

Joint Spatial Geometric and Max-margin Classifier Constraints for Facial Expression Recognition Using Nonnegative Matrix Factorization

^aPhan Trong-Thanh, ^bDoan Van Thang

^a*Vietnam Korea University of Information and Communication Technology, The University of Da Nang, Da Nang, Viet Nam*

ptthanh@vku.udn.vn

^b*Industrial University Of Ho Chi Minh City*

vanthangdn@gmail.com

Abstract

Based on the constrained non-negative matrix factor algorithm, the article presents a new approach to facial recognition recognition. Our proposed method incorporated two tasks in an automatic expression analysis system: facial feature extraction and classification into expressions. To obtain local and geometric structure information in the data as much as possible, we amalgamate max-margin relegation into the constrained NMF optimization, resulting in a multiplicative updating algorithm is additionally proposed for solving optimization quandary. Experimental results on JAFFE dataset demonstrate that the effectiveness of the proposed method with improved performances over the conventional dimension reduction methods.

Keywords

facial expressions;
classification;
nonnegative matrix factorization;
graph regularization;
spatial constraints

1. Introduction

Facial expression recognition (FER) is increased the attention from psychologist's anthropologists, and computer scientists [1, 2, 3]. The computer researchers attempt to create complicated human-computer interfaces which are able to recognize automatically and classify human expressions or emotions. Fasel et al. [1] has defined facial features that will deform over time and the compression of facial muscles will give skin texture. Research has also shown that the general changes of muscle activity occur quickly and usually only a few seconds. The study also provides for the fact that emotions are the only source of facial expressions in addition to other things such as verbal and nonverbal correspondence, or physiological activities.

Usually facial expressions and emotions are different (and the terms are commonly not correctly exchanged), for PC vision group, the expression " facial expression recognition" frequently refers to the characterization of facial components in one of the six alleged essential feelings: sadness, disgust, happiness, surprise, fear and anger, as presented by Ekman in 1971 [2]. The endeavor of an elucidation depends on the suspicion which the appearances of feelings are widespread crosswise over people and also human ethnics and societies.



Figure 1. Six universal emotions

An automatic expression analysis system has two tasks that are compulsory [4]: facial feature extraction and relegation into expressions. There are generally two kinds of methods in facial feature extraction: appearance-predicated and geometric feature-predicated methods. After focalizing the face, the exhibited countenance has to be extracted as much information as possible. In visage apperception, most automatic expression analysis systems endeavor to apperceive a diminutive set of prototypic expressions (i.e. fear, anger, joy, surprise, sadness, and disgust).

Table 1. Universal emotion identification

Universal emotion identification		
Emotion	Definition	Motion of facial part
Anger	Anger shows the emotion which is the most dangerous, it could be very harmful, this emotion should be tried to avoid by human.	Pulling down eyebrows, ocular perceiver opening, teeth shutting and lips tightened, upper and lower lids pulling up.
Fear	Fear is also the emotion of danger. It may lead to physical or psychological harms.	Outer eyebrow going down, inner eyebrow going up, mouth opening and jaw dropping.
Happiness	Happiness is most desired expression by human.	Eyes opening, mouth edging up, mouth opening, lip corner pulling up, cheeks raising up and wrinkles around eyes.
Sadness	Sadness is opposite emotion of Happiness.	External eyebrow going down, internal corner of eyebrows raising up, mouth edge going down, eye shutting, lip corner pulling down.
Surprise	This emotion comes when unexpected things happens.	Eyebrows going up, eye and mouth opening, jaw dropping down.
Disgust	Disgust is a feeling of dislike such as taste, smell, sound or tough.	Lip corner depressor, nose wrinkle, lower lip depressor, Eyebrows pulling down

2. Facial expression recognition

2.1. Feature Extraction Using Nonnegative Matrix Factorization

In this section, we describe the most important step in a facial expression recognition system which is feature extraction step that it can be analyzed in terms of facial action occurrence after the face has been located in the image or video frames. Over the last several decenniums, massive endeavor has been made and exceptional. Immense endeavor has been made over the past several years and consequential results have been successful finished in FER. A main step in FER is to engender or abstract features of verbalization from the pristine facial images. A main step in FER is to build or abstract features of verbalization from the pristine images of the ocular perceivers. Several wide-ranging methods of extraction of features such as the key element analysis (PCA) [5], Eigen-face [6], Singular value decomposition (SVD) [7] Non-negative matrix factorization (NMF) [8].

In NMF algorithm, it was described as follow:a non-negative $m \times n$ matrix has been given: $X=(x_0,x_1,\dots,x_{(m-1)}) \in \mathbb{R}^{(m \times n)}$ is exactly the facial data will be analyzed, it's the need of finding non-negative matrix factor $U (m \times k)$ and matrix factor $V=(k \times n)$ such that $X \approx UV$ where k is smaller than m and n .

In the original X matrix a column vector can be considered as the sum of the weights of all the vectors in the left matrix U , but the opposite the elements of the corresponding column vector in the right matrix V are weight coefficients. The un-negativity constraints of U and V consistent with the intuitive notion of combining parts to form a whole, which is how NMF learns a part-based representation.

$$\min_{U, V} \frac{1}{2} \|X - UV\|_F^2, \text{ s.t. } U \geq 0, V \geq 0 \quad (1)$$

where $\|\cdot\|_F$ is Frobenius norm of a matrix, and the product UV^T is the non-negative matrix factorization approximation of X of rank at most k . The non-negativity constraints on U and V enables only additive (non-subtractive) combine of parts to build entire data.

In the utility for consciousness of facial expression, the grey values on each facial picture are non-negative and saved as a type of matrix $X=[X_1, X_2, \dots, X_n]$, the place X_j is a column vector of m dimensions, consisting of the non-negative gray facial picture. The X matrix can be divided into the product of a non-negative U matrix representing simple NMF images and a non-negative weight coefficient matrix V , NMF decomposition renders expression pix reconstructed in a nonsubtractive manner and a great deal like the approach of forming harmony from sections.

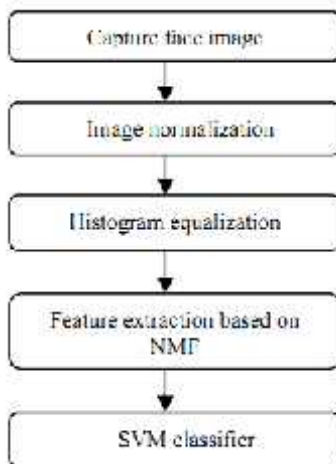


Figure 2. Basic structure of facial expression analysis system [4]

2.2. Classification by SVM Classifier

The last task of the FER system is classification based on machine learning theory. The input to the classification process is a combination of features retrieved from the face area in the earlier stage. The earlier stage which will contain a set of features formed to describe facial expressions. Classification is in need of supervised training, so the training set has to be made up of labeled data.

Support Vector Machines (SVM) are data classification techniques based on optimization theory and statistical theory. In the SVM technique the initial input space will be mapped into the feature space and in this feature space the optimal dividing hyperplane is determined. To divide multiple class, the original SVM technique will divide the data space into two parts and the process is repeated many times. Thus, the multi-class problem using the SVM method can completely perform like the two-class problem. Assuming the problem needs to be classified with k classes ($k > 2$), the "one-against-one" strategy will proceed to implement $k(k-1)/2$ binary classifications using the SVM method. Each class will conduct separation with the remaining $k-1$ class to determine the $k-1$ separation function based on the two-class division problem by SVM method.

2.3. Adaptive Feature Extraction and Classification Method

There are few existed works that use constrains the aim of increasing the the discriminatory ability for extracted features. Several variants of NMF with discriminant constraints imposed were proposed in [9, 10]. Kumar et al [11] introduced an adaptive feature extraction and classification method which propose a method to obtain the baseline matrix by constructing a soft max-margin constraints calculation for the objective function of NMF that maximized the classification margin using the features that are extracted using those bases. Inspired by this, they aim at finding a set of basis vectors that maximizes the margin of an SVM classifier.

Let $\{x_i, y_i\}_{i=1}^L$ denote a set of data vectors and their corresponding labels, where $x_i \in R^m, y_i \in \{-1, 1\}$. To extract features based on criteria max-margin classifier we use base matrix U . To perform this problem, we impose constraints on the features vectors derived from the matrix U . Features extracted from projections

of example data x on base vectors are stored in $U(U^T X)$ and optimizations are calculated as follows:

$$\min_{U, V, W, b, \varepsilon_i} \lambda \|X - UV\|_F^2 + \frac{1}{2} w^T w + C \sum_{i=1}^L \varepsilon_i \quad (2)$$

$$s.t. y_i(w^T U^T x_i + b) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0, 1 \leq i \leq L, V \geq 0$$

where $w = (w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_L)$ is variable vector, λ is a scalar that controls the relative significance for the NMF cost and C a scalar that controls the overall significance of the punishment forced for the preparation models that are either excessively near the isolating hyper-plane or misclassified.

3. Methodology

We propose a unity of objective function for the model which archives the upper objectives by combining the benefit of max-margin classifiers and NMF constraints together, through adding the pixel dispersion penalty and manifold regularization into the objective function. Following, we drive a multiplicative update rules using optimized gradient method and describe how the systems use this algorithm to perform the classification task we expect it to do.

3.1. Max-margin Nonnegative Matrix Factorization via Spatial Constraints and Graph Regularization

The unified objective function is constructed by jointing the data reconstruction objective function:

$$\min_{U, V, w, b, \varepsilon} \|X - UV\|_F^2 + \lambda_1 \left(\frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^L \varepsilon_i \right) + \lambda_2 Tr(VLV^T) + \lambda_3 Tr(UEU^T) \quad (3)$$

$$\text{s.t. } \mathbf{U} \geq 0, \mathbf{V} \geq 0, \varepsilon_i \geq 0, y_i(w^T v_i + b) \geq 1 - \varepsilon_i, i = 1..n$$

All variables are divided into three terms: the coefficient matrix (V), the basis matrix (U) and variables about max-margin projection (w, b,). Where $\|\cdot\|_F$ is Frobenius norm of a matrix, and the product \mathbf{UV} is the non-negative matrix factorization approximation of \mathbf{X} of rank at most k ; $\varepsilon = (\varepsilon_1, \dots, \varepsilon_i, \dots, \varepsilon_n)$ is the slack variable vector, λ_1 and C are scalars; the regularization parameter $\lambda_2 \geq 0$ controls the smoothness of the new representation; $\lambda_3 \geq 0, L \ll d$ and c_0 is a simple positive constant bound parameter; L is called graph Laplacian, E is called the dispersion kernel matrix.

Multiplicative Update Rules.

Update the Projection Vector and Slack Variables.

When the coefficient matrix and the basis matrix are fixed, MMNMF_MR optimization problem changes into the standard binary soft-margin SVM classification.

$$\min_{w,b,\varepsilon} \lambda_1 \left(\frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^n \varepsilon_i \right), \text{s.t. } \varepsilon_i \geq 0, y_i(w^T v_i + b) \geq 1 - \varepsilon_i, i = 1..n \quad (4)$$

The hyper-plane parameters w , b and slack variable vector ε are received by the use of SVM classifier.

Update the Coefficient Matrix.

When other variables are fixed, optimizing the coefficient matrix is a quadratic program solution:

$$\min_{\mathbf{V}} \|\mathbf{X} - \mathbf{UV}\|_F^2 + \lambda_2 \text{Tr}(\mathbf{V}\mathbf{L}\mathbf{V}^T) \quad (5)$$

$$\text{s.t. } \mathbf{V} \geq 0, y_i(w^T v_i + b) \geq 1 - \varepsilon_i, i = 1..n$$

The Lagrangian of above objective function is

$$L(\mathbf{V}, \alpha, \beta) = \text{Tr}(\mathbf{X} - \mathbf{UV})(\mathbf{X} - \mathbf{UV})^T + \lambda_2 \text{Tr}(\mathbf{V}\mathbf{L}\mathbf{V}^T) - \alpha^T \mathbf{V} - \sum_{i=1}^n \beta_i [y_i(w^T v_i + b) - 1 + \varepsilon_i]$$

where α, β are Lagrangian multipliers, specifically α is Lagrangian multipliers vector. Under the Karush-Kuhn-Tucker (KKT) conditions, we get

$$\begin{cases} 2\mathbf{U}^T \mathbf{UV} - 2\mathbf{U}^T \mathbf{X} + 2\lambda_2 \mathbf{V}\mathbf{L}^T - \alpha - \beta y w = 0 \\ \mathbf{1}^T \mathbf{V} = 0 \\ y(w^T \mathbf{V} - b) - 1 + \varepsilon = 0 \end{cases}$$

Transform the equation into a matrix for

$$\begin{pmatrix} 2\mathbf{U}^T \mathbf{U} + 2\lambda_2 \mathbf{L}^T & -\mathbf{1}^T & -y w \\ \mathbf{1}^T & 0 & 0 \\ y w^T & 0 & 0 \end{pmatrix} \times \begin{pmatrix} \mathbf{V} \\ \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 2\mathbf{U}^T \mathbf{X} \\ 0 \\ y b + 1 - \varepsilon \end{pmatrix}$$

where $\mathbf{1}$ is a unit vector whose size is the same as \mathbf{v} , $\mathbf{0}$ is the zero vector. We can derive \mathbf{v} by solving this equation.

Update the Basis Matrix.

When other variables are fixed, the model is transformed to a non-negative matrix factorization:

$$\mathcal{O}_2 = \min_{\mathbf{U}, \mathbf{V}} \|\mathbf{X} - \mathbf{UV}\|_F^2 + \lambda_2 \text{Tr}(\mathbf{UEU}^T), \text{ s.t. } \mathbf{U} \geq \mathbf{0} \quad (6)$$

Because of the non-negative constraints, we use gradient descent methods to solve this problem. The gradient of equation (6) is

$$\mathbf{V} = 2\mathbf{UVV}^T - 2\mathbf{X}^T\mathbf{V}^T + 2\lambda_2\mathbf{EU}$$

Classification

To obtain the feature vectors during the implementation then input test vectors x_{test} will be projected onto the base matrix \mathbf{U} , $f_{test} = \mathbf{U}^T x_{test}$. The feature vector is used by the max-margin classifier which predicts the class $y_{test} = \text{sign}(\mathbf{U}^T f_{test} + b)$ where w, b, \mathbf{U} are calculated during training.

Algorithm for MNMF_SGR

```

Input: Matrix X, rank k, maxIter; positive constants  $\lambda_1, \lambda_2, \lambda_3$ 
Output: U, V, w, b

Begin

Initial the basis matrix U0 and the coefficient V0, let t=0
Let s=1, U=Us, V=Vs
Repeat
Fix U and Vs to find ws+1, bs+1 via equation (4)
Fix V, ws and bs+1 to find Us+1 via equation (5)
s=s+1
Until reaches the maximal iteration number;
Let t=t+1, Vt=Vs, wt=ws, bt=bs
Learning the new basis matrix Ut via minimizing equation (3.1)
End

```

Figure. 1. Algorithm for Max-margin Nonnegative Matrix Factorization via Spatial Constraint and Graph Regularization

4. Experiments

In this subsection, proposed MNMF_SGR method would be experimented and compared against several popular subspace learning algorithms, specifically the unsupervised methods (NMF [8], Spatial NMF [9] and Graph NMF [10]). We also compared with the supervised algorithm Semi-NMF [12] and Max-margin NMF [11].

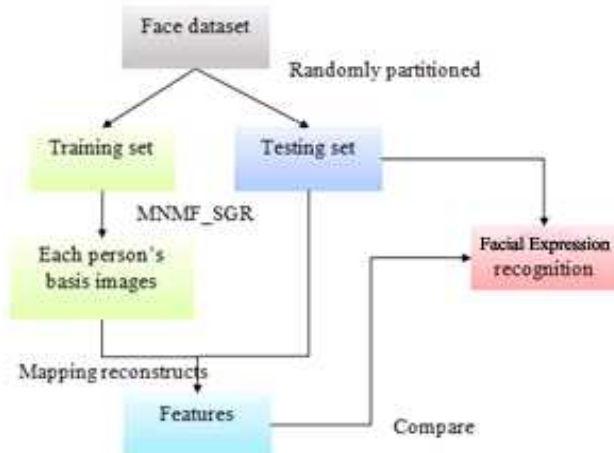


Figure. 2. The flow chart of MNMF_SGR based image reconstruction for facial expression recognition

4.1. Datasets

We using Japanese Female Facial Expression (JAFPE) database [13] to experiments. The dataset have been collected and staging by Michael Lyons, Miyuki Kamachi, and Jiro Gyoba.

4.2. Preprocessing

Due to the background is more immensely colossal than face image; firstly, the Viola-Jones algorithm will be used to find faces. Based on the partitions placed set on already detected frontal faces, we using cascade object detector to detection ocular perceivers, nasal perceiver and mouth. The reality shows that the use of Viola-Jones algorithm as a preprocessing step and want to achieve a good classification then this step needs to be done. Each original image from both databases is cropped and down-sampled in a such way that the final image size is 16×16 pixels.

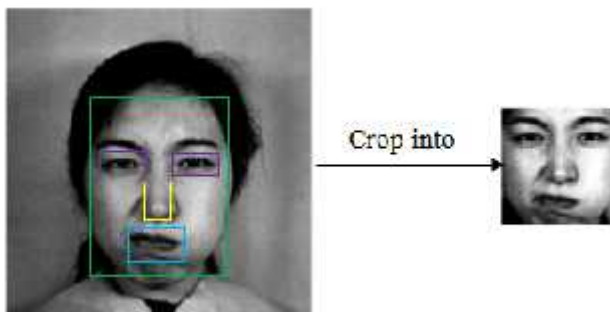


Figure. 3. Face and facial parts detection

All algorithms were initialized with 20 random U and V matrices, each of them was trained for 20 iterations and the one with the minimum objective function value was further trained for 1000 iterations.

4.3. Parameter Settings

For training and testing splits, we repeated the following procedure for ten times. Each time we randomly selected two-thirds of number of image per individual and labeled them. All the other images were unlabeled and used as the testing set.

In MNMF_SGR, λ_1 was tested for the following values {0.01, 1, 10} and λ_2 was tested for {1, 100} and λ_3 was tested for $\{10^{-5}, 10^{-4}, \dots, 10^2\}$. Firstly, the dimensionality reduction process with NMF, SpaNMF, GNMF and Semi-NMF algorithms, the trained coefficient matrix is ready to be used for classifying a testing face image. Then we use SVM algorithm for the classifiers in the face recognition.

With MNMF and MNMF_SGR, after training process we compute the feature vector from the input test vector that the basis matrix projected onto. After that, this feature vector is used in predicting class of face recognition. All algorithms were initialized with 20 random U and V matrices, each of them was trained for 20 iterations and the one with the minimum objective function value was further trained for 1000 iterations.

4.4. Classification Results.

The results of facial expression recognition for JAFFE dataset shown in Figure 6. Semi-supervised algorithms outperform all un-supervised ones. MNMF_SGR has highest accuracy, and then followed by MNMF_FA, MNMF, SpaNMF, GNMF, SemiNMF and standard NMF. MNMF_SGR outperforms NMF by 21.87%. The highest classification accuracy of 86.94% is achieved with $k=30$.

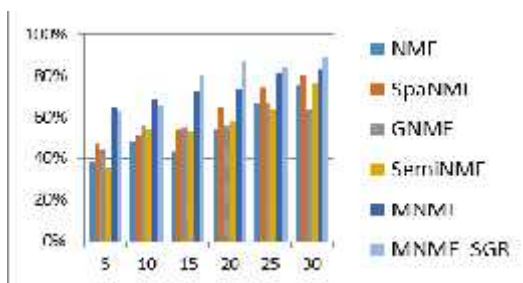


Figure. 4. JAFFE dataset facial expression recognition average accuracies (%) of different algorithms of five iterations

The disarray lattice of outward appearance acknowledgment appeared in Table II utilizing proposed technique with 30 number of highlight vectors ($k=30$). A portion of the pitiful and satisfaction outward appearance are mistaken for one another. The distinction of satisfaction and dismal fizzled in light of the fact that these articulations had a comparable movement of mouth.

Table 1. The 7-class facial about confusion matrix expression recognition using proposed method MNMF_SGR on JAFFE

An	Di	Fe	Ha	Sa	Su	Ne
0.6268	0.1813	0.7285	0	0.0812	0.0322	0.0333
0.1234	0.4510	0.1009	0.0426	0.1268	0	0.0224
0	0.1191	0.4825	0.0918	0.0810	0	0.0421
0	0.1289	0.0199	0.7490	0.0211	0.0635	0.1289
0.1713	0.1270	0.1197	0.0423	0.5343	0.0524	0.1756
0	0	0	0.0203	0.0417	0.7417	0.0722
0	0.0000	0.0198	0.0486	0.1222	0.0711	0.5221

5. Conclusion

In this paper, we have presented the constrained NMF approach for problem the facial expression recognition. The proposed MNMF_SGR performs well in facial expression recognition task and its effectiveness has been proven in our model. To summarize, With many constraints allows us to build models effectively and specifically on high dimensional, sparse and noisy datasets. For future work more sophisticated and efficient way to tune kernel functions will be explored. We will also apply the proposed method to problems in other fields, such as bioinformatics and computer vision. Studying the convergence rate for MNMF_SGR and increasing the efficiency, they should be all in consideration.

References

- [1]. B. Fasel, J. Luettin, "Automatic facial expression analysis: a survey", *Pattern Recognition*, 36 (1) (2003), pp. 259-275.
- [2]. Ekman. P, Friesen W.V, "Constants across cultures in the face and emotion", *Journal of Personality and Social Psychology*, 17: 124–129, 1971.
- [3]. L. Mai, "Joint Support Vector Machine with Constrained Nonnegative Matrix Factorization and Its Applications", Master Thesis, National Central University, July 2017.
- [4]. Ying Zilu, Zhang Guoyi, "Facial Expression Recognition Based on NMF and SVM", *IEEE: International Forum on Information Technology and Applications*, pp. 612-615, 2009.
- [5]. H. Moon and P. J. Phillips, "Computational and performance aspects of PCAbased face-recognition algorithms.," *Perception*, vol. 30, no. 3. pp. 303–21, Jan- 2001.
- [6]. H. Wechsler, "Enhanced Fisher linear discriminant models for face recognition," *Proceedings. Fourteenth Int. Conf. Pattern Recognit.* (Cat. No.98EX170), vol. 2, pp. 1368–1372.
- [7]. L.-F. Chen, H.-Y. M. Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognit.*, vol. 33, no. 10, pp. 1713–1726, Oct. 2000.
- [8]. D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization", *Advances in Neural Information Processing System*, 2000.
- [9]. W. S. Zheng, J. Lai, S. Liao, R. He, "Extracting non-negative basis images using pixel dispersion penalty", *Pattern Recognition*, pp. 2912-2926, 2012.
- [10]. D. Cai, X. He, J. Han and T. Huang, "Graph regularized nonnegative matrix factorization for data representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33(8), pp. 1548–1560, 2011.
- [11]. B. G. Kumar, I. Kotsia and I. Patras, "Max-margin nonnegative matrix factorization", *Image Vision Computing*, pp. 279–291, 2012.
- [12]. C. Ding, T. Li, and M. Jordan, "Convex and semi-nonnegative matrix factorizations for clustering and low-dimension representation", *Technical Report LBNL-60428*, Lawrence Berkeley National Laboratory, University of California, Berkeley, 2006.
- [13]. Matthew N Dailey, Carrie Joyce, Michael J Lyons, Miyuki Kamachi, Hanae Ishi, Jiro Gyoba, & Garrison W Cottrell, "Evidence and a computational explanation of cultural differences in facial expression recognition". *Emotion*, Vol 10(6), Dec 2010, 874-893.

Author's Biography



Phan Trong Thanh holds a Master of Information System and Management in 2016 from Lughwa University of Science and Technology – Taiwan He works as a lecturer in Vietnam Korea University of Information and Communication Technology, Danang, Vietnam. His research interests are Information System and Networking.



Doan Van Thang obtained his doctorate in Mathematical Foundation for Computers and Computing Systems in 2014 from the Institute Of Information Technology, Vietnamese Academy Of And Training Science And Technology Viet Nam. He has published over 30 research papers. His main research interests include the fuzzy database, hedge algebras and detection, prevention, avoidance of cloud computing and distributed systems. He serves as a technical committee program member, track chair, session chair and reviewer of many international conferences and journals.

How to Cite

Thanh, Phan Trong, and Thang, Doan Van(2019). Joint Spatial Geometric and Max-margin Classifier Constraints for Facial Expression Recognition Using Nonnegative Matrix Factorization. *International Journal of Machine Learning and Networked Collaborative Engineering*, 3(03) pp 125-134.

doi : <https://doi.org/10.30991/IJMLNCE.2019v03i03.001>
