International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.12, (2016), pp.103-112
http://dx.doi.org/10.14257/ijsip.2016.9.12.11

Multi Feature Information Fusion Target Image Recognition Based on Hyper Plane Fusion of Learning Prototype

Zhou Gaiyun[1], Zhang Guoping[1], Chang Cunhong[2] and Ma Li[1]

[1]Software Academy, Pingdingshan University, Pingdingshan, Henan, 467000, China;

[2]School of computer and information technology, Shangqiu Normal University,Shangqiu, 476000, China

zhougypdu@126.com

Abstract

In view of the greater changes of posture, illumination, expression and scene in reality environment have a strong impact on wild face recognition algorithm to identify performance problem, and puts forward a kind of linear discriminant analysis side information (SILD) algorithm on hyperplane fusion of learning prototype. First of all, using support vector machine (SVM) to weak tag of data-concentrated sample is expressed as the middle-level characteristics of prototype hyperplane, using a learning combination coefficient to select sparse support vector set from untagged conventional data set; then, under the constraints of the combination sparse coefficient of SVM model, by using Fisher linear discriminant criterion to maximize discriminant ability of untagged data set, and using the iterative optimization algorithm to solve the objective function; in the end, using SILD for feature extraction, cosine similarity measure to complete the final face recognition. In two general face data sets of wild face recognition (LFW) and YouTube, it makes comparison of PHL+SILD method and low-level features + SILD method on some characteristics, such as strength, LBP, Gabor feature and Block Gabor feature, average accuracy, area under the curve (AUC) and entire error rate (EER). The validity and reliability of the proposed algorithm is verified by the experiments.

Keywords: *Wild face recognition; Prototype hyperplane learning; Middle-level character representation; Support vector machine; Linear discriminant analysis information*

1. Introduction

Over the past two decades, under the non-limited conditions face recognition has made significant progress, and obtained good experimental results on data sets such as FERET [1], CMU-PIE [2] and so on. Recently, wild face recognition has attracted a wide range of research interests of scholars, face image is usually collected under non-limited conditions, because there has larger changes on posture, illumination, expression and scenario, wild face recognition has become a more challenging task [3]. Inspired by literature [14], this paper proposes a linear discriminant side information algorithm of hyper plane fusion of learning prototype, by using attached untagged conventional data set to build Support Vector Machine (SVM) model (that is prototype hyperplane), thus obtain the middle-level characteristics expression of prototype hyperplane, and proposes an iterative optimization algorithm to solve the objective function, the obtained non-zero combination coefficient automatically determine each prototype hyperplane used for wild face recognition, using SILD for dimension reduction, cosine similarity metric to complete face recognition. Experimental results verify the effectiveness and superiority of the proposed algorithm.

2. Proposed Algorithm

2.1. Learning Algorithm of Prototype Hyper Plane

1) Define Equation(1)tracking problem as regression problem：given weak-tagged different subject M_0 corresponding sample $\left\{(z_i^0, \hat{z}_i^0)\Big|_{i=1}^{M_0}\right\}$ in dataset, define two data matrixes ： $D=[(z_1^0-\hat{z}_1^0),(z_2^0-\hat{z}_2^0),...,(z_{M_0}^0-\hat{z}_{M_0}^0)] \in \mathbb{R}^{D\times M_0}$ and $H_b = D^T X \in \mathbb{R}^{M_0\times N}$, in order to define another matrix $R_w \in \mathbb{R}^{N\times N}$, first conduct singular value decomposition SVD to S_w , that is $S_w = R_w^T R_w$. By introducing intermediate variable $A=[a_1,a_2,...,a_c] \in \mathbb{R}^{N\times C}$ to redefine tracking problem as regression problem：

$$[A^*, B^*] = \arg\min_{A,B} \sum_{i=1}^{C} \left\| H_b R_w^{-1} a_i - H_b \beta_i \right\|^2 + \sum_{i=1}^{C} \lambda \beta_i^T S_w \beta_i \\ \text{s.t.} A^T A = I_{C\times C}, \|\beta_i\|_1 < t, i=1,...,C. \tag{1}$$

2) Optimize the regression problem in Equation(1): using the iterative optimization method for iterative optimization of A and B , A is given, solving the following problem to obtain B ：

$$B^* = \arg \min_{\beta_1,\beta_2,...,\beta_C} \sum_{i=1}^{C} (\left\| H_b R_w^{-1} a_i - H_b \beta_i \right\|^2 + \lambda \beta_i^T S_w \beta_i) \\ \text{s.t.} \|\beta_i\|_1 < t, i=1,...,C. \tag{2}$$

$\beta_1, \beta_2, ..., \beta_C$ are independent in Equation (2), by optimizing the following problem to solve each β_i separately

$$\begin{aligned} \beta_i^* &= \arg\min_{\beta_i} \|s_i - H_b \beta_i\|^2 + \lambda \beta_i^T S_w \beta_i \\ &= \arg\min_{\beta_i} \|\tilde{s}_i - \tilde{W}\beta_i\|^2, s.t. \|\beta_i\|_1 < t, \end{aligned} \tag{3}$$

$$s_i = H_b R_w^{-1} a_i,\ \tilde{s}_i = [s_i^T, 0_N^T]^T,\ \tilde{W} = [H_b^T, \sqrt{\lambda} R_w^T]^T$$

Using minimum angle regression to solve the optimal β_i .

Given B, ignore the constraint condition on β_i , and directly calculate A by solving the following problem：

$$\begin{aligned} A^* &= \arg \min_{a_1,a_2,...,a_c} \sum\nolimits_{i=1}^{C} \left\| H_b R_w^{-1} a_i - H_b \beta_i \right\|^2, \\ &= \arg\min_{A} \left\| H_b R_w^{-1} A - H_b B \right\|^2, \text{s.t.} A^T A = I_{C\times C} \end{aligned} \tag{4}$$

Using SVD can obtain the optimal A, that is

$$R_w^{-T}(H_b^T H_b)B = U\Sigma V^T, and\ A^* = \tilde{U}V^T \tag{5}$$

Where, $\tilde{U} = [u_1, u_2, ..., u_N]$ includes front C dominant feature vectors of matrix $U=[u_1,u_2,...,u_N]$. In this study, iterative solving Equation(2) and (4), until the absolute error of B after twice continuous iterations is less than the preset threshold, untagged dataset $X = [x_1, x_2, ..., x_N] \in \mathbb{R}^{D \times N}$, weak-tagged datasets of including M_1 corresponding same subject sample $\left\{(z_i^0, \hat{z}_i^0)\Big|_{i=1}^{M_0}\right\}$ and M_0 corresponding different subject sample $\left\{(z_i^1, \hat{z}_i^1)\Big|_{i=1}^{M_1}\right\}$, relevant data matrix is defined as : $D = [(z_1^0 - \hat{z}_1^0), (z_2^0 - \hat{z}_2^0), ..., (z_{M_0}^0 - \hat{z}_{M_0}^0)] \in \mathbb{R}^{D \times M_0}$. Detailed prototype hyperplane learning (PHL) algorithm as shown below.

Step1: Initialize $A \in \mathbb{R}^{N \times C} + B \in \mathbb{R}^{N \times C} = 1$.

Step2: Using Equation (1) to calculate S_b and S_w.

Step3: By performing SVD of S_w, that is $S_w = R_w^T R_w$ to calculate $H_b = D^T X \in \mathbb{R}^{M_0 \times N}$ and $R_w \in \mathbb{R}^{N \times N}$.

Step4: $\beta_i^* = \arg\min_{\beta_i} \left\|\tilde{s}_i - \tilde{W}\beta_i\right\|^2, s.t. \left\|\beta_i\right\|_1 < t, i = 1, ..., C.$ 。 Given A, to solve independent Lasso problem in Equation (10) to obtain C, using least angle regression: $\beta_i^* = \arg\min_{\beta_i} \left\|\tilde{s}_i - \tilde{W}\beta_i\right\|^2, s.t. \left\|\beta_i\right\|_1 < t, i = 1, ..., C.$ 。

Step5: Given B, establish SVD, that is $R_w^{-T}(H_b^T H_b)B = U\Sigma V^T$, using $A^* = \tilde{U}V^T$ to solve A, where, $\tilde{U} = [u_1, u_2, ..., u_N]$ including front C domain feature values of matrix $U = [u_1, u_2, ..., u_N]$.

Step6: Repeat Step1 to Step5 until the changes of B in twice continuous iterations is less than ε (in this study $\varepsilon = 0.001$)。

2.2. Dimensionality Reduction and Recognition

With learning prototype hyperplane, using Equation (3), each sample can be represented as its middle-level decision value eature, in order to further reduce the feature dimension and improve performance, using the proposed recently SLID [14] for dimension reduction, use only weak-tagged training data can learn discriminant projection matrix, when the class label information of each sample is known, SILD and Fisher linear discriminant analysis are equivalent. In SILD training process, only use the samples of the same object to define the inner-class scattering matrix, using the sample pair of different objects to define inter-class scattering matrix, use the general feature value decomposition method to determine the projection matrix for dimension reduction. Testing process (as shown in Figure 1 (b)), for each pair of test data z and $\hat{z}$, use learning prototype hyperplane to respectively produce corresponding middle-level feature to express $f(z)$ and $f(\hat{z})$, and then use SILD training process learning of projection matrix $f(z)$ and $f(\hat{z})$ mapped to a space, and finally, before the execution of face recognition, using cosine function to calculate to similarity of test sample pair, the entire process is shown in Figure 1.

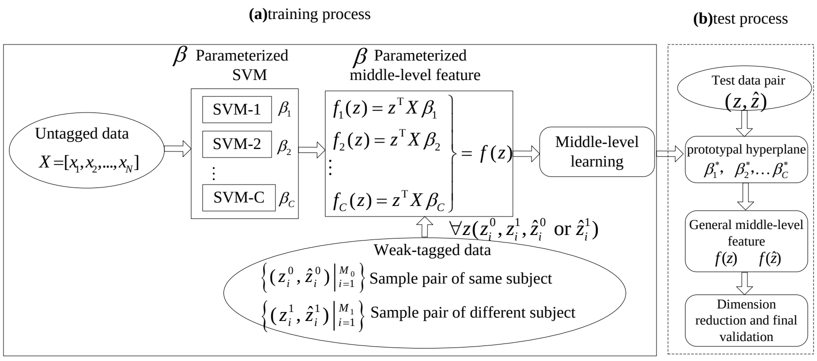

Figure 1. Proposed Algorithm Flow Chart: (1) Training Process (b) Test Process

3. Experiment

On two unconstrained conditions data sets: wild flagged face image (LFW) [4] and YouTube face dataset [15], this section makes comparison on the prototype hyperplane learning algorithm (PHL) proposed in this paper and several other relatively advanced algorithms.

3.1. Dataset

LFW database is a large database, is composed of 13,233 face images of 5749 people, standard assessment protocol focus on two points: the model selection and performance evaluation. In this experiment, the central area of each face image is cut to 80×150 pixel by removing the background.

YouTube face database is a large video data set without constraints, contains 3425 videos of 1595 objects, each object has average 2.15 video, each video clip length is about 181 frames at 24fps.

Using image-limited training model on two data sets in experiment, namely, only know some sample belongs to the same object or a different object, and have no idea about the class label of each sample, in order to build untagged conventional data set $\mathcal{X}$, randomly select 3000 untagged samples from LFW dataset as conventional data set, select image from YouTube face dataset as the training dataset, it is important to note that there is no repeat image between conventional data set and test set. Every cycle of experiment accuracy is defined as the correct classified sample pairs number divided by the total number of test sample pairs, the standard deviation is defined as $\frac{\hat{\sigma}}{\sqrt{10}}$, where $\hat{\sigma}$ is the standard deviation.

3.2. Parameter Discussion

Take YouTube face dataset as example, it explores the influence of the number of different prototype hyperplane C and sparse parameter t on the performance of the algorithm in this paper, including the recognition accuracy and time-consuming of the whole algorithm, C respectively takes 100, 200, 400 and 200, and t respectively takes 0.1,0.2,…,0.8. Experiment uses MATLAB7.0 and implements on a personal computer, the computer configuration is: Windows XP operating system, Centrino core 2 processor, 3.10 GHz basic frequency, 8GB RAM, the influence of parameter C and t on recognition accuracy of proposed algorithm as shown in Figure 2, the influence of parameter C and t on time-consuming of proposed algorithm as shown in Figure 3.

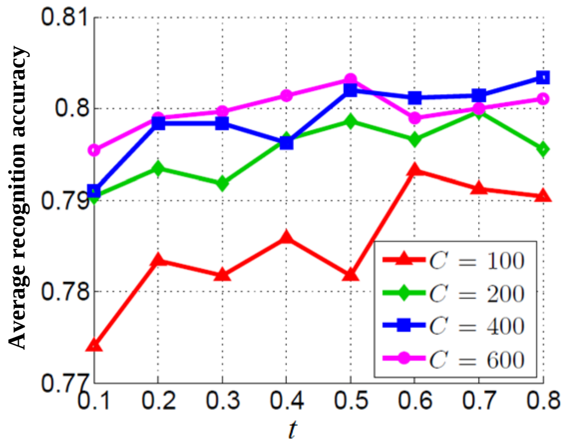

Figure 2. The Influence of Parameter C and t on Recognition Accuracy of Proposed Algorithm

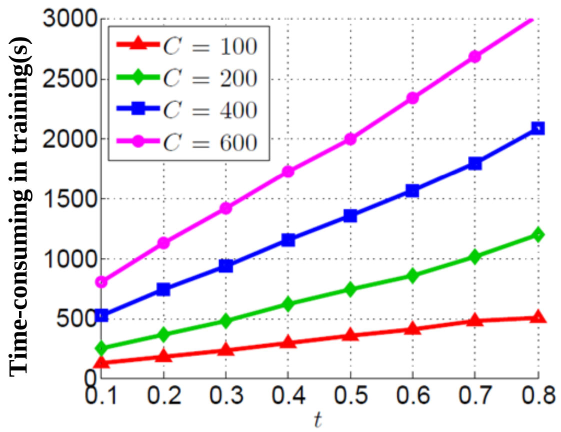

Figure 3. The Influence of Parameter C and t on Time-Consuming of Proposed Algorithm

It can be seen from the Figure 2, when C is set to be relatively larger, the average accuracy of the proposed algorithm will be better, but the training time will increase at the same time (as shown in Figure 3). This paper also uses other features on YouTube data sets and makes similar observation with LFW data set, in order to weigh the effectiveness and efficiency, when using all types of features, set C=400 on the two data sets. When parameter t is set between 0.2 to 0.8, the results of this study has become relatively stable, considering that no predefined attached dataset used for model selection, and in following experiment set parameter t as 0.5 on YouTube face dataset.

3.3. Comparison and Analysis

This section makes comparison of the proposed algorithm with other relatively advanced algorithms on LFW and Youtube dataset.

3.3.1. LFW Dataset: In LFW data sets, this study uses eight types of features, including strength, LBP, Gabor feature and Block Gabor feature and the root of these features, and contrasts performance of PHL + SILD and low-level feature + SILD algorithm.

Strength feature is directly extracted by vectorization of each gray image to 12000-dimensional feature vector. For LBP feature, it first extracts 59 frames histogram from each 10×10 non-overlapping pixel blocks, then all the histograms series into a single

7080-dimensional feature vector. Use 40 Gabor kernel functions to extract Gabor feature from eight directions of five grades, in order to reduce the feature dimension, further using 10×10 conversion factor under-sampling Gabor filtering image, however, this important under-sampling process may reduce the performance of face recognition. Using Block Gabor feature additionally, before under-sampling with each Gabor filtering image be divided into six non-overlapping blocks, each block of the Gabor filtering sub-image only using 2×2 conversion factor for under- sampling, and then separate each block of Gabor feature, rather than put them in series into a long feature vector, for each pair of face image, the Gabor feature of six blocks using cosine function to calculate the six similarity, and then output a average score. In order to mix 8 class features, each pair of images is expressed as a 8-dimensional similarity features, then calculate the final similarity of each pair of images by using linear SVM, the experimental results are shown in Table 1.

Table 1. Performance of Different Types Low-Level Feature on LFW Dataset (Average Accuracy ±Standard Deviation)

feature name	feature type	low-level feature+SILD	PHL+SILD
strength	original	0.8020±0.0067	0.8097±0.0072
	root	0.8010±0.0056	0.7925±0.0045
LBP	original	0.8412±0.0034	0.8442±0.0062
	root	0.8485±0.0035	0.8542±0.0064
Gabor	original	0.7902±0.0059	0.8030±0.0065
	root	0.8102±0.0064	0.8335±0.0056
Block Gabor	original	0.8233±0.0052	0.8343±0.0067
	root	0.8452±0.0044	0.8510±0.0052
Merge result		0.8768±0.0050	0.8867±0.0070

It can be seen from Table 2, "PHL + SILD" uses SILD to perform dimension reduction is helpful to the improvement of recognition rate, it is worth mentioning that "PHL + SILD" used is middle-level feature, while “low-level feature+ SILD "are using the original low-level feature, in addition to the root of strength feature, execution effect of " PHL + SILD " in all other types of features are better than" low-level features + SILD "[14], when using Gabor feature root, performance increases 2.55%, indicates that the use of PHL method is very effective to middle-level feature extraction when learning the optimal separating hyperplane. Single feature based on "single LE” [6] method gets only 81.22% of recognition accuracy," PHL + SILD "using LBP root results is 85.42%," PHL + SILD "combined all eight classes of feature, the recognition rate can be as high as 88.67%.

In addition, this paper makes comparison of the proposed algorithm with several other relatively advanced algorithms, including: " Multi-area histogram " [5], " Method based on combined b/g sample" [10], "attribute and similar classifier"[9], "LE + small kind figure" [6], "CSML + SVM" [12], " Qualcomm brain inspiration characteristics " [8] and " interaction prediction "[13]. Comparison results as shown in Table 2, at the same time, ROC curve is also given in Figure 4.

Table 2. Performance Comparison of Proposed PHL+SILD and Other Advanced Algorithms on LFW Datasets (Average Accuracy ±Standard Deviation (SE))

Type	Method	Average accuracy ±standard deviation
No attached data	Multi-area histogram [5]	0.7925±0.0055
	Multi-LE + small sample figure [6]	0.8445±0.0046
	Low-level feature + SILD [14]	0.8768±0.0050

Tagged attached data	CSML + SVM [12]	0.8800±0.0037
	Qualcomm brain inspiration features [8]	0.8813±0.0058
	Properties and similar classifier [9]	0.8529±0.0123
	Interaction prediction [13]	0.9057±0.0056
Untagged attached data	Method based on combined b/g sample [10]	0.8683±0.0034
	PHL + SILD (proposed algorithm)	0.8867±0.0070

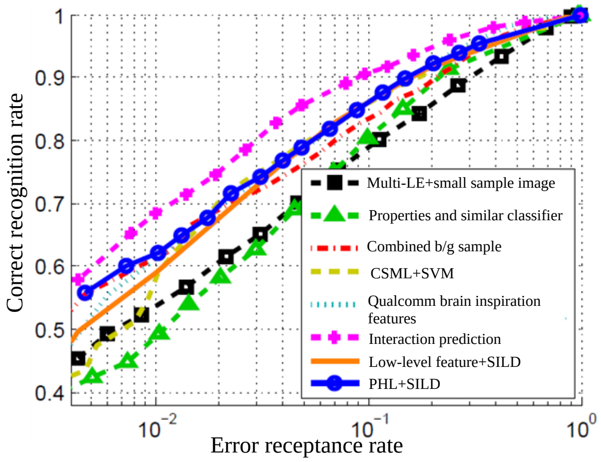

Figure 4. ROC Curve of Each Algorithm on LFW Dataset

It can be seen from Table 3, and Figure 4, PHL + SILD is a bit poor than " interaction prediction" algorithm in the literature [13], but, the " interaction prediction" requires a personal internal changes of strong-tagged attached data sets, and the proposed PHL + SILD only need attached untagged data sets. Compared with several other comparison algorithms, this algorithm has achieved better recognition effect, it does not need to use the attached data in literature [5-8, 12-14].

3.3.2. YouTube Dataset: On YouTube dataset, the experiment directly uses three characteristics provided by literature [15] (*i.e.*, LBP, CSLBP and FPLBP), considering all the face images are all aligned by the fixed face key point detection, so it extracts the average characteristics from all frames in a video clip, so as to output average vectors for follow-up process of PHL + SILD and low-level feature + SILD.

Experiment compares PHL+SILD, "MBGS" [15] algorithm with low-level feature + SILD algorithm, average precision, area under the curve (AUC) and entire error rate t (EER) data of the three kinds of algorithms are listed in Table 3.

Table 3. Recognition Results of Three Algorithms Using LBP, CSLBP and FPLBP Characteristic on YouTube Face Data Set (Average Accuracy ±Standard Deviation std),AUC and EER)

Character	Method	Average accuracy ±standard deviation	AUC	EER
LBP	MBGS[15]	0.764±0.018	0.826	0.253
	Low-level feature +SILD[14]	0.773±0.019	0.840	0.236
	PHL+SILD			
	PHL+SILD (proposed algorithm)	0.802±0.013	0.872	0.203
CSLBP	MBGS[15]	0.724±0.020	0.789	0.287
	Low-level feature +SILD[14]	0.736±0.015	0.804	0.286
	PHL+SILD	0.752±0.010	0.823	0.248

FPLBP	MBGS[15]	0.726±0.020	0.801	0.277
	+SILD[14] Low-level feature +SILD[14]	0.729±0.024	0.796	0.283
	PHL+SILD PHL+SILD (proposed algorithm)	0.759±0.015	0.825	0.244

It can be seen from Table 4, that, three kinds of characteristics performances used in PHL + SILD are better than "low-level feature + SILD ", and the average accuracy is 3% higher than " low-level feature + SILD", once again shows that using PHL learning classification hyperplane to extract characteristic is beneficial. Using LBP characteristic, the performance improvement of PHL + SILD algorithm in aspects of ACC, AUC and EER are 3.8%, 4.6% and 5% higher than MBGS, respectively, using the CSLBP and FPLBP characteristics, PHL + SILD algorithm is obviously better than that of MBGS.

As shown in Figure 5, are three algorithms using ROC curve of LBP characteristic and FPLBP characteristic. Also can be seen from the figure that the PHL + SILD method proposed in this paper are superior to "low-level feature + SILD" and MBGS in various cases.

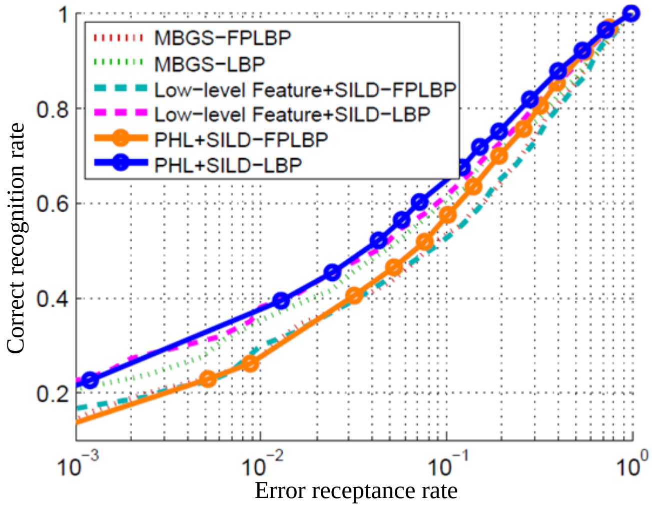

Figure 5. ROC Curve of Each Algorithm on YouTube Dataset

4. Conclusion

In order to better resolve the problem of influence of illumination, expression, posture, and scene larger changes on wild face recognition, this paper uses weak-tagged data sets to learn some binary classification hyperplane of the SVM model, of which support vector sparse set can automatically select from un-tagged conventional dataset, each sample in weak-tagged dataset expresses a middle-level characteristics, the term of characteristics is the decision-making value corresponding to learning SVM model, then, in each of the SVM model only select support vector sparse set from conventional dataset under constraint conditions, by using FLD-like objective function to maximize the discriminant ability of weak-tagged dataset to learn the optimal prototype hyperplane, and using SILD for dimension reduction, cosine similarity metric to complete human face recognition. Based on the two unconstrained conditions of datasets: wild tagged face image (LFW) and YouTube face data set, this paper makes contrast experiment on proposed prototype hyperplane learning algorithm (PHL) and several other relative advanced algorithms, the results verify the effectiveness and superiority of the proposed algorithm. It will apply the proposed algorithm to other non-constraint face data sets in the future, and combined with other advanced technologies, and carries on a large number of experiments to further improve the recognition performance.

Acknowledgements

The research is supported by National Natural Foundation Project (61503206).

References

[1] Z. Lv, X. Li and B. Zhang, "Managing Big City Information Based on WebVRGIS", IEEE Access, vol. 4, **(2016)**, pp. 407-415.

[2] J. Yang, Y. Lin and Z. Gao, "Quality Index for Stereoscopic Images by Separately Evaluating Adding and Subtracting", PloS one, vol. 10, no. 12, **(2015)**, e0145800.

[3] D. Jiang, Z. Xu and Z. Lv, "A multicast delivery approach with minimum energy consumption for wireless multi-hop networks", Telecommunication Systems, **(2015)**, pp. 1-12.

[4] Y. Liu, J. Yang and Q. Meng, "Stereoscopic image quality assessment method based on binocular combination saliency model", Signal Processing, vol. 125, **(2016)**, pp. 237-248.

[5] Z. Lv, A. Tek and F. Da Silva, "Game on, science-how video game technology may help biologists tackle visualization challenges", PloS one, vol. 8, no. 3, **(2013)**, e57990.

[6] X. Zhang, Y. Han, D. Hao and Z. Lv, "ARPPS：Augmented Reality Pipeline Prospect System", 22th International Conference on Neural Information Processing), Istanbul, Turkey. In press, ICONIP (**2015)**.

[7] W. Gu, Z. Lv and M. Hao, "Change detection method for remote sensing images based on an improved Markov random field", Multimedia Tools and Applications, **(2016)**.

[8] Z. Lu, C. Esteve, J. Chirivella and P. Gagliardo, "A Game Based Assistive Tool for Rehabilitation of Dysphonic Patients", 3rd International Workshop on Virtual and Augmented Assistive Technology (VAAT) at IEEE Virtual Reality 2015 (VR2015), Arles, France, IEEE, **(2015)**.

[9] Z. Lv, A. Halawani, S. Feng, H. Li and S. U. Rehman, "Multimodal Hand and Foot Gesture Interaction for Handheld Devices", ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol.1, no. 1, Article 10, **(2014)** October, pp. 19 .

[10] Y. Lin, J. Yang, Z. Lv, W. Wei and H. Song, "A Self-Assessment Stereo Capture Model Applicable to the Internet of Things", Sensors, **(2015)**.

[11] W. Ou, Z. Lv and Z. Xie, "Spatially Regularized Latent topic Model for Simultaneous object discovery and segmentation", The 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC **2015**).

Authors

Zhou Gaiyun, received her Bachelor's degree in Electrical and Information Engineering School from Zhengzhou university in Zhengzhou, HenanProvince,China. She is currently a lecturer in the Software College of Pingdingshan university. Her Research interests is mainly in the area of image processing, pattern recognition, and game development. She has published several research papers in scholarly journals in the above research areas and has participated in several books.

