NO-REFERENCE IMAGE QUALITY ASSESSMENT FOR ZY3 IMAGERY IN URBAN AREAS USING STATISTICAL MODEL

Y. Zhang [a, *], W.H. Cui [a], F. Yang [a], Z.C.Wu [a]

[a] School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan, Hubei, P.R.China-
(ivory2008, whcui)@whu.edu.cn, -yangfan_whu@sina.cn, -zcwoo@whu.edu.cn

ICWG III/VII

ABSTRACT:

More and more high-spatial resolution satellite images are produced with the improvement of satellite technology. However, the quality of images is not always satisfactory for application. Due to the impact of complicated atmospheric conditions and complex radiation transmission process in imaging process the images often suffer deterioration. In order to assess the quality of remote sensing images over urban areas, we proposed a general purpose image quality assessment methods based on feature extraction and machine learning. We use two types of features in multi scales. One is from the shape of histogram the other is from the natural scene statistics based on Generalized Gaussian distribution (GGD). A 20-D feature vector for each scale is extracted and is assumed to capture the RS image quality degradation characteristics. We use SVM to learn to predict image quality scores from these features. In order to do the evaluation, we construct a median scale dataset for training and testing with subjects taking part in to give the human opinions of degraded images. We use ZY3 satellite images over Wuhan area (a city in China) to conduct experiments. Experimental results show the correlation of the predicted scores and the subjective perceptions.

1. INTRODUCTION

Chinese high resolution remote sensing satellites have generated a large number of remote sensing (RS) images every day. However users are still lack of domestic high quality RS data in application. To change this situation, in addition to improve the hardware conditions, satellite image quality evaluation and enhancement is the fundamental assurance for the commercial application.

In the chain of optical image acquisition for RS satellites the platforms, sensors, atmospheric environment, the surface albedos all have influence on image quality. Of above factors satellite platform and sensors' impact on the geometry precision could be eliminated by sensor calibration while the image quality degradation caused by path scattering and the spectral mixture from the surrounding ground is difficult to eliminate, especially in urban areas. Compared to the mountainous area and the cultivated region the RS images in urban areas are mainly featured with artificial structures such as buildings, roads, tree, and lawns. The RS city images often contain complicated structures, high contrast objects and large information contents. Moreover with the industrial development and environmental pollution in cities the frequently occurred fog and haze has made RS city image quality evaluation and improvement a very difficult task.

For one scene of a RS image the data quality can be divided into three parts: the geometric quality, the radiation quality and the quality of auxiliary data. The geometric quality is generally evaluated by internal and outside geometric positioning accuracy. The auxiliary data quality can provide the RS image information described by the data completeness, correctness and logic check results. Digital image based radiation quality expresses the ability of holding the surface reflectance characteristics (Wang, 2014). Depending on whether or not there are observers, the existing RS image quality evaluation is divided into two categories: the subjective evaluation and objective evaluation. For objective RS image quality evaluation a lot of previous work have been conducted and are mainly concentrated in the following four aspects (Chen, 2011 Ma, 2014): (1) Task and application oriented RS image quality evaluation. The most typical example is the U.S. NIIRS which is used to assess the quality of military RS image classification (Irvine, 1997). The quality of image fusion, the classification accuracy and the scale of surveying and mapping are all task oriented. However, to fulfil the RS task successfully is not only decided by the quality of RS images but also by the contribution of data processing system and the visual interpretation done by workers. (2) Methods based on imaging system performance evaluation. In addition to the spatial resolution, temporal resolution, spectral resolution and radiation resolution, such methods focus on Modulation Transfer Function (MTF) based indexes and Signal to noise ratio (SNR) assessment (Zhang, 2002 Ferzli, 2009). Good imaging system does generate high quality RS images. But the ground targets and imaging environment also play essential roles to the RS images acquisition. (3) Methods based on the image feature extraction. Such methods are mostly inherited from image processing field. They endeavour to find out the image features to represent the image quality degradation. Besides the traditional image features such as mean, variance, entropy, edges, contrast,

---

* Corresponding author

moments some recent features include JNB(Ferzli, 2009), CPBD(Narvekar, 2011), MLV(Bahrami, 2014), BRISQUE (Mittal, 2012), RIQMC(Gu, 2016) etc.. From the point of view of information theory and signal processing the above features try to capture the image edges and details, simulate the image blurring and contrast degradation and provide powerful tools to assess the image quality. However, the majority of the above methods are for close range visible images. They seldom consider the image degradation caused by atmospheric blurring, clouds, reflectance, sampling, quantization and other complicated conditions for RS imaging process. So it will be meaningful to adopt those methods to RS images and explore the effective features to express the characteristics of RS images. (4) Learning based methods. Some recent research includes CNN method (Kang, 2014), deep learning network (Gu, 2014), DOG model and random forest (Pei, 2015). These methods try to reveal the image quality degradation mechanism in feature space by training an empirical model then the prediction model is used to calculate the test image quality score. However most of these methods have strong image distortion dependence and have certain requirements on the sample selection and sample numbers. In most cases the computational complexity is also high which has restricted the use of such methods.

In addition, due to the change of weather and air pollution there often exist clouds and haze in RS urban images, which leads to the loss of the local feature information. So the RS urban image quality evaluation must consider the degradation caused by cloud and haze. The existing cloud-index based features for RS images lack of adaptability for thin cloud and mist. And at least four bands are required to calculate the cloud index which is not suitable for panchromatic RS images.

In this paper we adopt a general purpose framework to construct RS urban image quality indices based on natural scene statistic (NSS) models based on multi spectral RS images. Inspired by the work of Moorthy (Moorthy, 2010a) and Mittal (Mittal, 2013) we establish a model that can learn to predict human judgments of image quality from databases of human rated degraded images. Cloud and haze-caused image degradation are primarily considered but not limited to them. We do not explicitly seek to characterize the structure of degradation using local filters, but instead utilize concepts from NSS to produce an easily extensible approach to other kinds of distortions. Once trained, our algorithm does not require further foreknowledge of the distortion affecting the RS images to be assessed.

The first contribution of this work is the development of a modular framework for image quality assessment of cloud and mist-deteriorated multi-spectral RS urban images. The modularity of the proposed method implies that the approach is extensible in that other distortion categories beyond the cloud and mist may be easily accomplished. The second contribution is the image quality index for RS urban images. Due to the diversity of the ground types, the multi-scale structures, the irregularity of spatial distribution of artificial buildings for urban areas the traditional filter based feature representation is ineffective. Our model is founded on perceptually relevant spatial domain NSS features extracted from local image patches that effectively capture the essential low-order statistics of RS images. We construct a dataset with image patches extracted from images acquired by multi-spectral sensors loaded on Chinese ZY3 satellite. We invited remote sensing image interpretation expert to give a quality score for each image patch and test our algorithm on the dataset. We demonstrate that our

algorithm performs well in terms of correlation with human perception.

2. QUALITY ASSESSMENT MODEL FOR REMOTE SENSED IMAGES OF URBAN AREA

2.1 Image normalization based on NSS

Compared to other images the RS urban images have rich feature categories, complex structures and abundant information content. To extract the perceptually relevant spatial NSS features from the image patches. The spatial NSS model (Moorhy, 2010b) that we use begins by preprocessing the image by processes of local mean removal and divisive normalization:

$$\tilde{I}(x,y,\lambda) = \frac{I(x,y,\lambda) - \mu(x,y,\lambda)}{\sigma(x,y,\lambda) + 1} \qquad (1)$$

Where $x \in \{1,..M\}$, $y \in \{1,..N\}$, $\lambda \in \{\lambda_1,..\lambda_n\}$ are spatial and spectral indices. For multi-spectral RS images $n = 4$, and

$$\mu(x,y,\lambda) = \sum_k \sum_l w_{k,l} I(x+k,y+l,\lambda) \qquad (2)$$

$$\sigma(x,y,\lambda) = \sqrt{\sum_k \sum_l w_{k,l}[I(x+k,y+l,\lambda) - \mu(x,y,\lambda)]^2} \qquad (3)$$

estimate the local mean and contrast respectively, where $w = \{w_{k,l} | k = -K,..K, l = -L,..L\}$ is a 2D circularly-symmetric Gaussian weighting function sampled out to 3 standard deviations($K = L = 3$) and rescaled to unit volume. The normalized image (1) has been observed to reliably follow a Gaussian distribution when computed from natural images.

Given one scene of RS urban image of size 8856 by 8476 for example. The image patches with size $P \times P$ are cropped randomly from the entire image. The normalization is computed on the image patches.

2.2 Multi-scale image quality indices

Given a collection of RS image patches, their qualities are characterized by the features computed from each patch. It has been generally noticed that the histogram indicates the probability distribution of image gray level and when computed over multi scales it reveals the statistical features in scale space. The types and degrees of image degradation caused by blur or cloud cover will generally affect the mean and shape of the local histograms in diverse ways.

Region 1 (a) B1   (b) B2

(c) B3　　　　　　　　(d) B4

(e) Histograms of (a-d).
Blue-B1,Green-B2,Red-B3,Black-B4

Region 2　(a) B1　　　　　(b) B2

(c) B3　　　　　　　　(d) B4

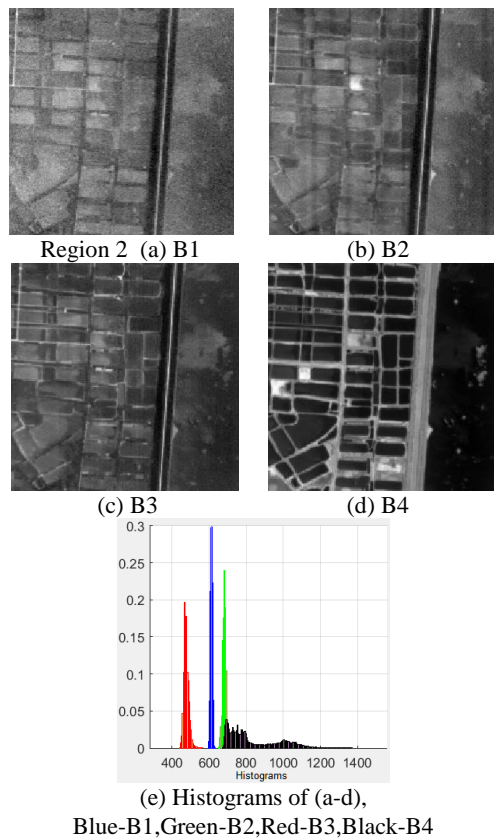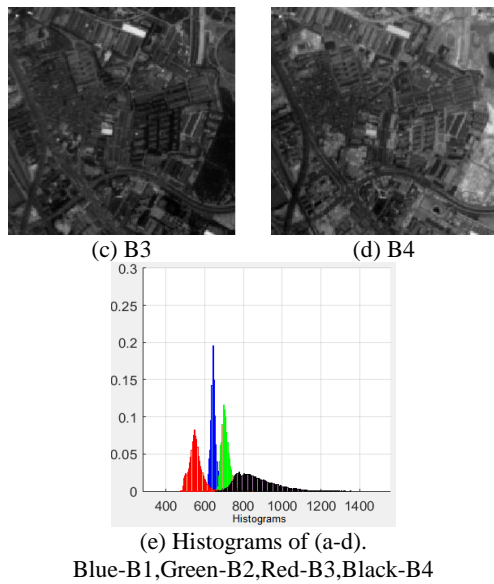(e) Histograms of (a-d),
Blue-B1,Green-B2,Red-B3,Black-B4

Figure 1. Two regions from ZY3 multispectral image (four bands with spatial resolution 5.8m) in Wuhan city and their histograms. The image patches are 200x200 and have been stretched and zoomed for display purpose.

As shown in Fig.1, for Region 1 from ZY3 multispectral image of Wuhan city the image patch has relatively high quality (according to RS image interpretation expert). In region 2, there has fog and blur in band1 and band2 so their histograms (blue and green plots in Region2 (e)) are significantly compressed in a very narrow area with high peaks due to loss of structural properties. The pinnacle pattern in histogram represents the influence of imaging noise or environmental fogs over city. It is

also noticed that for band 4 (black plot in Region2 (e)) the histogram has salient structural shapes since the infrared band is less affected by the fog. So the image quality of band 4 is thus relatively high. In order to express this characteristic, we use the mean and skew as descriptive quality features. For each image patch we extract these 2 features (f_1, f_2) from each scale, yielding $2 \times NS$ features for NS scale levels.

Prior studies of NSS based image quality have shown that the generalized Gaussian distribution effectively captures the behaviour of the normalized natural images (Mittal, 2013). It has been observed that the model would be violated when the images do not derived from a natural source or when natural images are subjected to unnatural distortions. The degree of modification can be indicative of perceptual image degradation severity. So we use GGD based parameters as features to describe the gray distribution characteristics of the urban image caused by system noise and irregular cloud and mist coverage. The generalized Gaussian distribution (GGD) with zero mean is given by:

$$f(x; \gamma, \beta) = \frac{\gamma}{2\beta\Gamma(1/\gamma)} \exp((-\frac{|x|}{\beta})^\gamma) \tag{4}$$

where $\Gamma(\cdot)$ is the gamma function:

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt \quad x > 0 \tag{5}$$

The parameter γ controls the 'shape' of the distribution. For example, $\gamma = 2$ yields a Gaussian distribution and $\gamma = 1$ yields a Laplacian distribution. The parameters of (4) ((γ, β)) can be estimated using moment-matching based approach (Sharifi, 1995) and be used as image quality features (f_3, f_4). The signs of the normalized image (1) have been observed to follow a regular structure. This deviation can be captured by analysing the sample distribution of the products of pairs of adjacent pixel values computed along spatial horizontal, vertical and diagonal orientation (Mittal 2012):

$$
\begin{aligned}
\tilde{J}_x &= [\tilde{I}(x,y) \cdot \tilde{I}(x+1,y)] \\
\tilde{J}_y &= [\tilde{I}(x,y) \cdot \tilde{I}(x,y+1)] \\
\tilde{J}_{d+} &= [\tilde{I}(x,y) \cdot \tilde{I}(x+1,y+1)] \\
\tilde{J}_{d-} &= [\tilde{I}(x,y) \cdot \tilde{I}(x+1,y-1)] \\
x &\in \{1,..M\}, y \in \{1,..N\}
\end{aligned} \tag{6}
$$

Displacement squared images $\tilde{J}_x, \tilde{J}_y, \tilde{J}_{d+}, \tilde{J}_{d-}$ are well-modeled as following a zero mode asymmetric generalized Gaussian distribution (AGGD):

$$
f(x; \gamma, \beta_l, \beta_r) =
\begin{cases}
\dfrac{\gamma}{(\beta_l + \beta_r)\Gamma(1/\gamma)} \exp(-(\dfrac{-x}{\beta_l})^\gamma) & \forall x \leq 0 \\[3mm]
\dfrac{\gamma}{(\beta_l + \beta_r)\Gamma(1/\gamma)} \exp(-(\dfrac{x}{\beta_r})^\gamma) & \forall x \geq 0
\end{cases} \tag{7}
$$

The mean of the distribution can also be used as a statistics:

$$\zeta = (\beta_r - \beta_r) \frac{\Gamma(2/\gamma)}{\Gamma(1/\gamma)} \qquad (8)$$

For each $\tilde{J}$ there are four features obtained $(\gamma, \beta_r, \beta_l, \zeta)$. By extracting estimates along the four orientations, 16 parameters are arrived at yielding ($f_5, ... f_{20}$). All features are computed at NS scales to capture multi scale behaviour, by low pass filtering and down sampling by a factor of 2, yielding a set of $20 \times NS$ features. This feature vector characterizes the distortion that the image patch is subject to.

2.3 SVM training and test

Machine learning has been applied in the field of image quality assessment for a long time. In order to map feature vectors to predicted quality scores we use support vector machine (SVM). SVMs are popular as classifiers since they perform well in high-dimensional spaces, avoid over-fitting and have good generalization capabilities. In our demonstration in order to produce a quality index, the $v-SVM$ is utilized to perform such a regression (Sch ölkopf 2000). Specifically, for each image patch we consider a $v-SVM$ is trained using quality scores from training set to learn the mapping from the feature space to subjective quality. When presented a test image the algorithm will produce a quality score which correlates with human perception. Finally, each $v-SVM$ constructed for quality assessment requires a set of parameters (C, γ_s) to be determined. They are set using 5-fold cross validation on the training set of images. We use the LIBSVM package (Chang 2011) to implement the SVR. In SVR, all the experiments use the radial basis function (RBF) kernel.

3. EXPERIMENTAL RESULTS

3.1 Dataset

In order to evaluate the performance of proposed algorithm we construct a median scale dataset for training and testing. The dataset was derived from a set of source RS images acquired by a multispectral camera loaded on ZY3 satellite. ZY3 is China's first high-resolution optical mapping satellite for civil use. It was launched on 9 Jan. 2012, equipped with four optical cameras. One panchromatic TDI CCD of 3.6m for nadir view, two panchromatic TDI CCD cameras of 2.1m for front and rear view and one multispectral camera of 5.8m for nadir view. We used 4-bands multi-spectral images of Level 1 as source without doing enhancement and radiometric correction. Some information of one image used are shown in Table 1.

Attribute	Value
Sensor type:	MUX
Image size (WidthInPixel×HeightInPixel）:	8856×9716
CenterTime:	20120620111111
Center Lon/Lat:	114.22/30.65
Coverage area:	Wuhan city in China
PixelByte	uint16
SolarAzimuth:	34.915783
SolarZenith:	161.861008
CloudPercent:	10%

Table 1. The information of one source image used for dataset

A total of 250 patches of 200×200 from each band were cropped from three source images. When do cutting, be sure that the sub-images contain a variety of ground types such as water, vegetation, artificial buildings, roads etc. and have different levels of noise and varying degrees of fog coverage. The subjects taking part in scoring are academic teachers of remote sensing science and technology from School of Remote Sensing and Information Engineering of Wuhan University in China. The average number of subjects ranking each image was 3. Each subject was individually told the goal of the scoring. In order to highlight the impact of fog on image quality we assumed that the images containing mist or fog have low scores. Therefore, for the four bands of one image band 4 always had the highest score and band 1 the lowest one. The subjects reported their judgements of quality by dragging a slider on a quality scale according to their estimation on image contrast, clearness and more importantly the mist and fog content. The decimal number for the score is in the range 0-1. All the scores for the same image were averaged. Raw difference score for an image was considered to be an outlier if it was outside an interval of deviations about the mean score for that image. After rejecting 33 outliers a total of 217 images and their corresponding scores were kept and constitute our dataset. Figure 2 shows one example of the subjective scores for blue, green, red and infrared band of a randomly selected image patch.
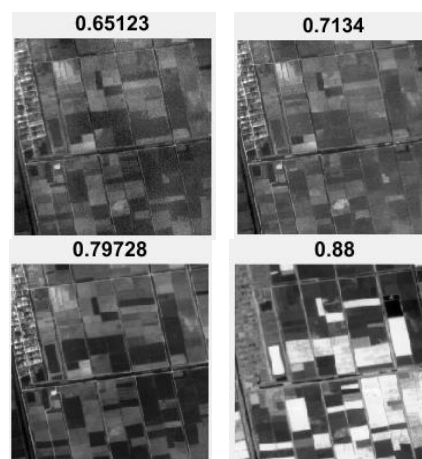

Figure 2. Four bands of an image patch with their subjective scores on title. Upper-left blue band, upper-right green band, lower-left red band and lower-right infrared. Notice that the infrared band has the highest quality score due to its less sensitivity to mist.

3.2 Results

For each sample of 217 images (four bands per image) in dataset we did normalization and calculated 20-D features for NS scales for each band as described in 2.1 and 2.2. Since our method is training based, the database needs to be partitioned into the training and test sets. The training set is used to train the regression model and the test is used to evaluate the performance of algorithm. Training and testing were implemented on each band dataset respectively. In our trial about 60% of the samples and their associated human subjective score are used to train and the remaining 40% used to test. The training and test set are disjoint and do not share content.

Therefore, our algorithm is independent of content and specific image degradation types. The performance indices contain the Person (Linear) Correlation Coefficient (LCC) and the Sum of Absolute Difference (SAD) between the predicted quality score and the subjective score provided by database. The better correlation with human perception means a smaller number for SAD and a value close to 1 for LCC.

NS	Num. of F	Train (%)	log₂C	LCC	SAD
3	60	75	20	0.7297(b1)	8.3924
				0.6970(b2)	8.4844
				0.7643(b3)	6.6389
				0.6924(b4)	7.5321
4	80	75	16	0.7909	7.0819
				0.7344	7.2698
				0.8203	5.3182
				0.8655	4.9128
5	100	75	18	0.8068	6.3966
				0.8166	6.0844
				0.9016	4.3893
				0.9212	3.9949
3	60	70	16	0.6539	9.9776
				0.6331	10.2159
				0.7370	7.6143
				0.5979	10.4009
4	80	70	13	0.7468	8.4356
				0.7856	8.9051
				0.7201	6.8098
				0.8400	5.9063
5	100	70	11	0.7283	8.4868
				0.7721	7.3881
				0.8239	5.7810
				0.9019	4.5539
3	60	65	16	0.4868	12.8444
				0.6120	10.4912
				0.6621	9.6193
				0.5979	10.8408
4	80	65	10	0.6184	10.4838
				0.6619	9.3536
				0.6514	8.3649
				0.7682	7.3119
5	100	65	12	0.6762	9.5182
				0.7217	8.6287
				0.8104	7.3883
				0.8724	5.2684

Table 2. LCC and SAD results for different test setup

Table 2 shows the results of LCC and SAD for different test setup. 'NS' is the scale numbers for multi scale feature extraction. 'Num. of F' means the feature numbers extracted for every test. 'Train (%)' is the proportion of training samples to the total number of samples. 'Log2C' is the logarithm of C, the parameter of $v - SVM$ and has different values according to the result of 5-fold cross validation. γ_s =0.0625 for all the tests. In LCC and SAD columns in table 2, there are four numbers which are values for band 1-4 respectively.

We designed three sets of experiments on different size training sets. The first experiment uses 75% of 217 images as training set and the remaining as test set by randomly selection, then the second and third uses 70% and 65% training samples. It shows that the regression accuracy increases with the increase of the

number of training samples. For NS=3, the correlation coefficient for 65%,70% and 75% samples participating in training of band 1 is 0.4868, 0.6539 and 0.7297. For a certain training and test the prediction accuracy increases with the increases of the number of scale layers. The high dimensionality of the features brings the rise of precision but also makes the algorithm time become longer. Also we note that the quality of band 4 has always been higher than that of other bands which is consistent with our scoring rules used in subjective score acquisition process. For NS=5, Train(%)=75 the linear correlation coefficient and sum of absolute difference reach the optimal values in the experiments. The LCC (0.8068 0.8166 0.9016 0.9212) values show that the algorithm matches well with human subjective opinions of image quality.

We believe that the current approach may not be ideal. The algorithm should also be widely compared with other methods. Future work will involve finding more effective quality indices according to the characteristics of RS image in urban area. Under our framework more experiments should also be done to make comparisons.

4. CONCLUSION

In this paper, we proposed a general-purpose image quality assessment method for RS image of urban areas. This is achieved by using two types of image features in multi scales. One is from the shape of histogram the other is from the natural scene statistics. A 20-D feature vector for each scale is extracted and is assumed to capture the RS image quality degradation caused mainly by cloud and mist. We use SVM to learn to predict image quality scores from these features. And the results show the correlation of the predicted scores and the subjective perception.

ACKNOWLEDGEMENTS

REFERENCES

Bahrami K, Kot A C, 2014. A fast approach for no-reference image sharpness assessment based on maximum local variation. *Signal Processing Letters, IEEE*, 21(6), pp. 751-755.

Chang, ChihChung, Lin, & ChihJen, 2011. Libsvm: a library for support vector machines. *Acm Transactions on Intelligent Systems & Technology*, 2(3), pp. 389-396.

Chen Shiping, 2011. Some issues about space remote sensing (In Chinese). *Spacecraft recovery & remote sensing*, 32(3), 1-8.

Fiete R D, Tantalo T, 2001. Comparison of SNR image quality metrics for remote sensing systems. *Optical Engineering*, 40(4), pp.574-585.

Ferzli R, Karam L J, 2009. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). *Image Processing, IEEE Transactions on*, 18(4), pp. 717-728.

Gu K, Zhai G, Yang X, et al, 2014. Deep learning network for blind image quality assessment//Image Processing (ICIP), *IEEE International Conference on*. IEEE, pp. 511-515.

Gu K, Zhai G, Lin W, et al, 2016. The analysis of image contrast: From quality assessment to automatic enhancement. *Cybernetics, IEEE Transactions on*, 46(1): 284-297.

Irvine J M, 1997. National imagery interpretability rating scales (NIIRS): overview and methodology//Optical Science, Engineering and Instrumentation'97. International Society for Optics and Photonics, pp. 93-103.

Kang L, Ye P, Li Y, et al, 2014. Convolutional neural networks for no-reference image quality assessment. *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pp. 1733-1740

Ma Xudong, 2014. *Research on quality assessment of image compression and image fusion for optical remote sensing image*. Doctoral dissertation, Wuhan University.

Mittal A, Moorthy A K, Bovik A C, 2012. No-reference image quality assessment in the spatial domain. *Image Processing, IEEE Transactions on*, 21(12), pp. 4695-4708

Mittal, A., Soundararajan, R., & Bovik, A. C, 2013. Making a "completely blind" image quality analyzer. *Signal Processing Letters, IEEE*, 20(3), pp. 209-212.

Moorthy, A. K., & Bovik, A. C, 2010a. A two-step framework for constructing blind image quality indices. *Signal Processing Letters, IEEE*, 17(5), pp. 513-516.

Moorthy, A. K., & Bovik, A. C, 2010b. Statistics of natural image distortions. *In Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 962-965.

Narvekar N D, Karam L J, 2011. A no-reference image blur metric based on the cumulative probability of blur detection (CPBD). *Image Processing, IEEE Transactions on*, 20(9), pp. 2678-2683

Pei S C, Chen L H, 2015. Image quality assessment using human visual DOG model fused with random forest. *Image Processing, IEEE Transactions on*, 24(11), pp. 3282-3292.

Sharifi A, Leon-Garcia A, 1995. Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video. *IEEE Trans. Circuits Syst. Video Technol.*, 5(1), pp. 52-56.

Schölkopf, B., Smola, A., Williamson, R., & Bartlett, P, 2000. New support vector algorithms. *Neural Computation*, 12(5), pp. 1207-1245.

Wang Rongbin, 2014. *Comprehensive evaluation and adaptive restoration of remotely sensed images (In Chinese)*. Doctoral dissertation, Wuhan University.

Zhang Xia, Zhang Bing, Zhao Yongchao, Tong Qiongxi, Zheng Lanfen, 2002. Image quality assessment for the infrared multi-spectral scanner of the Chinese-Brazil Earth Resources Satellite. *Journal of image and graphics*, 7(6), pp. 581-586.