

AN EFFECTIVE PERFORMANCE RANKING MECHANISM TO IMAGE DEHAZING METHODS WITH PSYCHOLOGICAL INFERENCE BENCHMARK

Zi'ang Hu^{1,2}, Qingsong Zhu^{1,3*}

¹Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China,

²South China Agricultural University, Guangzhou, China,

³The Chinese University of Hong Kong, Hong Kong, China

ABSTRACT

In this paper, we proposed a novel quality assessment framework for the performance ranking of image dehazing algorithms by employing prior features and radial basis function (RBF-based) classifier. First, we formulate the evaluation problem within a novel comparison framework by using classification methods. Second, prior information is used to extract the inherently feature of hazy image and these feature values are normalized with psychological inference benchmark (*PIB*) to eliminate the cognitive bias of individuals. Finally, a cost-compensation classification network is cyclically utilized to rank the image dehazing performance until the iteration ends and updated the *PIB* in every loop. Experiments show that the proposed method is more effective for evaluating the image dehazing performance than the existing blind image quality assessment methods, and the evaluation results correlate well with human judgments of image quality.

Index Terms— Image dehazing, no-reference image quality assessment, *PIB*, prior feature, classification

1. INTRODUCTION

Along with the advance of haze removal techniques in the past few years, how to effectively compare the performance of image dehazing becomes a novel task. The current assessment method of image dehazing performance is mainly based on subjective evaluation [1-3], which is strongly affected by the cognition of individuals. This could cause the cognitive bias between individuals and it is not always suitable for applications. Consequently, the proper quantitative assessment methods should be studied to mimic the perceive ability of human and to evaluate images automatically [4].

Image haze removal is inherently an ill-posed problem, because a haze-free reference image in the same scene is often not available. Hence, we should employ the no reference image quality assessment (NR-IQA/ Blind-IQA) methods for evaluating image dehazing performance. The most existing approaches of NR-IQA are to use a single indicator to calculate

This work was supported by the National Natural Science Foundation of China under 61303166. **Corresponding author:** qs.zhu@siat.ac.cn

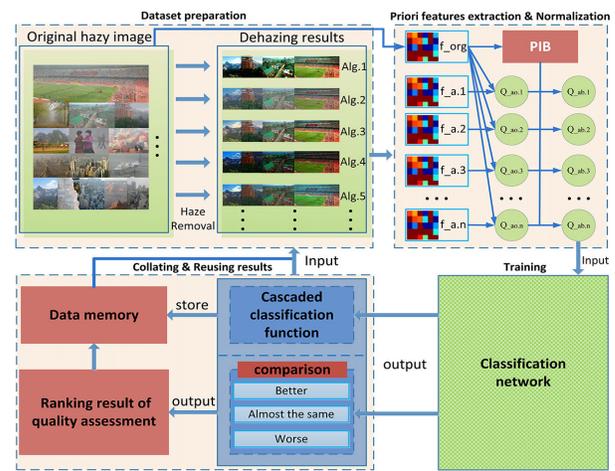


Fig. 1. The framework of the proposed approach.

an absolute quality score for image, such as BRISQUE [6], BLINDS [7], CORNIA [8], etc. Moreover, some researchers employ full reference image quality assessment metrics to quantify the visual quality of the dehazed image comparing with the original hazy image [10,11].

Recently, machine learning technology has been applied in image quality assessment. Most of the state-of-the-art frameworks follow two steps: global feature extraction and model regression by human scores [12]. Mittal [6] trained a regression-based support vector machine (SVM) for IQA based on the locally normalized luminance coefficients in the spatial domain. Chen [14] proposed a framework for comparing the different image enhancement algorithms by using rank-SVM. In [16], Kang described a convolutional-neural-network (CNN) algorithm, which combines feature learning and regression to predict image quality score. They claim that the modified network structure is more accurate than traditional NR-IQA methods. However, global feature extraction is time consuming and labeling an image with numerical score is consequentially affected by individual experience and background so that it is ambiguous.

Nowadays, there is no a commonly acceptable method for evaluating image dehazing performance [17]. This task

faces with several challenges. First, it is important to extract the truly intrinsically salient features to describe hazy images and distinguish a hazy image from natural haze-free image. Second, comparing single indicator scores and utilizing regression-based prediction model are not consistent with human visual perceptual mechanism. Third, there is no public quality assessment benchmark dataset for image dehazing. Therefore, improving the evaluation method of image dehazing performance needs to be solved and it will benefit many image understanding and computer vision applications such as haze removal algorithm [1,3], video segmentation [5,9] and image matting and classification [13,15,18], etc.

In this paper, we present a cost-compensation classification framework with psychological inference benchmark (*PIB*) for evaluating the image dehazing performance. We find that comparison or ranking mechanism of evaluation procedure is more close to a classification process rather than just predicting an absolute quality score. In this framework, prior information is used to extract the inherently features of hazy image firstly. And then, we compared these prior features with *PIB* to eliminate the cognitive bias of individuals. Finally, training a 3-grade classification network circularly with normalized features until the iteration ends and updating the *PIB* in every loop. The main contribution of our method is to formulate the evaluation of image dehazing problem as a comparison framework by using classification methods, rather than previous prediction-based approaches for IQA.

2. PROPOSED METHOD

2.1. Framework formulation

Image dehazing performance evaluation can be converted to a classification learning problem. Let $Q = \{Q_{ab.n_i}, n = 1, \dots, N\}$ denote the set of input features and $E = \{e_i, i = 1, \dots, M\}$ is the label set (ranking result). Then, the training set is given by $X = \{Q_{n_i}, e_{n_i}\}_{n=1}^N$ with $Q_{n_i} \in \mathbb{R}^N$. Where n stands for the different priori features (Section 2.1), M is the number of input dehazing algorithms. This classification problem can be formulated as the estimation of the rank labels by solving a convex optimization problem.

$$F = \{f : Q \in \mathbb{R}^n \mapsto E\} \quad (1)$$

The optimization problem is given by $\min_i f(Q_{n_i})$, *st.* $h_j(Q_{n_i}) = 0, j = 1, \dots, p, g_k(Q_{n_i}) \leq 0, k = 1, \dots, q, Q_{n_i} \in X \subseteq \mathbb{R}^N$. Where p and q are the number of constraint. In this formulation, the goal is to learn a decision function f from training set X to classify the candidate algorithms in rank. In this paper, we focus on the framework of dehazing quality assessment method yet specific classifier.

Fig.1 shows the flowchart of the proposed framework. Where $f_{.a.n}$ denote the extracted priori feature value (Section 2.1) of different dehazing algorithms, $f_{.org}$ is for the priori feature value which extracted from original hazy image. $Q_{.ao.n}$ is the comparison value which $f_{.a.n}$ divided by

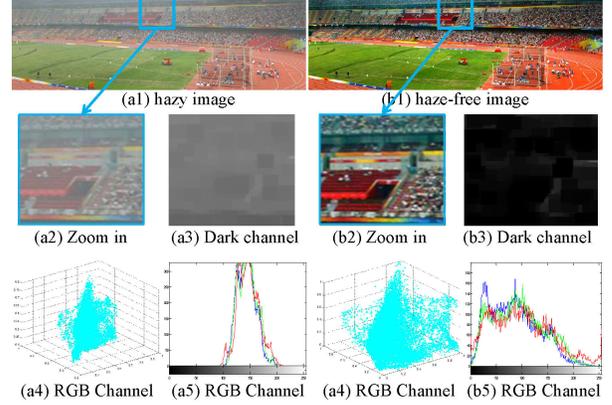


Fig. 2. Comparison of prior features.

$f_{.org}$, $Q_{.ab.n}$ denote the $Q_{.ao.n}$ compared with *PIB* for normalization. Input the beforehand extracted prior features into classification network, and output the ranking result of compared dehazing performance with classification function.

2.2. Framework components

• **The prior features extraction.** Prior feature [19] contains the intrinsic information of degraded image. It can be used to distinguish a hazy image from natural haze-free image and compare the image dehazing performance. Extracting the salient prior features from hazy image will contribute to improve the classification performance from both accuracy and efficiency. Generally, lower prior feature values correspond to better dehazing performance in our case.

Our classification-based network is founded on perceptually relevant prior features, including *dark channel sparseness*, *dispersion of color channels* and *textural similarity*. These features are extracted from local image patches to capture the intrinsic statistics of natural images. Fig.2 illustrates the concept comparison of prior features from a sample window Ω . There are two images of the same scene under different weather conditions. Fig.2(a) shows the hazy group and Fig.2(b) exhibits the haze-free group. A sample window Ω is located on the same place in each image. The prior features of each zoom-in patch are illustrated as follow.

Fig.2(a3),(b3) indicate the dark channel [1] of images. He [1] claims that low-intensity pixels in at least one color channel often exist in most of the local regions. For any local patch $\Omega(x)$ and any color channel $\tau \in \mathbb{R}\{r, g, b\}$, the dark channel J^{dark} of any image J is given by:

$$J^{dark}(x) = \min_{y \in \Omega(x)} \left(\min_{\tau \in \mathbb{R}^3} J^\tau(y) \right) \quad (2)$$

The dark channel map of a haze-free image Fig. 2(b3) is visually darker than the one of a hazy image Fig.2(a3). Thus, we could utilize an appropriate threshold θ_0 to build the sparse matrix of dark channel. If the dark channel value of a local patch is lower than θ_0 , it could be redefined as zero.

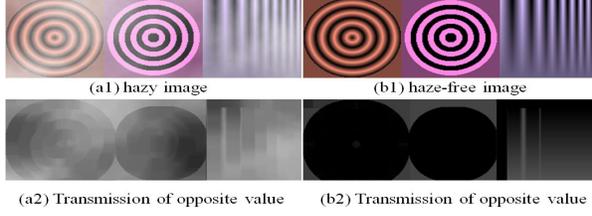


Fig. 3. Comparison of artificial images and their transmission in hazy and haze-free condition.

In practice, assign value zero to every $J^{dark} - \theta_0 < 0$ element, and then update the redefined dark channel matrix as \hat{J}^{dark} . Consequently, to properly extract the prior feature of dark channel sparseness D_s , we can simply calculate ratio of non-zero elements in \hat{J}^{dark} . It is defined as follow:

$$D_s(\tau, X, \Omega) = \frac{\sum_{\Omega} \|\hat{J}^{dark}(x)\|_0}{I_{h \times w}}, x = 1, 2, \dots, X \quad (3)$$

Where $\|\hat{J}^{dark}(x)\|_0$ is the L0-norm of $\hat{J}^{dark}(x)$, $I_{h \times w}$ is the size of image and $x \in X$ is each pixel in the image. Hereinafter, we denote $D_s(\tau, X, \Omega)$ simply as D_s . Theoretically, the D_s of haze-free image is much sparser than the D_s of hazy image.

Fig.2(a4),(b4) present the color (RGB) channels 3D scatterers consisting of all pixels in the sample window Ω . It reveals the spatial distribution of the colors. Generally speaking, the haze-free image is associated with a higher dynamic range. The degree of dispersion can be used as prior feature descriptor. Since the values of RGB channel is statistically independent, the dispersion of color channels D_σ is given by:

$$D_\sigma = \frac{1}{3} \sum_{\sigma_\tau} \sqrt{\frac{1}{I_{h \times w}} \sum_{x=1}^{I_{h \times w}} (J_\tau(x) - \mu_\tau)^2} \quad (4)$$

Where σ_τ is the standard deviation of each color channel and μ_τ is the mean value of each channel.

Fig.2(a5),(b5) show the color histogram of the zoom-in patch. Compared with the haze-free image, it is obvious that the color histogram of hazy image is more centralized and skewing towards the bright side holistically.

In addition, texture is an intrinsic attribute of natural image. Fig.3 illustrates two artificial images containing particular texture patterns from the TID database [20] and their redefined transmission. For easier observation, Fig.3(a2),(b2) show the redefined transmission which the value $\bar{t}(x)$ is redefined as opposite value ($\bar{t}(x) = 1 - t(x) \propto J^{dark}$). We find that, in most of the local patch which texture adhere to the surface, the texture detail in transmission of haze-free image Fig.3(a2) is barely to see. However, the surface texture is partially remained in transmission of hazy image Fig.3(b2). The reason lies on the fact that transmission t is just related to the scene depth d (expressed as $t(x) = e^{-\beta d(x)}$, where β is a constant) and it should not be influenced by any surface

texture. Therefore, the textural information of transmission can be regarded as a prior feature to distinguish dehazing performance. In our case, we choose a grey level co-occurrence matrix (GLCM) of the transmission to be a prior feature and employ the correlation D_c , homogeneity D_h and contrast D_e of GLCM as hue-shifting and textural features.

• **Psychological inference benchmark (PIB).** The new psychological evidence shows that human prefers to evaluate images by comparing candidate with benchmark rather than assess with numerical scores[21]. These comparing/ranking results of assessment tend to qualitative evaluation, such as better, worse or almost the same. Hence, the logic of existing learning-based evaluation methods, which totally depend on regression-prediction models, is clearly different from human visual perceptual mechanism and thus unreliable. Following [21], we believe that human visual assessment is related to psychological activity of individuals. The candidate image should be compared with a benchmark in reality. Evaluating image dehazing performance in the absence of reference image, we are subconsciously compared the candidate image with our psychological expectation in fact. Hereinafter, we call the psychological expectation as psychological inference benchmark (PIB). It is not hard to imagine that different observer will not always agree with the same PIB, but it does not determine the final evaluation results (proved in sect. 3.3). In practice, PIB can be appointed and also can be learnt from the learning-network.

• **Training network.** As an example, a cost-compensation classifier [22] is cyclically used to output the ranking results until the iteration ends. The penalty weight of every empirical Hinge Loss term in loss function is very sensitive to the cost of top sequence and it is compensated by empirical value. In the implementation, we utilize the LIBSVM package to implement the SVC with a radial basis function (RBF) kernel.

2.3. Monotonicity of prior feature

Note that we proved the prior features are valid for distinguishing hazy image from natural haze-free image in sect. 2.2, but it does not mean that value of extracted prior feature is consistently keep the same in variation trend with the degradation increasing. In this case, we define a sign function $sgn(n)$ to verify the monotonicity of prior feature for further study. Let $H = [h_i, i = 1, \dots, D]_{1 \times D}$ denote the degree of haze density and $V = [V_{n_i}, i = 1, \dots, D]_{n \times D}$ represent the value of each prior feature in different hazy density. For any $i < j$, define a decision function as:

$U = \Delta V \cdot \Delta H^T |_{n \times 1}; \Delta H = h_j - h_i > 0; \Delta V = v_j - v_i$
Then the sign function can be written as:

$$sgn(i, j) = \begin{cases} 1, & U > 0 \\ -1, & U < 0 \\ 0, & else \end{cases} \quad \forall 0 < i < j \quad (5)$$

If $sgn(i, j) = 1$, the series of extracted feature value is monotonic. Thus, it can be employed as potential prior feature. Else, the candidate feature is not available.

3. EXPERIMENT

3.1. Data and experimental protocol

Experiments are done using 120 most popular hazy images in previous research and 4 classic dehazing algorithms [23][1][24][25] are chosen to generate the dehazed dataset. As the inputs of our training network, we normalized the feature vector with *PIB* and regarded the results of algorithm [1] as *PIB* for simplifying the calculation. Label set is obtained by comparing different dehazing performance in pair and using sorting method to rank the dehazing performance of candidate algorithms. The threshold value θ_0 (sect.2.2) is set as 64 because the first 64 gray-scale is pretty close to the dark channel intensity according to the Mach Band [26].

3.2. Performance evaluation

In order to evaluate the performance of proposed method, we use correlation coefficients (SROCC, KROCC, PLCC, shown in Tab.1) to calculate the correlation between objective scores (output results) and the subjective opinion scores (labels). Table.1 presents the average correlation coefficients with 1000 iterations of blind IQA methods test on the database described in sect.3.1. Compared with previous works, our method performs better than the previous methods [6][7][8][27][17].

Table 1. Performance comparison of the blind IQA methods

Blind IQA method	SROCC	KROCC	PLCC
BRISQUE [6]	0.3877	0.3126	0.3721
DIVINE [27]	0.5295	0.5078	0.5138
BLINDS-II [7]	0.4526	0.3967	0.4469
CORNIA-II[8]	0.4933	0.4072	0.4638
Ma[17]	0.6259	0.6017	0.6435
Ours	0.7371	0.6733	0.7041

3.3. Effects of *PIB* on accuracy and iteration times

Whether the evaluation method with *PIB* really works and what effects did the *PIB* bring to our framework? To figure out this question, we did some experiments as follow.

- **With *PIB* and without *PIB*.** The major difference between evaluation methods with *PIB* and without *PIB* is the input label format. Some comparing experiments are designed on the same dataset described in sect.3.1. Following [28], we employ a ranking SVM to directly predict the ranks of test set. The ranking SVM input labels which value from 1 to the number of candidate algorithms. In contrast, the labels of our method with *PIB* are ranged from 1 to 3, which stand for simple semantic: better, worse and almost the same. In practice, we calculate the SROCC between predicted ranks of former method and labels. The highest SROCC of the ranking SVM method is 0.3984 which is lower than the average performance of the proposed method. We find the accuracy of a multi-class classification method will be rapidly decreased with the increasing of candidate algorithms.

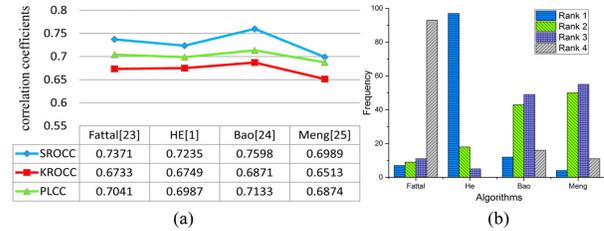


Fig. 4. Performance evaluation of proposed method using different *PIB* and performance comparison of some different image dehazing algorithms.

- **Effect on accuracy and iteration times.** We choose the results of each candidate dehazing algorithms as *PIB*, and evaluate the performance using correlation coefficients. As shown in Fig.4 (a), the correlation coefficients should keep in the same level with the changing of different *PIB*. So the choice of *PIB* will neither influence the accuracy of evaluation nor determine the final ranking result. But our method could not give the final ranking result after first iteration when the number of candidate dehazing algorithms is more than three. We circularly utilize the same structure of classifier and update the *PIB* in every loop. If M is the number of candidate dehazing algorithms, then it is not hard to prove that the maximum iteration time is $M-1$. The iteration times will increase linearly with the increasing number of candidate algorithms.

3.4. Image dehazing performance comparison

Since the evaluation method is established hereinbefore, we test the average performance of each candidate dehazing algorithm on the dataset described in sect.3.1. In Fig.4 (b), we employ a frequency histogram to illustrate the rank results of each candidate algorithm. It can be seen that He[1] ranks in the first place and Fattal[23] came in the last upon most occasions, the performance of other two algorithms is very close. It comes to a conclusion that, on average, the dehazing performance in Fig.4 (b) is ordered in He[1]>Meng[25] > Bao[24]> Fattal[23].

4. CONCLUSION AND FUTURE WORK

This paper presents a novel framework employing prior features and RBF-based classifier to rank the performance of different dehazing algorithms. Experiments show that the method is able to evaluate the image dehazing performance and rank them in order successfully. The framework is quite simple but effective, and the evaluation results correlate well with human judgments of visual quality. Our method have some limitations as well. For instance, 1) The hazy image dataset does not include large bright regions which usually exists in natural scene. 2) The classifier can be improved to give the final result by only one iteration. Therefore, the future improvements of the method will deal with some more effective prior that can accurately handle bright regions as well as improve the structure of classifier.

5. REFERENCES

- [1] K. He, J. Sun and X. Tang, Single image haze removal using dark channel prior, in *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33: pp. 598-605, 2011.
- [2] Q. Yan, L. Xu, J. Jia, Dense scattering layer removal, in *Proc. of SIGGRAPH Asia*, 2013.
- [3] Q. Zhu, J. Mai, L. Shao, A fast single image haze removal algorithm using color attenuation prior, in *IEEE Trans. Image Process*, vol. 24 (11), pp. 3522-3533, 2015.
- [4] W. Hou, X. Gao, D. Tao, et al, Blind image quality assessment via deep learning, in *IEEE Trans. Neural Networks and Learning Systems*, vol. 26 (6), pp. 1275-1286, 2015.
- [5] Q. Zhu, Z. Song, Y. Xie, L. Wang, A novel recursive bayesian learning-based method for the efficient and accurate segmentation of video with dynamic background, in *IEEE Trans. Image Processing*, vol. 21 (9): pp. 3865-3876, 2012.
- [6] A. Mittal, A. Krishna and A. Bovik, No-reference image quality assessment in the spatial domain, in *IEEE Trans. Image Processing*, vol. 21: pp. 4695-4708, 2012.
- [7] M. Saad, A. Bovik and C. Charier, Blind image quality assessment: a natural scene statistics approach in the DCT domain, in *IEEE Trans. Image Processing*, vol. 21: pp. 3339-3352, 2012.
- [8] P. Ye, J. Kumar, L. Kang and D. Doermann, Real-time no-reference image quality assessment based on filter learning, in *Proc. of CVPR*, 2013.
- [9] Q. Zhu, L. Shao, Q. Li, Y. Xie, Recursive kernel density estimation for modeling the background and segmenting moving objects, in *Proc. of ICASSP*, 2013.
- [10] Y. Wang and C. Fan, Single image defogging by multi-scale depth fusion, in *IEEE Trans. Image Processing*, vol. 23: pp. 4826-4837, 2014.
- [11] L. Caraffa, J.P. Tarel, Markov random field model for single image defogging, in *Proc. of Intelligent Vehicles Symposium (IV)*, 2013.
- [12] W. Xue, L. Zhang and X. Mou, Learning without human scores for blind image quality assessment, in *Proc. of CVPR*, 2013.
- [13] Q. Zhu, L. Shao, X. Li, L. Wang, Targeting accurate object extraction from an image: a comprehensive study of natural image matting, in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26(2): pp. 185-207, 2015.
- [14] Z. Chen, T. Jiang and Y. Tian, Quality assessment for comparing image enhancement algorithms, in *Proc. of CVPR*, 2014.
- [15] Z. Zhang, Q. Zhu, Y. Xie, Learning based alpha matting using support vector regression, in *Proc. of ICIP*, 2014.
- [16] L. Kang, P. Ye, Y. Li, et al, Convolutional neural networks for no-reference image quality assessment, in *Proc. of CVPR*, 2014.
- [17] K. Ma, W. Liu and Z. Wang, Perceptual evaluation of single image dehazing algorithms, in *Proc. of ICIP*, 2015.
- [18] Q. Zhu, Z. Zhang, Z. Song, et al, A novel nonlinear regression approach for efficient and accurate image matting, in *Signal Process. Lett.*, vol. 20(11): pp. 1078-1081, 2013.
- [19] T. Wu, M. Bae, M. Zhang, et al, A prior feature SVM-MRF based method for mouse brain segmentation, in *NeuroImage.*, vol. 59(3): pp. 2298-2306, 2012.
- [20] N. Ponomarenko, L. Jin, O. Ieremeiev, et al, Image database TID2013: Peculiarities, results and perspectives, in *IEEE Trans. Signal Process.*, vol. 30: pp. 57-77, 2015.
- [21] M. H. Waugh, Mentors, Muses, and Memories: Personal narratives from psychological assessment, *Journal of Constructivist Psychology*, 2015.
- [22] Y. Li, J.T. Kwok, Z. Zhou, Cost-sensitive semi-supervised support vector machine, in *Proc. of AAI*, 2010.
- [23] R. Fattal, Single image dehazing, *ACM Trans. Graphics*, vol. 34, pp 1-14, 2008
- [24] L. Bao, Y. Song, Q. Yang, et al, An edge-preserving filtering framework for visibility restoration, in *Proc. of ICPR*, 2012.
- [25] G. Meng, Y. Wang, J. Duan, et al, Efficient image dehazing with boundary constraint and contextual regularization, in *Proc. of ICCV*, 2013.
- [26] https://en.wikipedia.org/wiki/Mach_bands. September 23/ 2015
- [27] A.K. Moorthy and A.C. Bovik, Blind image quality assessment: From natural scene statistics to perceptual quality, in *IEEE Trans. Image Processing*, vol. 20: pp. 3350-3364, 2011.
- [28] Y. Cao, J. Xu T. Liu, et al, Adapting ranking SVM to document retrieval, in *ACM Proc. of SIGIR*, 2006.