

No-Reference JPEG image quality assessment based on Visual sensitivity

Zhang You-Sai

School of Electrical Information

Jiangsu University of science and technology, Zhen Jiang, China

Email: yszhang100@163.com

Chen Zhong-Jun

School of Electrical Information

Jiangsu University of science and technology, Zhen Jiang, China

Email: chen.magic@163.com

Abstract—in this paper, a novel human visual sensitivity measurement approach is presented to assess the visual quality of JPEG-coded images without reference image. The key features of human visual sensitivity (HVS) such as edge amplitude and length, background activity and luminance are extracted from sample images as input vectors. SVR-NN was used to search and approximate the functional relationship between HVS and mean opinion score (MOS). Then, the measuring of visual quality of JPEG-coded images was realized. Experimental results prove that it is easy to initialize the network structure and set parameters of SVR-NN. And the better generalization performance owned by SVR-NN can add the new features of the sample automatically. Compared with other image quality metrics, the experimental results of the proposed metric exhibit much higher correlation with perception character of HVS. And the role of HVS feature in image quality index is fully reflected.

Index Terms — Human visual sensitivity; support vector regression; neural network; image quality; No-reference assessment

I INTRODUCTION

Image quality evaluation plays an important role in processing image. With the extensive application of image, developing image quality metric without reference image has received widespread attention especially when it is difficult to obtain reference image. Thanks to image serving people, image quality assessment is more and more dependent on the characteristics of human visual system (HVS). Considerable volume of research has demonstrated that image quality evaluation methods considering human

visual characteristics is better than others not considering these characteristics [1]. Therefore, it is imperative to develop the no-reference image quality metric based on human visual factors.

In the last few decades, extensive valuable research has been carried out in developing this topic. Gastaldo et al. proposed a circular back propagation (CBP)-based image quality evaluation method [1], Venkatesh Babu et al. proposed a no-reference image quality index using growing and pruning radial basis function (GAP-RBF) [2] and Suresh et al. proposed a no-reference metric based on extreme learning machine classifier [3].

JPEG is one of the most popular and widely used image formats in internet and digital cameras. In this paper, for JPEG images, the extracted visual sensitivity approach is used to assess the visual quality of images without any reference. The key human visual sensitivity factors were used as input vectors of network. Image quality estimation includes computation of functional relationship between HVS features and subjective test scores. Here, the functional relationship is approximated using support vector regression neural network. The experimental results show that the proposed no-reference image quality metric has a good consistency with mean opinion score (MOS), really embodying the role of HVS features in image quality measurement.

Zhang You-Sai and Chen Zhong-Jun done the research work together in Jiang Su University of science and technology, and were grateful for LIVE database supported by Prof. H.R. Sheikh, etc.

Email: yszhang100@163.com; chen.magic@163.com

II HVS-BSED FEATURE EXTRACTION

A key distortion of JPEG images is horizontal and vertical blocking artifact generated by DCT-based transform coding for per 8×8 image block. In order to measure this kind of distortion accurately, several important human visual sensitivity factors such as edge amplitude and length, background activity and luminance [4] are taken into consideration. Given a gray scale image of size $M \times N$. The intensity of the image at any pixel location (i, j) is given by $I(i, j)$. The algorithm is explained for extracting human visual features along horizontal direction.

(1) Edge amplitude and length: Edge amplitude quantifies the strength of edge along the borders of 8×8 blocks; Edge length quantifies the length of continuous block edges. They are obtained by horizontal orthogonal sobel filter operator.

$$E = \begin{cases} |I * P|, & |I * P| < \tau_e \\ 0, & \text{others} \end{cases}, \quad (1)$$

P is the sobel horizontal filter, E is edge information. Threshold τ_e should be below 40, if choosing a lower threshold might result in missing the real blocky edges resulting from compression.

(2) Background activity: Background activity is denoted by the amount of high frequency texture content around the block edges. It is extracted by following method:

$$M_h = \begin{cases} 1, & |I * F_{ah}| < \tau_a \\ 0, & \text{others} \end{cases} \quad (2)$$

Here, F_{ah} is high-pass filtering, M_h is background activity. The value of threshold τ_a in our experiments is 0.3, the effect of blockiness is masked by the activity if τ_a is more than 0.3.

(3) Background luminance: Background luminance measures the amount of brightness around the block edges. It is obtained by following way:

$$W_l(i, j) = \begin{cases} \sqrt{\frac{I * fl_p}{128}}, & 0 \leq I(i, j) \leq 128 \\ 1, & \text{others} \end{cases} \quad (3)$$

Here, fl_p is low-pass filter, W_l is background luminance.

(4) Obtain the combined sensitivity difference along the horizontal direction (D_H):

$$P_v(i) = \sum_{j=1}^N E_h(i, j) \times M_h(i, j) \times W_l(i, j) \quad (4)$$

Here, $\bar{P}_v = \text{median}(P_v)$, $\bar{P}'_v = P_v(8n)$,

$$\bar{P}'_v(n) = \bar{P}_v(8n), D_H = \frac{1}{M}(P'_v - \bar{P}'_v),$$

Here, P_v is row vector of horizontal sensitivity features, $\text{median}(\cdot)$ is a five tap median filter, $P_v(8n)$ denotes only extracting values of the borders of 8×8 blocks.

In order to explain this process clearly, the vertical profile and median filtered profile of single image in horizontal direction is shown in Fig.1 and Fig.2, it is explicit that there is sharp rise between both neighboring locations because of block compression, which is in line with Reference [1]. Generally speaking, the more the compression ratio, the greater the rise value and the difference of both reflect human being's different sensitivity for different area in an image. The sensitivity difference is shown in Fig.3.

Similar steps are used to obtain the sensitivity difference along vertical direction (D_V). The final sensitivity difference (D_F) is got by combining the horizontal and vertical sensitivity difference ($D_F = [D_H, D_V]$). The HVS feature vector X is able to be obtained by quantizing D_F into 14 intervals. The 14 intervals are divided as following way: $[\text{min}, -1]$, $[-1, -0.6]$, $[-0.6, -0.3]$, $[-0.3, -0.1]$, $[-0.1, -0.05]$, $[-0.05, -0.01]$, $[-0.01, 0]$, $[0, 0.05]$, $(0.05, 0.1]$, $(0.1, 0.3]$, $(0.3, 0.6]$, $(0.3, 1.5]$, $(1.5, 2.5]$, $(2.5, \text{max}]$.

The extracted HVS feature from 233 JPEG images is showed in Fig.4. Here, x axis, y axis and z axis respectively denote dimension per image, the number of sample images and corresponding HVS feature. It can be observed from Fig.1 that most of HVS features fall in the range of (-1, 0.3) and evenly distributed feature could ensure difference of different images.

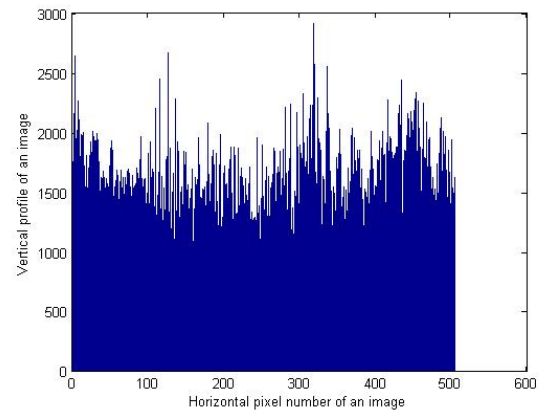


Fig.1 The vertical profile of an image

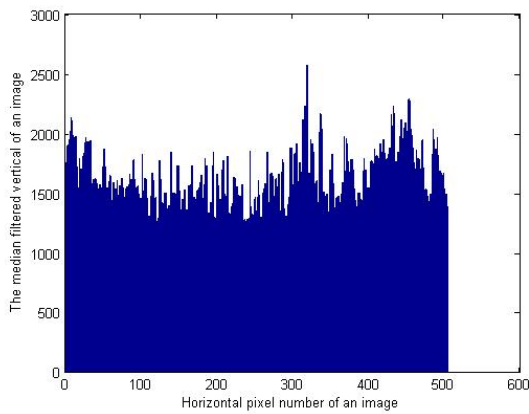


Fig.2 The median filtered profile of an image

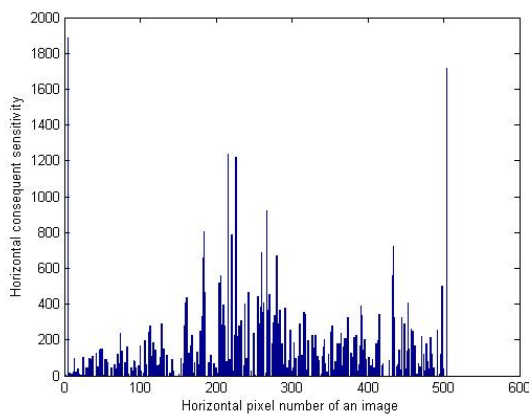


Fig.3 Horizontal consequent sensitivity difference

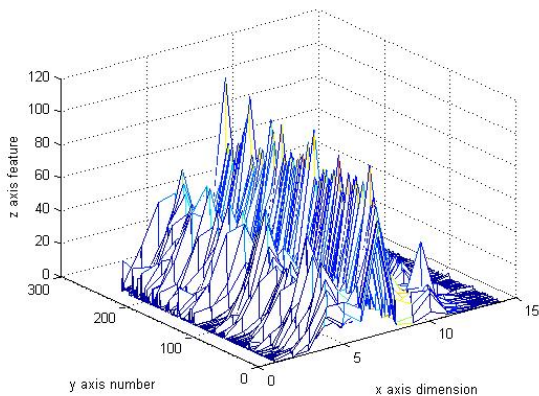


Fig.4 Illumination of HVS features

III NO-REFEREENCE IMAGE QUALITY MODEL

Because of the goal of image quality assessment for people, the performance of measurement approach considering human visual sensitivity is highly better than those without considering them. Therefore, the human visual sensitivity objective score (HVSOS) is presented to assess image quality. The support vector regression neural network (SVR-NN) is utilized to approximate the functional relationship between human visual sensitivity

features and mean opinion score. The HVSOS index considers human visual sensitivity factors as a criterion to evaluate the image quality, so the image visual quality is emphasized and then the process of image quality assessment is to some extent simplified.

A. The basic principles of SVR-NN

The basic idea of SVR-NN algorithm is that architecture of SVR-NN is initialized by SVR and then corresponding parameters of network is updated using neural network. In this paper, we select ϵ -SVR for initializing architecture of SVR-NN because of its better sparsity performance and Gaussian function as kernel function of SVR-NN because of its better performance of fitting. Therefore, the architecture of SVR-NN can be written as [5]:

$$f(x) = \sum_{i=1}^s (\alpha_i - \alpha_i^*) \exp\left(-\frac{\|x - v_i\|^2}{2\gamma^2}\right), \quad (5)$$

Here, α_i is vector solution, α_i^* is dual vector solution, s denotes the number of support vector, v and γ are center vector and width of the hidden node.

B. HVSOS metric approach model

Mentioned above, HVSOS model applied in image quality is constructed as Fig.2:

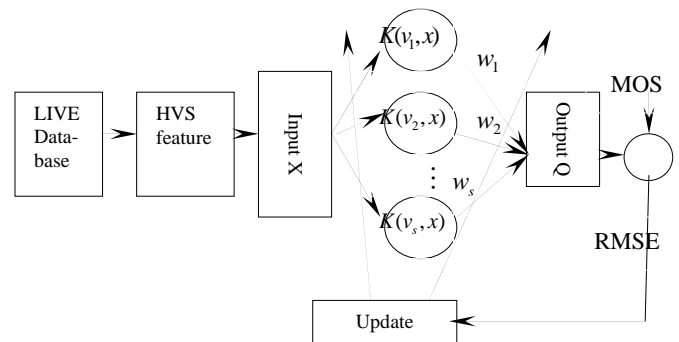


Fig.2 the HVSOS metric model of image quality

From Fig2, it is shown that the architecture of HVSOS index is regarded as a two-layer neural linear network and in order to beat the target of MOS convergence, neural network updates weight vector, center vector and width of the hidden node of SVR-NN based on the convergence of RMSE.

$$\hat{Q}_{HVSOS} = \sum_{i=1}^s w'_i \exp\left(-\frac{\|x - v_i\|^2}{2\gamma^2}\right) + b, \quad (6)$$

Eq. (6) shows the output result of HVSOS assessment metric. Here, $x = [x_1, x_2, \dots, x_q]^T$ is input sample matrix extracted in section II, s denotes the number of support vector, $w'_i = (\alpha_i^* - \alpha_i) \neq 0$ is weights needed

updating, v and γ are center vector and width of the hidden node needed updating, b is bias of network.

To obtain better assessment accuracy, neural network algorithm based on gradient-decent method is used to update SVR-NN network. The lost function for HVSOS is defined here [6]:

$$E^R = \sum_{i=1}^L \frac{\mu(t)}{2} \ln \left(1 + \left(\frac{e_i^2(t)}{\mu(t)} \right) \right), e_i(t) = y_i - \hat{y}_i(t), \quad (7)$$

Here, t is the epoch number; i is the layer number, $\hat{y}_i(t)$ is assessment value. Update the weight vector, center vector and width of the hidden node according to [7]:

$$\begin{aligned} \delta w'_j &= -\frac{\partial E_j^R}{\partial w'_j} = -\frac{e_j(t)}{1 + \frac{e_j^2(t)}{\mu(t)}} \cdot \frac{\partial e_j(t)}{\partial w'_j} \\ &= -\frac{e_j(t)}{1 + \frac{e_j^2(t)}{\mu(t)}} \cdot \exp \left\{ -\frac{\|x - v_j\|^2}{2\gamma_j^2} \right\}, \quad (8) \end{aligned}$$

$$\begin{aligned} \delta v_{i,j} &= -\frac{e_j(t)}{1 + \frac{e_j^2(t)}{\mu(t)}} \cdot \frac{\partial e_j(t)}{\partial v_{i,j}} \\ &= w'_j \cdot \frac{x_i - v_{i,j}}{\gamma_j^2} \cdot \exp \left\{ -\frac{\|x - v_j\|^2}{2\gamma_j^2} \right\} \cdot \frac{e_j(t)}{1 + \frac{e_j^2(t)}{\mu(t)}} \quad (9) \end{aligned}$$

Here, x_i is i th dimension of x , $i = 1, 2, \dots, q$.

$$\begin{aligned} \delta \gamma_j &= -\frac{e_j(t)}{1 + \frac{e_j^2(t)}{\mu(t)}} \cdot \frac{\partial e_j(t)}{\partial \gamma_j} \\ &= w'_j \cdot \frac{\|x - v_j\|^2}{\gamma_j^3} \cdot \exp \left\{ -\frac{\|x - v_j\|^2}{2\gamma_j^2} \right\} \cdot \frac{e_j(t)}{1 + \frac{e_j^2(t)}{\mu(t)}} \quad (10) \end{aligned}$$

$$w_j^{(k+1)} = w_j^{(k)} + \eta \cdot \delta w'_j + \alpha \cdot (w_j^{(k)} - w_j^{(k-1)}), \quad (11)$$

$$v_{i,j}^{(k+1)} = v_{i,j}^{(k)} + \eta \cdot \delta v_{i,j} + \alpha \cdot (v_{i,j}^{(k)} - v_{i,j}^{(k-1)}), \quad (12)$$

$$\gamma_j^{(k+1)} = \gamma_j^{(k)} + \eta \cdot \delta \gamma_j + \alpha \cdot (\gamma_j^{(k)} - \gamma_j^{(k-1)}), \quad (13)$$

Here, k is the number of iteration η is the learning rate, α is momentum coefficient.

The $\frac{e_j(t)}{1 + \frac{e_j^2(t)}{\mu(t)}}$ in Eq. (8), (9), (10), which resulted

from lost functions such as Eq. (7) has a great impact on the approximation between objective assessment results and MOS especially when there are outliers

exist. If $\frac{e_j(t)}{1 + \frac{e_j^2(t)}{\mu(t)}}$ weaken too quick, there is not

enough time for most of training data to converge and then evaluation results of HVSOS index are mostly smaller than actual values, that is to say, under fitting occurs; if this decay is too slow, there is too a long time for training data to converge in time, and then assessment results are greater than actual results. Ultimately, the outliers are mistaken as normal training data. That is to say, the over fitting happens.

Therefore, the suitable decay method is that $u(t) = A/t$ is selected, and then the better generalization performance is achieved, especially when A is a constant [6].

IV EXPERIMENTS AND DISCUSSIONS

In our simulations, we have used the LIVE image quality assessment database I [8] which is constituted of 29 JPEG original images and 204 JPEG images with different compression rates from original images. The 233 images are divided into two disjoint sets for training and testing. The corresponding MOS need be transformed to 1-10 range. In order to validate the effect of evaluation of proposed HVSOS, Wang's NR [9], ELM [3] and SSIM index [10] are compared to proposed HVSOS algorithm.

After training phase, the parameters are set as follows:

$\varepsilon = 1.25$, $C = 31$, $\gamma = 27$, $\eta = 0.1800$, $\alpha = 0.0200$, the number of iteration is 300. The error RMSE convergence curve is shown in Fig.5.

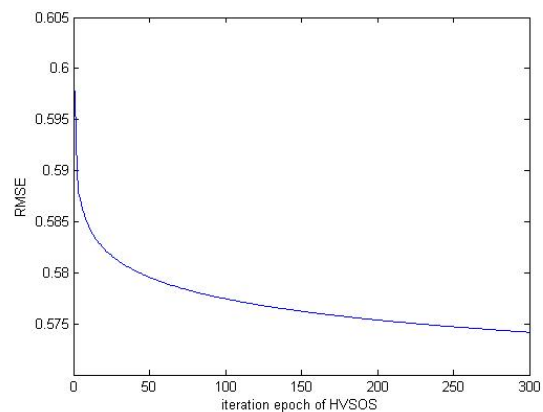


Fig.5 Error convergence curve of HVSOS

A. Visual analysis of HVSOS

In our experiments, the LIVE image database is divided into train set and test set. The PSNR and HVSOS are respectively used to predict the train and test sets. Fig.6 and Fig.7 show comparison measurement results for train sets, Fig.8 and Fig.9 for test sets.

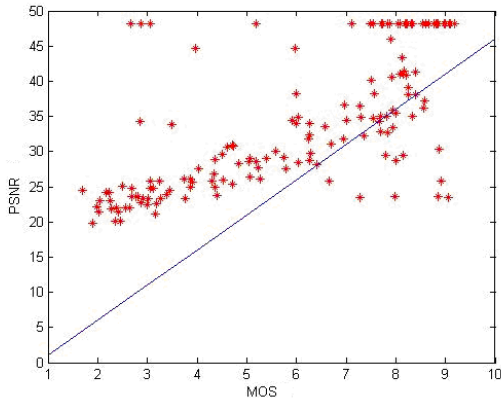


Fig.6 PSNR quality assessment of training sets

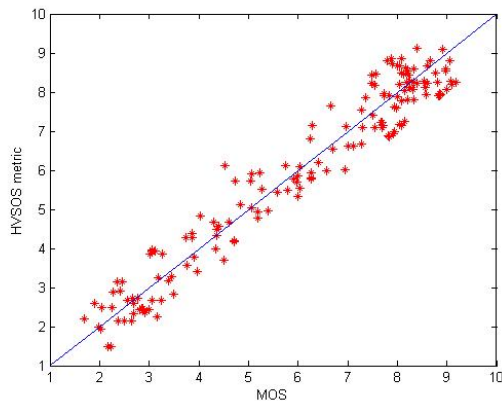


Fig.7 HVSOS quality assessment of training sets

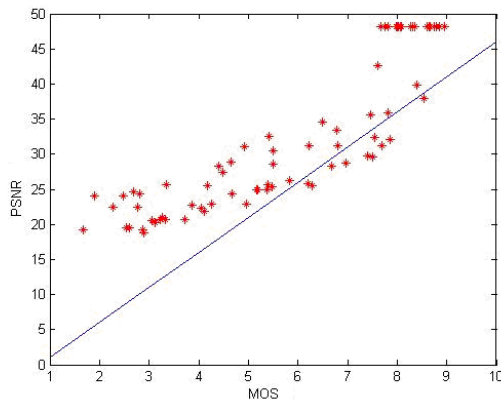


Fig.8 PSNR quality assessment of testing sets

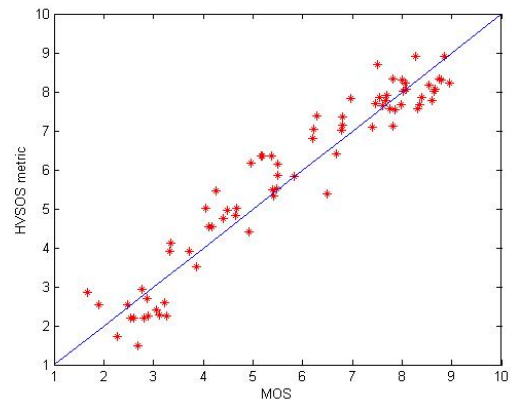


Fig.9 HVSOS quality assessment of testing sets

As can be seen from Fig.6 to Fig.9, whether is train sets or test sets, the evaluation results of HVSOS index are all able to precisely vary with diversification of MOS values, more evenly distributing around the ideal straight line and then there are few outliers interfering with experimental results.

In order to vividly explain this phenomenon, here show four group pictures, such as Fig.10 (a) (b), Fig.11 (a) (b), Fig.12 (a) (b), Fig.13 (a) (b), each of which is composed of both same pictures except the blocking effect in different aspects. These pictures reveal human visual sensitivity effect in different aspects. There is better visual effect in Fig.10 (a) and Fig.11 (a) than in Fig.10 (b) and Fig.11 (b), because of the (b) pictures' bad blocking in them; while the both group of Fig.12 and Fig.13 owe the same high visual quality and little visual difference between (a) and (b) of each group.



(a) (b)

Fig.10 LIVE: churchandcapitol



(a) (b)

Fig.11 LIVE: floweronih35



(a) (b)

Fig.12 LIVE: woman hat



(a) (b)

Fig.13 LIVE: cemetery

The table 1 shows each evaluation result that is predicted in HVSOS and PSNR two metrics.

TABLE 1

The comparison of various pictures between HVSOS and PSNR

LIVE Image database		Index		
		HVSOS	PSNR	MOS
churchand-capitol	(a)	7.0136	29.8147	6.2800
	(b)	2.5153	26.5907	2.6809
floweronih35	(a)	4.6156	22.2849	4.0492
	(b)	1.5945	20.2892	1.6803
woman hat	(a)	8.4810	48.1308	8.9675
	(b)	7.9951	38.0535	8.2422
cemetery	(a)	7.7467	48.1308	7.8412
	(b)	7.7244	32.2077	7.3663

Seen from table 1, it is obvious and precise that the HVSOS index gave right difference for big difference between Fig.10 (a) and Fig.10 (b), so is it for Fig.11 (a) and Fig.11 (b); while the PSNR index assesses the very similar evaluation results, which is not consistent with the corresponding MOS values. For the Fig.12 (a) and

Fig.12 (b), they have close visual quality. This is highly consistent with HVSOS index's evaluation result and MOS value, which is contrary to PSNR giving assessment results for Fig.13 (a) and Fig.13 (b).

B. The comparison between HVSOS approach and other popular metrics

TABLE 2

Comparison of various metrics

Metrics	Training sets		Testing sets		Type
	RMS E	R-square	RMS E	R-square	
Wang's NR	4.6900	-2.3930	5.0900	-2.5170	NR
ELM	0.6900	0.9010	0.7000	0.9230	NR
SSIM	0.6300	0.9408	0.6500	0.9431	FR
HVSOS	0.5776	0.9378	0.6131	0.9201	NR

From TABLE 2, obviously, the RMSE value of HVSOS model is much less than that of other three algorithms and the R-square value of HVSOS are much more excellent than that of Wang's NR and SSIM algorithms. Therefore, we could conclude that the proposed HVSOS quality model owns better generalization performance and its measurement result has a higher consistency with HVS features.

V CONCLUSIONS

In this paper, a novel HVSOS metric approach is presented to assessment the visual quality of JPEG-coded images considering various HVS features. The functional relationship between the extracted HVS features and MOS is modeled by SVR-NN. The network can be updated easily and the performance of the proposed metric is better than others. This metric can be easily extended to measure the quality of videos which also use similar block DCT-based compression.

ACKNOWLEDGMENT

The authors would like to express their thanks to Prof. H.R. Sheikh and Prof. Bovid for providing the JPEG image quality assessment database to validate our metric.

REFERENCES

- [1] P.Gastaldo , R.Zunino, “No-Reference Quality Assessment of JPEG Images by Using CBP Neural Networks,” *IEEE International Symposium on Circuits and Systems*, pp.772-775, May2004.
- [2] R. Venkatesh Babu, S. Suresh, Andrew Perkis, “No-reference JPEG-image quality assessment using GAP-RBF,” *Signal Processing of ScienceDirect*, A87, 2007, pp.1493-1593.
- [3] S. Suresh, R. Venkatesh Babu, “H.J. Kim. No-reference image quality assessment using modified extreme learning machine classifier,” *Applied Soft Computing of ScienceDirect*, pp.541–552, September2009,
- [4] S.A. Karunasekera, N.G. Kingsbury, A distortion measure for blocking artifacts in images based on human visual sensitivity, *IEEE Transactions on Image Processing*, 4th ed., pp.713-724, June1995.
- [5] C-C Chuang, S-F Su, “Robust Support Vector Regression Networks for Function Approximation with Outliers,” *IEEE Trans on Neural Networks*, 13th ed., pp.1322-1330, June2002.
- [6] C-C Chuang, S-F Su, C-C Hsiao, “The annealing robust backpropagation (ARBP) learning algorithms,” *IEEE Trans on Neural Networks*, 11th ed., pp.1067-1077, May2000.
- [7] Li Jun, Liu Junhua, “Identification of dynamic systems using support vector regression neural networks,” *Journal of Southeast University (English Edition)*, pp.228-233, June 2006,.
- [8] H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik, “Live image quality assessment database”, website: <http://live.ece.utexas.edu/research/quality>.
- [9] P.Gastaldo , R.Zunino, “No-Reference Quality Assessment of JPEG Images by Using CB PNeural Networks,” *IEEE International Symposium on Circuits and Systems*, pp.772-775, May2004.
- [10] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.* 3rd ed., pp.600-612, April2004.

Prof. Zhang You-Sai, who was born in 1959, Zhenjiang city, Jiangsu Province, researches on image processing and 3D visualization;

Chen Zhong-Jun, who was born in 1983, He Nan province, researches on image and video processing.

Communication author:

Zhang You-Sai, Jiang Su University of science and technology, 2# Meng Xi Road, Zhen Jiang city of Jiang Su province in China.

Mobile Phone: 13862442688

Email address: yszhang100@163.com