

C-35

by Armin Lawi

Submission date: 07-Nov-2021 12:02PM (UTC+0700)

Submission ID: 1695204113

File name: C.1 c-35 Facial expression recognition using multiclass ensemble least-square support vector machine.pdf (883.96K)

Word count: 3425

Character count: 17598

PAPER · OPEN ACCESS

Facial Expression Recognition using Multiclass Ensemble Least-Square Support Vector Machine

To cite this article: Armin Lawi and M Sya'Rani Machrizzandi 2018 *J. Phys.: Conf. Ser.* **979** 012032

View the [article online](#) for updates and enhancements.

You may also like

- [Facial Expression Recognition Based on Gabor Multi-orientation Feature Fusion](#)
Xuechen Liang, Lei Yang and Shengda Luo
- [Research on Facial Expression Recognition Based on Active Learning](#)
Yan Wan, Hongjie Ni, Xiaodong Peng et al.
- [Computational Simulation on Facial Expressions and Experimental Tensile Strength for Silicone Rubber as Artificial Skin](#)
Andi Amijoyo Mochtar



IOP ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Facial Expression Recognition using Multiclass Ensemble Least-Square Support Vector Machine

Armin Lawi¹, M Sya'Rani Machrizzandi²

¹Department of Computer Science, Hasanuddin University, Makassar

²Department of Informatics, Al-Asy'ariah Mandar University, Polewali Mandar

E-mail: machrizandi@gmail.com

Abstract. Facial expression is one of behavior characteristics of human-being. The use of biometrics technology system with facial expression characteristics makes it possible to recognize a person's mood or emotion. The basic components of facial expression analysis system are face detection, face image extraction, facial classification and facial expressions recognition. This paper uses Principal Component Analysis (PCA) algorithm to extract facial features with expression parameters, i.e., happy, sad, neutral, angry, fear, and disgusted. Then Multiclass Ensemble Least-Squares Support Vector Machine (MELS-SVM) is used for the classification process of facial expression. The result of MELS-SVM model obtained from our 185 different expression images of 10 persons showed high accuracy level of 99.998% using RBF kernel.

1. Introduction

Facial expression in everyday human life is used as a form of natural human response that describes the feelings felt by a person in interacting with a certain thing. In the interaction between fellow human, expression is use as a part of communication. On the other hand, technological developments from time to time have great leaps when a new interaction is implemented such as the use of keys, monitor screen, GUI, touch screen, to voice commands. The key to technological development lies in the user, therefore computer interaction is an important part in technological development and understanding the user's response becomes very important. One approach taken to understand the response of the user is by a facial recognition system. By understanding the expression of the user, the computer can be made as if it can understand human feelings.

The research on facial expressions has been started since 1978 [1] and was extensively researched in the computer science area of the 90s and continues to the present day. an expression recognition system of two important stages, namely facial features representation and classification design. Facial feature representation is obtained by extracting input in the form of digital image by using certain method. The method of feature extraction used by researchers to build a facial expression recognition system that is Principal Component Analysis Algorithm (PCA). Principal Component Analysis (PCA) algorithm is one method that can be used to process the image of a person's face so that the system will automatically recognize a person's face through its main features such as eyes, nose, lips, and eyebrows as identity [2]. The identity of the person's face image by the system will be recognized through various training (training) stored in the database. The training phase is the result of extraction from different collections of different faces then collected and stored in a database. The results of facial images that have been extracted using PCA algorithm will be compared with the new face image



as the face image to be tested and testing whether it has similarity or almost similar to recognizable by the system [3].

2. Related Works

Yubo Wang, *et al.*, [4] have proposed a new method of recognition of facial expressions. The facial expressions are extracted from the human face by the expression classifier learned from the Haar feature based on the LUT Weak Classifiers. The expression recognition system consists of three modules, face detection, facial feature landmark extraction and facial expression recognition. This system is applied automatically can recognize seven expressions in real time that include anger, disgust, fear, pleasure, neutrality, sadness and surprise. Reda Shbib and Shikun Zhou in [5] studied the introduction of facial expressions using the Active Shape Model method by adopting AdaBoost classifier and Haar-Like feature to detect images. AdaBoost is a method to increase classifier accuracy gained from the component learner of the Support Vector Machine (SVM) method [6, 7, 8].

Data mining is a process that uses statistical techniques, calculations, artificial intelligence and machine learning to extract and identify useful information and related knowledge from large databases [9]. SVM initially can only classify data in two classes [8, 10, 11]. However, further research SVM was developed so that it can classify data over two class (multiclass) [12, 13, 14]. Classifying M-classes means predicting the class labels $C_m, m = 1, \dots, M$ one way to solve the M-class problem by formulating it into binary L classification problems [15]. SVM concept is simply trying to find the best hyperplane that serves as a separator of two classes in the input space. Pattern that is a member of two classes into +1 and -1, and share alternate field separators. The best dividing fields can not only separate the data but also have the largest margins. Margin is the distance between the fields of separator (hyperplane) with the closest pattern of each class.

Let $\{x_1, \dots, x_n\}$ be the dataset and $y_i \in \{+1, -1\}$ is the class label of the x_i data. A pair of parallel bounding plane separates the two classes. The first delimiter field limits the first class while the second delimiter field limits the second class, it is obtained therefore as follows [16].

$$\begin{aligned} x_i \cdot w + b &\geq +1 \text{ for } y_i = +1 \\ x_i \cdot w + b &\leq -1 \text{ for } y_i = -1 \end{aligned} \quad (1)$$

The best separator field has the largest margin values that can be formulated into a quadratic programming problem as follows.

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (2)$$

with constraints: $y_i(x_i \cdot w + b) \geq 1 - \xi_i$ and $\xi_i \geq 0$,

where $l = 1, \dots, n$ is a slack variable that determines the level of misclassification of the data samples. Whereas, $C > 0$ is a parameter.

Kernel method is used when the ordinary SVM classifier cannot separate the data linearly. The kernel method transforms the data into the features space dimension so that it can be linearly separated on the features space. The kernel method can be formulated as follows.

$$K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j) \quad (3)$$

Commonly used kernel functions are as follows.

- The linear kernel: $K(x_i, x_j) = x_i^T x_j$
- Kernel polynomial: $K(x_i, x_j) = (\gamma \cdot x_i^T x_j + r)^p, \gamma \geq 2$
- RBF Kernel (Radial Basis Function): $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$

LS-SVM is one of the SVM modifications that is solved linearly [7, 8, 11, 13, 14]. If the SVM separator field is given as in (3), then the LS-SVM is given as follows.

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + \frac{C}{2} \xi^T \xi$$

with constraint: $y_i(x_i \cdot w + b) \geq 1 - \xi_i$ (4)

The above equation can be solved with Lagrangian multiplier.

$$L = \frac{1}{2} \|w\|^2 + \frac{C}{2} \xi^T \xi - \sum_{i=1}^l \alpha_i (y_i(\varphi(x_i) \cdot w + b) - 1 + \xi_i)$$
 (5)

where the Lagrangian multiplier α_i can be either positive or negative.

To optimize the conditions in (5), we decrease of w , b , ξ , and α is equal to zero. The results of the process are as follows.

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^l \alpha_i y_i \varphi(x_i)$$
 (6)

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^l \alpha_i y_i = 0$$
 (7)

$$\frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha = \gamma \xi_i, i = 1, \dots, N$$
 (8)

$$\frac{\partial L}{\partial \alpha} = 0 \rightarrow y_i(\varphi(x_i) \cdot w + b) - 1 + \xi_i = 0, i = 1, \dots, N$$
 (9)

Using the One-Against-All (OAA) method, k binary model can be constructed (in which k is the number of classes). Each i -class model is trained by using the entire data. For example, there is a classification problem with 3 classes. Training step constructs 3 pieces of binary classification with the following objective function.

$$\min_{w_i, b_i, \xi_{t,i}} \frac{1}{2} \sum_i^m (w_i)^T w_i + \frac{C}{2} \sum_i^m \xi_{t,i}^2$$

with constraint: $y_{t,i}(\varphi_i(x_t)(w_i)^T + b_i) \geq 1 - \xi_{t,i}$. (10)

Confusion matrix is a table that expresses the amount of test data that is correctly classified and misclassified. The table using the following parameters in order to calculate accuracy, sensitivity and false discovery rate. True Positive (TP) is the number of documents from class 1 is correctly classified as class 1. True Negative (TN) is the number of documents from class 0 is correctly classified as class 0. False Positive (FP) is the number of documents from class 0 incorrectly classified as class 1. False Negative (FN) is the number of documents from class 1 that are misclassified as class 0.

Performance measurements of accuracy, sensitivity and false discovery rate can be calculated using the following formulas respectively.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100\%$$
 (11)

$$Sensitivity = \frac{TP}{TP + FN} \times 100\%$$
 (12)

$$False Discovery Rate = \frac{FP}{FP + TP} \times 100\%$$
 (13)

3. Experimental Details

The process of facial classification is explained by the following activities:

1. Face image capture
2. Face image processing
3. PCA feature extraction
4. Perform Data Training Process
5. Pattern model Classification in get after doing the data training process.
6. Conducting Testing Process/Data Classification after obtained model pattern from result of training process data.

3.1. Principal Component Analysis (PCA)

The sequence of features extraction process using PCA is given as follows.

1. Read input data

2. Calculate the average of each column

Input of all image of training data is input parameter on PCA function that is "ImageTraining" mean column image can be calculated using matlab program code:

```
I = imread ('YM_AN1_61.tiff');
```

```
Average = mean (I);
```

3. Calculate the zero mean

To calculate the zero mean, the average matrix of training data must be doubled as much as the sample data, which is multiplied by 240 lines with the command:

```
ZeroMean = bsxfun (@ minus, 256, Average);
```

4. Calculate the covariance

To calculate the covariance matrix by multiplying the zero mean against its transpose, and the result is divided by "Amount Sample-1" with the command:

```
CovMatrix = (1 / (10-1)) * (ZeroMean * ZeroMean ');
```

5. Calculate eigen value and eigen vector

The dimensions resulting from the covariance matrix have 240 row dimensions and 240 column dimensions. The dimension has been significantly reduced from 10304 to 240, so the dimension number has decreased by 10064 (10304-240 = 10064). The value of the covariance matrix is used further process to determine the value of eigen value and eigen vector with the command:

```
[eigvector, eigvalue] = eig (CovMatrix);
```

6. Sort the eigen value in descending order

To perform a descending order of "eigen value", simply take the diagonal value followed by the sequencing process. The "eigen value" sorting result is stored in a variable, that is, the "Result" variable and the sorting index are stored in the "Index" variable with the command:

```
eigvalue = diag (eigvalue);
```

```
[Results, index] = sort (eigvalue, 'descend');
```

```
eigvalue = eigvalue (index);
```

7. Sort the eigen vector column to change as the eigen value index changes

8. Check the value of eigen value close to 0

The most non-dominant eigenvector values tend to have a correlation of the value of eigen value that is close to zero, therefore it is necessary to check the value of eigen value and the number of close to zero with the command:

```
NotDominant = find (eigvalue < 0.0000000001);
```

```
Amount = length (Not Dominate);
```

9. Calculate the projection matrix

The projection matrix can be calculated by multiplying the transpose data variable by "eigen vector". The code of the program is as follows:

```
ProjectionMatrix = ZeroMean ' * eigvector;
```

The result of the program code is used to form a diagonal matrix with the command:

```
RootSumProjectionMatrix = (1 ./ (sum (ProjectionMatrix.^ 2). ^ 0.5));
```

New projection matrix can be obtained with the command:

```
ProjectionMatrix = (bsxfun (@ times, ProjectionMatrix, RootSumProjectionMatrix));
```

10. Throw away the eigen vector column whose position is equal to the eigen value position which has a value close to 0.

If the value of eigen value is found to be close to zero, then the eigenvector column corresponding to the eigen value position need not be used for the next process with the command:

```
eigvector = eigvector (:, 1: end-Amount);
```

3.2. Dataset

We conducted a simulation for the proposed method; i.e., face detection, feature extraction using PCA algorithm, and facial expression classification process using Ensemble Multiclass Least-Square Support Vector Machine (LS-SVM). The data used in this simulation is data about facial expressions from the Olivetty Research Laboratory (ORL) and Jaffe Images can be downloaded, respectively, on the following URL addresses.

1. <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html> [16], and
2. <http://www.kasrl.org/jaffe.html> [17] and

The type of expression is divided into six classes, i.e., angry, disgust, fear, happy, neutral, and sad. The sample data on this dataset amounts to 240 images of facial expressions.

3.3. Classification Process

In the process of expression classification, starting with image taking of the file, then image is processed using image processing software to get numerical data value which will be used as input data for learning process and validation. After the learning data in the next can be the result of data learning is used for the testing process.

3.4. Multiclass Algorithm Method

The analysis used in this multiclass classification is the One Against All classification method. The multilingas method algorithm as follows:

1. Dataset input.
2. Identify the input dataset
 - a. The values of the training data feature (x_i)
 - b. Class of training data (y_i)
 - c. Values feature test data (x_t)
 - d. Class of test data (y_t)
3. Initiate objects on LS-SVM before performing the training process with the `initlssvm` function
 - a. Specify data of training data feature (x_i)
 - b. Specifies the training data class (y_i),
 - c. Choose a classifier to classify data
 - d. Selects the *kernel* and its parameters to use
4. Selecting the multilingual method code used (`code OneVsAll` for *One Against All*)
5. Conduct training process with `train lssvm` function
6. Calculating values w
7. Make predictions based on the model obtained and determine data feature test data (x_t) with `simlssvm` function
8. Create a *confusion matrix*
9. Calculate the level of accuracy with the formula:

$$\lambda = \frac{C}{N} \times 100\%$$

where C is the correct total of predictions and N is the total of all data tested.

3.5. Implementation Results on dataset

The separator function for the one against all method with the RBF (Radial Basis Function) kernel using the parameter $\sigma = 0.5$ for dataset is as follows.

$$f_1(\mathbf{x}) = \mathbf{x}_i \cdot \mathbf{w}_1 + b = \mathbf{x}_i \cdot \begin{pmatrix} 10,5967 \\ 361,2859 \\ 31,8372 \\ 15,3328 \end{pmatrix} + (-0.8302), \quad (14)$$

$$f_2(x) = x_i \cdot w_2 + b = x_i \cdot \begin{pmatrix} 149,1081 \\ 285,1159 \\ 35,5804 \\ 148,3523 \end{pmatrix} + (-0.5724), \tag{15}$$

$$f_3(x) = x_i \cdot w_3 + b = x_i \cdot \begin{pmatrix} 60.8802 \\ 352,8413 \\ 71,1501 \\ 64,0558 \end{pmatrix} + (-0.6154). \tag{16}$$

$$f_4(x) = x_i \cdot w_1 + b = x_i \cdot \begin{pmatrix} -39,1510 \\ 324,6996 \\ 55,4214 \\ -38,4858 \end{pmatrix} + (-0.4468), \tag{17}$$

$$f_5(x) = x_i \cdot w_2 + b = x_i \cdot \begin{pmatrix} 4,7389 \\ 352,3402 \\ 81,7228 \\ 3,0785 \end{pmatrix} + (-0.8428), \tag{18}$$

$$f_6(x) = x_i \cdot w_3 + b = x_i \cdot \begin{pmatrix} -19,9091 \\ 376,3541 \\ 81,1863 \\ -19,6398 \end{pmatrix} + (-0.6924). \tag{19}$$

4. Results and Discussion

The values of w and b based on the kernel type and their parameters for the use of the One Against All method on the expression dataset can be seen in table 1:

Table 1: w and b values.

kernel dan parameter	w value			b value		
	w_1	w_2	w_3	b_1	b_2	b_3
RBF $\alpha = 0.5$	10,5967	149,1081	60.8802	-0.8302	-0.5724	-0.6154
	361,2859	285,1159	352,8413			
	31,8372	35,5804	71,1501			
	15,3328	148,3523	64,0558			
kernel dan parameter	w value			b value		
	w_4	w_5	w_6	b_4	b_5	b_6
RBF $\alpha = 0.5$	-39,1510	4,7389	-19,9091	-0.4468	-0.8428	-0.6924
	324,6996	352,3402	376,3541			
	55,4214	81,7228	81,1863			
	-38,4858	3,0785	-19,6398			

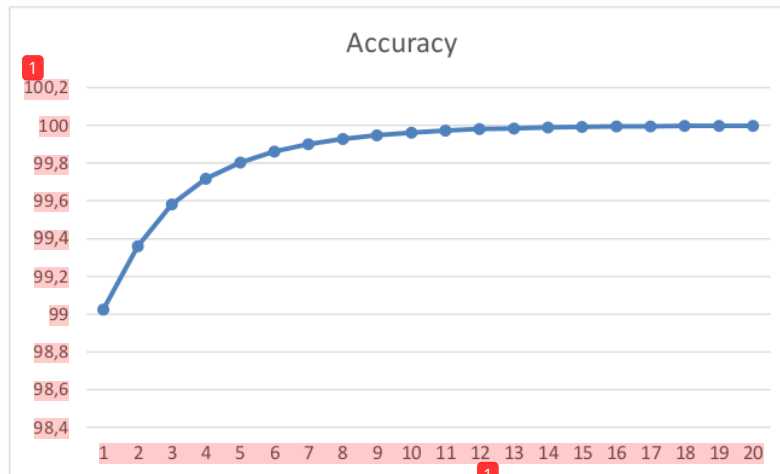


Figure 1. Accuracy level of the OAA for each kernel in dataset.

Based on table 1 and figure 1, the use of RBF kernel ¹ and using the parameter $\sigma = 0.5$ has the highest accuracy on dataset has the highest accuracy of 99.99832%. Accuracy for classification with the 6 number of classes.

5. Conclusion

The facial expression recognition have been proposed using the Multiclass Ensemble Least-Squares Support Vector Machine (MELS-SVM) method. The image of the face is extracted using Principal Component Analysis (PCA) algorithm to take the value of Projection Matrix and Weight Matrix. The classification process uses the MELS-SVM to construct the classifier model from the training images. Finally, the testing images is used to measure accuracy using RBF kernel and the our result showed the accuracy level is 99.99832%.

Acknowledgments

The authors would like to express heartfelt thanks to Department of Computer Science, Hasanudin University and the Department of Informatics, Al-Asy'ariah Mandar University, Polewali Mandar for the research support and assistances.

References

- [1] Ojala T, Pietikäinen M and Harwood D 1994 *Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR1994)* **1** p 582 – 585
- [2] Pratt, William K 2007 *Digital image processing, PIKS Scientific inside* (John Wiley) 4th Edition
- [3] Nugraheny D *Hasil Ekstraksi Algoritma Principal Component Analysis (PCA) Untuk Pengenalan Wajah* (Yogyakarta: STTA)
- [4] Wang Y, Ai H, Wu B, Huang C 2004 *Real Time Facial Expression Recognition With Adaboost* (China : Tsinghua University)
- [5] Shbib R, Zhou S *Facial Expression Analysis using Active Shape Model* (United Kingdom: School of Engineering, University of Portsmouth)
- [6] Kumiawan D and Supriyanto C 2013 *Journal Teknologi Informasi*
- [7] Marliani R R, Lawi A and Thamrin S A 2016 *Algoritma Adaboost untuk Optimasi Ensemble*

- Least Squares Support Vector Machine* (Makassar : Hasanuddin University)
- [8] Zhou L, Lai K K and Yu L 2010 *J. Expert Systems with Applications* 127-133
 - [9] Turban E et.al 2005 *Decision Support Systems and Intelligent Systems* (Yogyakarta: Andi Offset) (Bahasa Indonesia translation version)
 - [10] Vapnik V and Cortes C 1995 *Support-Vector Networks. Machine Learning* 273-297
 - [11] Xu Y, Lv X, Wang Z and Wang L 2014 *Journal of Information Science and Engineering* 1773-1787
 - [12] Sembiring K 2007 *Tutorial SVM Bahasa Indonesia* (Bandung: Institut Teknologi Bandung)
 - [13] Jafar N, Thamrin S A, Lawi A 2016 *Proceeding of International Conference on Computational Intelligence and Cybernetic*
 - [14] Nurkamila J, Lawi A and Thamrin S A 2016. *Least Squares Support Vector Machine Untuk Klasifikasi Data Multikelas* (Makassar : Hasanuddin University)
 - [15] Gestel T, Suykens J, Lanckriet G, Lambrechts A, Moor B and Vandewalle J 2002 *Multiclass LS-SVMs: Moderated Outputs and Coding-Decoding Schemes. Neural Processing Letters* 45-58
 - [16] <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>
 - [17] <http://www.kasrl.org/jaffe.html>

ORIGINALITY REPORT

42%
SIMILARITY INDEX

41%
INTERNET SOURCES

9%
PUBLICATIONS

5%
STUDENT PAPERS

PRIMARY SOURCES

1 **eudl.eu** **39%**
Internet Source

2 **es.scribd.com** **1%**
Internet Source

3 **Submitted to King Fahd University for
Petroleum and Minerals** **1%**
Student Paper

4 **Submitted to University of Southampton** **1%**
Student Paper

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On