth of happy faces and 3rd,4th are first,2nd eigen nose of neutral faces.

This way it is found that to achieve a good recognition result using NMF 60% of the data should be trained and decomposed to make a reduced feature space. On the other hand, when PCA has been used nearly 70% of the data should be trained to make the feature space. The recognition result does not only depend upon the accuracy rate but also it depends upon how much data the programme is using for train and test data. To produce the graph(figure.8) we combine all the two datasets of CK+ and JAFFE to get a proper insight. It is shown in the graph that more training sample is needed for PCA based face recognition than NMF based face recognition to get a good result.

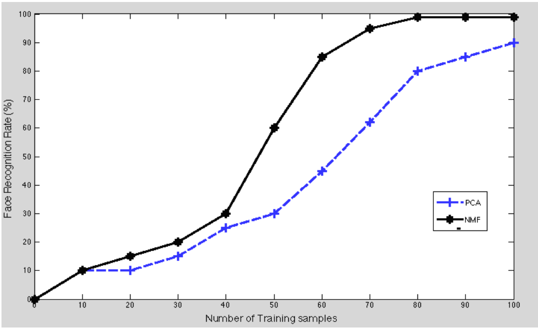

Figure 8: Face recognition rate vs. number of training samples.

Face vs. Facial Expression Recognition	*Dataset/Algorithm*	*Expressions*	*Accuracy with Brightness Adjustment*	*Images/ Subjects*	*Informedness*
Face	CK+/ NMF	N/A	99.00%	1500/100	98.93%
Face	JAFFE /NMF	N/A	96.24%	213/10	96.05%
Face	CK+/ PCA		94.00%	1500/100	91.90%
Face	JAFF/ PCA		90.00%	213/10	87.87%
Facial Expression	CK+/ PCA	Happy	80.00%	1200/100	77.70%
		Sadness	60.00%		57.74%
		Fear	60.00%		57.74%
		Surprise	80.00%		77.70%
		Disgust	70.00%		67,90%
		Angry	80.00%		77.70%
Facial Expression	JAFF/ PCA	Happy	60.00%	210/10	57.74%
		Sadness	58.00%		55.10%
		Fear	28.00%		25.75%
		Surprise	60.00%		57.74%
		Disgust	30.00%		27.77%
		Angry	58.00%		55.10%
Facial Expression	CK+/ NMF	Happy	**85.00%**	1200/100	82.17%
		Sadness	**83.00%**		80.16%
		Fear	**74.50%**		71.91%
		Surprise	**86.00%**		82.50%
		Disgust	**78.00%**		75.83%
		Angry	**83.50%**		80.71%
Facial Expression	JAFFE /NMF	Happy	**66.67%**	219/10	64.11%
		Sadness	**66.67%**		64.11%
		Fear	**33.33%**		30.21%
		Surprise	**70.00%**		67.90%
		Disgust	**33.33%**		30.21%
		Angry	**68.67%**		66.11%

Figure 9: Comparison of NMF and PCA for Face and Facial Expression Recognition.

7.3 Performance Measurement: Accuracy and Informedness

To evaluate the classifier's performance, accuracy is a usual measurement criterion. However, due to the variability of number of classes and bias of the systems, accuracy does not show reliable the measurement. (Powers, 2003) first introduced the concept of informedness which is a concept of probabilistic measurement based on decision, prediction or contingency is informed, rather than due to chance. Therefore we also adopt here informedness besides accuracy to enhance a better understanding of classifier's performance. Accuracy is calculated as the following equation which indicates the proportion of right prediction amount from the whole sample data set.

$$Accuracy = \sum_{i=1}^{m} a_i i / N. \tag{13}$$

Where m is the number of expression (here m=6) and N is the total number of images. To estimate the informedness, bookmaker is an algorithm, which calculates from a contingency table encountering the idea of betting with fair odds (Powers, 2011) and (Powers, 2012). It is shown that informedness subsumes chance corrected accuracy estimates based on other techniques that allow for chance, including Receiver Operating Characters (ROC), Correlation and Kappa, all of which are identical when bias is matched to prevalence. Informedness calculates the probability that the programme makes an informed decision versus guessing. It is calculated by the following equation.

$$Informedness = \frac{winloss}{N} \tag{14}$$

Where $winloss = \sum_{i \neq j}(a_{ij} * bias[j]/(prev[j]-1)) + \sum_{i=j}(a_{ij} * bias[j]/(prev[j]))$ and $prev[i] = X_i/N$, $bias_i = Y_i/N$. For clarity $prev$= prevalence, N is the total samples in the dataset, X_i and Y_i are the derived values which is the number of samples in original and predictaed set correspondingly. Figure. 9 shows the comparison of PCA and NMF based face and facial expression recognition. It is identified that facial expression recognition is not as good as face recognition rate. So we propose OEPA based method and it is shown that our system has a better performance.

7.4 Evaluation: OEPA based Facial Expression Recognition

From Figure. 9 it is clear that PCA and NMF based facial expression recognition rate is not as good as face recognition rate. So we propose OEPA based method and it is shown that our system has a better performance in Table1 and Table2.

This is because that some face parts make the test image to be confused with two or three near similar expressions, like sad and disgust face has some similarity. So for near similar expressions it is tough to get the good recognition rate with whole face or using the fusion of some face parts while some other combination of facial parts have good distinguishable criteria than others. Table1 and Table2 shows the recognition rate of facial expression using OEPA-PCA and OEPA-NMF. This result is from the joint dataset of CK+ and JAFFE. As we achieve good recognition rate for face recognition using PCA and NMF, we are not interested to apply OEPA based method for face recognition.

8 CONCLUSIONS

In this work we propose a multi feature fusion based algorithm to fuse different combination of facia feature subspace and to analyze how it improves the facial expression recognition rate. We name this method as Optimal Expression Specific Parts Accumulation (OEPA). Here our main work is to implement OEPA based NMF algorithm and to compare it with OEPA based PCA. As written before, we only apply OEPA based approach for facial expression recognition, not for face recognition. As for face recognition we achieve a reasonable result using straight PCA and NMF. Oue result shows OEPA-PCA and OEPA-NMF outperforms the predominant PCA and NMF method.

ACKNOWLEDGEMENTS

This work was supported in part by the Chinese Natural Science Foundation under Grant No. 61070117, the Beijing Natural Science Foundation under Grant No. 4122004, and the Australian Research Council under ARC Thinking Systems Grant No. TS0689874 as well as the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions.

REFERENCES

Ali, H. B. and Powers, D. M. W. (2013). Facial expression recognition based on weighted all parts accumulation and optimal expression specific parts accumulation. In *Digital Image Computing Techniques and Applications (DICTA),International Conference on. IEEE,Hobart, Tasmania,Vol.2,No. 1.pp.1-7*.

Ali, H. B. and Powers, D. M. W. (2014). Fusion based fastica method: Facial expression recognition. *Journal of Image and Graphics*, 2(1):1–7.

C. Hesher, A. S. and Erlebacher, G. (2003). A novel technique for face recognition using range images. In *Seventh Intl Symp. on Signal Processing and Its Applications*.

Charlesworth, W. R. and Kreutzer, M. A. (1973). Facial expressions of infants and children. In *In P. Ekman (Ed.), Darwin and facial expression: A century of research in review*, pages 91–138. New York; Academic Press.

Cichocki, Andrzej, e. a. (2009). *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation.* John Wiley and Sons.

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to russell's mistaken critique. In *Psychological Bulletin.*, pages 115(2): 268–287.

Ensari, Tolga, J. C. and Zurada, J. M. (2012). Occluded face recognition using correntropy-based nonnegative matrix factorization. In *Machine Learning and Applications (ICMLA),2012 11th International Conference on IEEE.vol 1*.

Frith, U. and Baron-Cohen, S. (1987). Perception in autistic children. In *Advances in Neural Information Processing Systems*, pages 85–102. Handbook of autism and pervasive developmental disorders.New York: John Wiley.

I. Buciu, N. N. and Pitas, I. (2007). Nonnegative matrix factorization in polynomial feature space. *IEEE Transactions on Neural Network*, 42:300–311.

J. Buhmann, M. J. L. and Malsburg, C. (1990). Size and distortion invariant object recognition by hierarchical graph matching. In *Proceedings, International Joint Conference on Neural Networks*, pages 411–416.

L. Zhao, G. Z. and Xu, X. (2008). Facial expression recognition based on pca and nmf. In *Proceedings of the 7th World Congress on Intelligent Control and Automation, Chongqing, China*.

Lee, D. D. and Seung, H. S. (2009). Learning the parts of objects by non-negative matrix factorization. In *Letters to Nature*, pages 788–791.

Li, J. and Oussalah, M. (2010). Automatic face emotion recognition system. In *Cybernetic Intelligent Systems (CIS), 2010 IEEE 9th International Conference.*

M. Lades, J. Vorbruggen, J. B. J. L. C. M. R. W. and Konen, W. (1993). *Distortion invariant object recognition in the dynamic link architecture*. IEEE Transaction on Computing, 42300-311.

Marco, Virgil R., D. M. Y. and Turner., D. W. (1987). The euclidean distance classifier: an alternative to the linear discriminant function. In *Advances in Neural Information Processing Systems*, pages 485–505. Communications in Statistics-Simulation and Computation.

Moghaddam, B. and Pentland, A. (1997). Probabilistic visual learning for object representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19:696–710.

Paul, V. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition*.

Powers, D. M. W. (2003). Recall and precision versus the bookmaker. In *Interna- tional Conference on Cognitive Science (ICSC-2003)*, page 529534.

Powers, D. M. W. (2011). Evaluation: From precision, recall and f-measure to roc., informedness, markedness

Table 1: Facial Eecognition based on OEPA-PCA(LE:Left Eye, RE:Right Eye, N:Nose, M:Mouth).

F-Parts	Surp.	Ang.	Sad.	Hap.	Fear	Disg.
LE	76%	58%	60%	68%	40%	48%
RE	76%	58%	60%	68%	40%	48%
LE+RE	76%	58%	60%	68%	40%	48%
N	12%	16%	50%	14%	30%	54%
M	**88%**	52%	52%	80%	70%	44%
LE+RE+N	60%	52%	50%	70%	54%	**80%**
LE+RE+M	80%	74%	72%	**86%**	**78%**	70%
N+M	64%	44%	44%	60%	40%	62%
LE+RE+N+M	76%	**86%**	**82%**	74%	74%	68%
OEPA-PCA	**88%** *(M)*	**86%** *(LE+RE +N+M)*	**82%** *(LE+RE +N+M)*	**86%** *(LE+RE) +M)*	**78%** *(LE+RE +M)*	**80%** *(LE+RE +N)*

Table 2: Facial Eecognition based on OEPA-NMF(LE:Left Eye, RE:Right Eye, N:Nose, M:Mouth).

F-Parts	Surp.	Ang.	Sad.	Hap.	Fear	Disg.
LE	84%	67%	68%	72%	44%	57%
RE	84%	67%	68%	72%	44%	57%
LE+RE	84%	67%	68%	72%	44%	57%
N	18%	20%	57%	18%	36%	59%
M	**96%**	52%	58%	88%	84%	54%
LE+RE+N	64%	58%	52%	78%	60%	**89%**
LE+RE+M	89%	82%	78%	**92%**	**88%**	80%
N+M	74%	44%	44%	60%	40%	72%
LE+RE+N+M	86%	**90%**	**86%**	85%	83%	80%
OEPA-NMF	**96%** *(M)*	**90%** *(LE+RE +N+M)*	**86%** *(LE+RE +N+M)*	**92%** *(LE+RE) +M)*	**88%** *(LE+RE +M)*	**89%** *(LE+RE +N)*

and correlation. *Journal of Machine Learning Technologies*, 2(1):3763.

Powers, D. M. W. (2012). The problem with kappa. In *Conference of the Euro- pean Chapter of the Association for Computational Linguistics,*. Avignon France.

Turk, M. A. and Pentland, A. P. (1991). Face recognition using eigenfaces. *Proceedings of International Conference on Pattern Recognition*, pages 586–591.

Yeasin, M. and Bullot, B. (2005). Comparison of linear and non-linear data projectiontechniques in recognizing universal facial expressions. In *proceedings of International Joint Conference on Neural Networks*.

Zilu, Y. and Guoyi, Z. (2009). Facial expression recognition based on nmf and svm. In *2009 International Forum on Information Technology and Applications*.