On Identifying Facial Expression using Multiclass Ensemble Least-Squares Support Vector Machine

M Sya'Rani[1], Armin Lawi[2], Rayuwati[3] and Mursalin[4]
{armin@unhas.ac.id}

[1]Department of Informatics, Al-Asy'ariah Mandar University, Polewali Mandar, Indonesia
[2]Department of Computer Science, Hasanuddin University, Makassar, Indonesia
[3]Informatics program study, Universitas Gajah Putih, Aceh, Indonesia
[4]Department of Mathematics Education, Universitas Malikussaleh, Aceh Utara, Indonesia

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 Linear Discriminant Analysis (LDA) 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.

Keywords: The Facial Expression, Multiclass Ensemble Least-Squares, Vector Machine

1 Introduction

Expression is one form of nonverbal communication that is the result of one or more movements or muscle position on the face and can convey the emotional state of a person to the person who observes it. Through facial expressions, it can be understood emotions that are churning on the individual self. Facial expression is one of the characteristics of behavior. The use of biometrics technology system with facial expression characteristics makes it possible to recognize a person's mood or emotion.

Research on facial recognition has been started since the 1970s [1]. In generalthere are 4 stages in the face recognition system is the acquisition of face image, preprocessing, feature extraction and classification of the face image. Of the four stages, the feature extraction and classification stages are the most important stages in facial recognition systems [2].

There are many feature extraction methods that have been developed to build facial recognition systems. Of all these methods, each has its own advantages and disadvantages. There are methods that require fast computing time but sacrifice recognition accuracy, and there is also a method with a high degree of recognition accuracy but requires a long computational time. Both computational time and accuracy of the face recognition system are both influenced by the method used at the feature extraction stage [2] and the method used in the classification stage [3]. Therefore, to optimize computational time without sacrificing

recognition accuracy, this research builds facial recognition recognition system by focusing on both phases of feature extraction stage and classification step.

This method is performed by using Linear Discriminant Analysis (LDA) algorithm to extract facial expressions of each image and Multiclass Ensemble Least-Square Support Vector Machine for facial classification.

2　Literature Review

Previous research as a basis, support or reference in this research comes from national and international academic journals published by various universities. In previous research and research that is now done there are some similarities and differences, i.e.

Yubo WANG, Haizhou AI, Bo WU, Chang HUANG, Department of Computer Science and Technology, Tsinghua University, China conducted a study entitled "Real Time Facial Expression Recognition with Adaboost". In this study the researchers 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. [4]

RedaShbib, Shikun Zhou, School of Engineering, University of Portsmouth, United Kingdom conducted a study entitled "Facial Expression Analysis using Active Shape Model". In this study the introduction of facial expressions using the Active Shape Model method by adopting AdaBoost classifier and Haar-Like feature to detect images. The one where AdaBoost classifier is a feature of the Support Vector Machine (SVM) method. [5]

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 [5]. SVM initially can only classify data in two classes. However, further research SVM was developed so that it can classify data over two classes (multiclass) [6]. Classifying M-classes means predicting the class label $C_m, m = 1, \ldots, M$ one way to solve the M-class problem by formulating it into binary L classification problems [7].

SVM concept is simply described as trying to find the best hyperplane that serves as a separator of two classes in the input space. Pattern which is a member of two classes: +1 and -1 and share alternate field separators. The best dividing fields can notonly separate the data but also have the largest margins. Margin is the distance between the field of separator (hyperplane) with the closest pattern of each class.

Let $\{x_1, \ldots, x_n\}$ be the dataset and $y_i \in \{+1, -1\}$ is the class label of the x_i data. The two classes are separated by a pair of parallel bounding plane. The first delimiter field limits the first class while the second delimiter field limits the second class, so it is obtained [8]:

$x_i.\, w + b \geq +1$ for $y_i = +1$

$x_i.\, w + b \leq -1$ for $y_i = -1$　　　　　　　　　　　　　　　　　(1)

The best dividing fields with the largest margin values can be formulated into quadratic programming problems:

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

with constraints $y_i(x_i.w + b) \geq 1 - \xi_i$ and $\xi_i \geq 0$, where $l = 1,..,n$ is a slack slack variable that determines the level of misclassification of the data samples. While $C > 0$ is a parameter.

Method for classifying data that can not be separated linearly is kernel method. The kernel method transforms the data into the feature space dimension so that it can be linearly separated on the feature space. The kernel method can be formulated:

$$K(x_i.x_j) = \varphi(x_i).\varphi(x_j) \tag{3}$$

Commonly used kernel functions are as follows:

a. The linear kernel: $K(x_i, x_j) = x_i^T x_j$

b. Kernel polynomial:

$$K(x_i, x_j) = (\gamma.x_i^T x_j + r)^p, \gamma \geq 2$$

c. RBF Kernel (Radial Basis Function): $K(x_i, x_j) = exp\left(-\gamma\|x_i - x_j\|^2\right), \gamma > 0$

LS-SVM was first introduced by Suykens and Vandewalle in 1999. LS-SVM is one of the SVM modifications that solves linear equations. If the SVM separator field is given as in (3), then for LS-SVM is given as follows:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + \frac{C}{2}\xi^T\xi \tag{4}$$

with constraint $y_i(x_i.w + b) \geq 1 - \xi_i$

The above equation can be solved after forming Lagrangian:

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

where α_i is a Lagrangian multiplier whose value can be either positive or negative.

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

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

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

$$\frac{\partial L}{\partial \xi} = 0 \rightarrow \alpha = \gamma\xi_i, \ i = 1, ..., N \tag{8}$$

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

Using the One Against All (OAA) method, a binary binary model is constructed k (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. For training use 3 pieces of binary classification. Its objective function:

$$\min_{w_i, b_i, \xi_{t,i}} \frac{1}{2}\sum_i^m (w_i)^T w_i + \frac{C}{2}\sum_i^m \xi_{t,i}^2 \tag{10}$$

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

Confusion matrix is a table that states the amount of test data that is correctly classified and the amount of test data being misclassified.

True Posstive (TP), ie the number of documents from class 1 is correctly classified as class 1.

True Negative (TN), ie the number of documents from class 0 is are correctly classified as class 0.

False Positive (FP), ie 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.

The calculation of accuracy is expressed by the equation:

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \times 100\%$$

The calculation of sensitivity is expressed by the equation:

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

The calculation of false discovery rate is expressed by the equation:

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

3   Methods

The process of facial classification is explained by the following activities:
1. Face image capture
2. Face image processing
3. LDA 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   Linear Discriminant Analysis (LDA)

The sequence of features extraction process using LDA as follows:
1. Enter the record number of training pose into the variable.
   NumberPoseTraining= InputLda.PostPoseTraining;
2. Enter the record number of classes into variables.
   Number of Classes = InputLda.Number of Classes;
3. Calculate the total amount of sample data.
   Number of Data = Number of Classes * NumberPoseTraining;
4. Cut the number of columns by the number of classes.
   NewProjectionsProject = InputLda.ProjectionMatric (: 1,:AmountData-TotalClass);
5. Forming Lda input
   Data EntryLda = double (InputLda.Faceline Training) * NewProject Projection;
6. Calculate the average of each class, from first grade to last class.
   Counter = 1;
   For k = 1NumberPoseTraining: NumberData
   Start = k
   Up = k + NumberPoseTraining-1;
   Avg Class (Counter,:) = mean (Data EntryLda (Start: Until, :));
   Counter = Counter + 1;
   end
7. Calculate the total averages.
   FlatAll = mean (Data EntryLda);
8. Calculating Between-Class Scatter (Sb)
   Sb = 0;
   for k = 1: Number Class

```
Zk = Classes (k,:) – AvgEverything;
Sb = Sb + Zk '* Zk ;.
end
```

9. Calculating Within-Class Scatter (Sw).
```
Sw = 0;
for j = 1: Number Class
for k = 1: NumberData
Zm = DataInputLda (k ':) - Avg Class (j' :);
Sw = Sw + NumberPoseTraining * Zm '* Zm;
end
end
```

10. Calculate eigen vector and eigen value.
```
[LdaVector, LdaEigValue] = svd (Sb * inv (Sw));
```

11. Reduced features up to number = Number of Classes-1.
```
wf1d = transpose (LdaVector(:, 1: NumberClass-1));
```

12. Calculate the matrix weight.
To obtain the weight of the new projection matrix value used to project the test data when the image test is performed with the command:
```
WeightMatric = ZeroMean '.*ProjectionMatric;
AveragePM = mean (ProjectionMatric);
AverageWM = mean (weightMatric);
```

3.2   Dataset

This type of research is simulation. Beginning with problem formulation, face detection, feature extraction using LDA algorithm, and facial expression classification process using Multiclass Ensemble Least-Square Support Vector Machine (MELS-SVM).

The data used in this research is data about facial expressions. The facial image used is the Olivetty Research laboratory known as ORL and Jaffe Image. Both data can be downloaded on the following sites:

1. http://www.kasrl.org/jaffe.html [9]
2.http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html [10]

This study uses 1 dataset with the output of facial expression. This type of expression is divided into six classes namely, first class angry, second class disgust, third class fear, fourth class happy, fifth class neutral, sixth grade 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 Methods

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 (yi)
   c. Values feature test data (xt_i)
   d. Class of test data (yt_i)
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 trainlssvm function
6. Calculating values w
7. Make predictions based on the model obtained and determine data feature test data (xt_i) 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:

$$f_1(\boldsymbol{x}) = \boldsymbol{x}_i.\boldsymbol{w_1} + b = \boldsymbol{x}_i.\begin{pmatrix} 31,1296 \\ 372,7066 \\ 82,6486 \\ 36,3001 \end{pmatrix} + (-0.9835),$$

(1.1)

$$f_2(\boldsymbol{x}) = \boldsymbol{x}_i.\boldsymbol{w_2} + b = \boldsymbol{x}_i.\begin{pmatrix} 132,8017 \\ 345,0284 \\ -22,2973 \\ 131,2886 \end{pmatrix} + (-0.3442),$$

(1.2)

$$f_3(\boldsymbol{x}) = \boldsymbol{x}_i.\boldsymbol{w_3} + b = \boldsymbol{x}_i.\begin{pmatrix} 57,8676 \\ 408,6087 \\ 77,2790 \\ 61,6941 \end{pmatrix} + (-0.4530).$$

(1.3)

$$f_4(x) = x_i.w_1 + b = x_i.\begin{pmatrix} -59,5390 \\ 435,8195 \\ 98,1523 \\ -59,0268 \end{pmatrix} + (-0.4588),$$

(1.4)

$$f_5(x) = x_i.w_2 + b = x_i.\begin{pmatrix} 6,5393 \\ 401,7964 \\ 81,7738 \\ 3,9947 \end{pmatrix} + (-1.0626),$$

(1.5)

$$f_6(x) = x_i.w_3 + b = x_i.\begin{pmatrix} -24,7938 \\ 445,8366 \\ 65,8424 \\ -24,0753 \end{pmatrix} + (-0,7571),$$

(1.6)

(1
.6)

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	$\alpha = 0.5$	31,1296 372,7066 82,6486 36,3001	132,8017 345,0284 −22,2973 131,2886	57,8676 408,6087 77,2790 61,6941	−0.9835	−0.3442	−0.4530

Table. 1 Continued

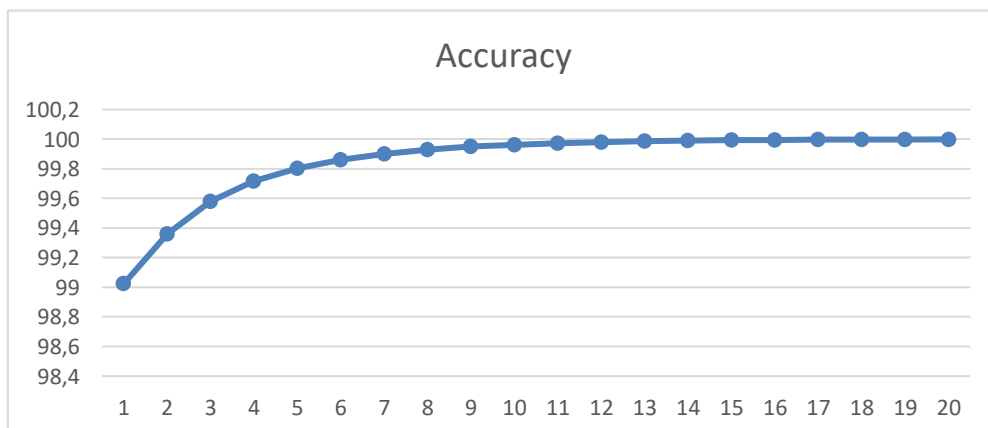| kernel dan parameter		w value			b value		
		w_4	w_5	w_6	b_4	b_5	b_6
RBF	$\alpha = 0.5$	−59,5390 435,8195 98,1523 −59,0268	6,5393 401,7964 81,7738 3,9947	−24,7938 445,8366 65,8424 −24,0753	−0.4588	−1.0626	−0.7571

Accuracy

Figure 1: Accuracy level graph of the One Against All method for each kernel in dataset
Based on Table 1 and Figure 1, the use of RBF kernel types and using the parameter σ = 0.5 has the highest accuracy on dataset has the highest accuracy (99.99832). Accuracy for classification with the number of classes : 6 classes.
.

5 Conclusions

From the research that has been done, it can be concluded that this research uses multiclass ensemble ensemble least-squares support vector machine method. The image of the face is extracted using Linear Discriminant Analysis (LDA) algorithm to take the value of Projection Matrix and Weight Matrix. Then the classification process using multiclass ensemble least-squares support vector machine method, to get the model of the image training and testing results of the image testing. The final result of this research by using RBF kernel with accuracy level that is (99.99832).

Acknowledgements

References

[1]   P.S. Hiremath dan M. Hiremath, "Linear Discriminant Analysis for 3D Face Recognition Using Radon Transform," in Swamy P., Guru D. (eds) Multimedia Processing, Communication and Computing Applications, Lecture Notes in Electrical Engineering, vol 213. Springer, New Delhi, (2013).

[2]   M. . Li, dan B. Yuan, "2D -LDA: A statistical linear discriminant analysis for image matrix," Pattern Recognition Letters 26.5 (2005): 527-532.

[3]   H. Ebrahimpour, dan A. Kouzani, "Face Recognition Using Bagging KNN," International Conference on Signal Processing and Communication Systems (ICSPCS'2007) Australia, Gold Coast, (2007).

[4]   Yubo Wang, Haizhou Ai, Bo Wu, Chang Huang, "Real Time Facial Expression Recognition With Adaboost", Departement of Computer Science and Technology, Tsinghua University, China (2004).

[5]   Reda Shbib, Shikun Zhou,"Facial Expression Analysis using Active Shape Model", School of Engineering, University of Portsmouth, United Kingdom .

[6]   Turban, E., dkk. 2005. Decision Support Systems and Intelligent Systems. Yogyakarta: Andi Offset.

[7]   Sembiring, K. 2007. Tutorial SVM Bahasa Indonesia. Bandung: Institut Teknologi Bandung.

[8] Gestel, T., Suykens, J., Lanckriet, G., Lambrechts, A., Moor, B., & Vandewalle, J. 2002. Multiclass LS-SVMs: Moderated Outputs and Coding-Decoding Schemes. Neural Processing Letters, 45-58.

[9] http://www.kasrl.org/jaffe.html

[10] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

[11] Jafar, N., Thamrin, S.A., Lawi, A., 2016, Multiclass Classification using Least Squares Support Vector Machine, Proceeding of International Conference on Computational Intelligence and Cybernetics.

[12] Kurniawan, D dan Supriyanto, C. 2013. Optimasi Algoritma Support Vector Machine (SVM) Menggunakan AdaBoost untuk Penilaian Risiko Kredit. Jurnal Teknologi Informasi.

[13] Marliani R, R., Lawi, A., & Thamrin, S. A. 2016. Algoritma Adaboost untuk Optimasi Ensemble Least Squares Support Vector Machine. Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

[14] Nurkamila, J. Lawi, A., & Thamrin, S. A. 2016. Least Squares Support Vector Machine Untuk Klasifikasi Data Multikelas. Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

[15] Vapnik, V., & Cortes, C. 1995. Support-Vector Networks. Machine Learning, 273-297.

[16] Pratt, William K. 2007, Digital image processing, PIKS Scientific inside, John Wiley, 4th Edition.

[17] Xu, Y., Lv, X., Wang, Z., & Wang, L. 2014. A Weighted Least Squares Twin Support Vector Machine. Journal of Information Science and Engineering, 1773-1787.

[18] Zhou, L., Lai, K. K., dan Yu, L. 2010. Least Squares Support Vector Machines Ensemble Models for Credit Scoring. Expert Systems with Applications, 127-133.