Haze Image Database and Preliminary Assessments

Ying Chu*, Zhiyuan Chen, Yu Fu and Hengyong Yu

***Abstract*– A database named HID2016 is built to provide a public benchmark to blindly evaluate image quality of haze images. The image set comes from a photographic exhibition held in China in 2014, and the single-stimulus method is adopted to develop the subjective evaluation experiments on those images. The contribution of this database not only lies in the fact that HID2016 is the first image database aiming at evaluating quality of haze images, but also lies in that the environmental indexes of each haze picture are provided, including air quality index (AQI), fine particulate matter (PM2.5) index and inhalable particles (PM10) index, which can help to explore the relationship between image quality and air quality. A three-order polynomial fitting function is adopted to simulate the experimental data, aiming at revealing the relationship between subjective score and environmental index. The experimental results show that the subjective scores in the HID2016 database keep a good accordance with the visual quality of haze images, and they roughly negative correlate with the air quality.**

***Keywords*—Haze image; subjective evaluation; mean opinion score; image database; image quality assessment.**

I. Introduction

In 2015, a haze documentary named "Under the Dome" was watched million times on the internet and it provoked a national discussion after going viral online. The documentary has sparked strong concerns on the haze problem by people in China. As a matter of fact, not only the healthy people but also our national security will be affected by the haze problem. For instance, during fog and haze weathers, it is particularly difficult for surveillance cameras to successfully capture clear images, which would in turn reduce the reliability of monitoring systems.

Although alternative solutions (e.g. infrared imaging) have been proposed, they only work under a limited extent in smog weathers. Besides, the amount and density of suspending particles in the air are quite large in haze weathers. Thus, various particles block the lights like a brick wall, which leads to vague pictures in the end. Studies on how to improve haze image quality through image post-processing have been developed in recent years [1-5].

Due to the absence of objective image quality assessment (IQA) index designed for haze images, most of the dehazing algorithms compared their results in a subjective way. This dramatically cut down the accuracy and reliability of comparison results [6]. To solve the problem, Fang *et al.* proposed an IQA metric based on a degree of contrast enhancement and structure similarity between the haze and dehazed images [7], which belongs to the category of full-reference IQA. For a haze image, the problem is that there is no way to acquire the original "perfect" reference image, and the quality evaluation should be in a no-reference way. Li *et al.* [8] and Guo *et al.* [9] used contrast enhancement assessment methods based on gradient ratioing at visible edges proposed by Nicolas *et al.* [10]. However, the suitable targets for this kind of methods are fog images. Strictly speaking, they cannot be applied to haze images. To conclude, state-of-art studies on haze images focus on how to improve visual quality instead of how to evaluate image quality. Besides, there are no public image databases of haze images, resulting the absence of generally accepted criterion for algorithm performance comparison in this field [7]. To shorten this gap, in this project, we build a haze image database (HID) named HID2016 and perform a preliminary study to demonstrate its usefulness.

The remainder of this paper is organized as follows. In Section II, the image contents in the database HID 2016 are introduced. In Section III, the details of the subjective experiments are described. Preliminary results and analysis are provided in Section IV and conclusions are made in Section V.

II. HID 2016 Image Contents

2.A. Image Set

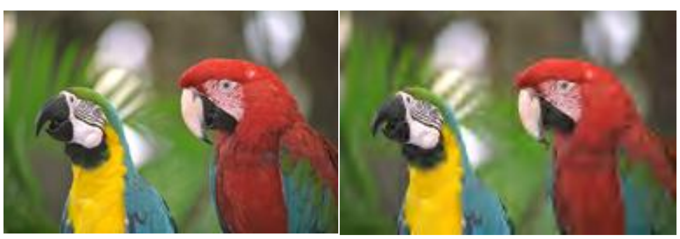
The details of most widely used subjective IQA databases in the world and the developed HID2016 are summarized in Table I. Taking the LIVE database as an example, there are five types of distortion, including JPEG2000 compression, JPEG compression, white Gaussian noise, Gaussian blur and fast fading Rayleigh. Distorted family of different degradation degree was generated from natural images with perfect visual quality, and subjective mean opinion score (MOS) or difference MOS (DMOS) were calculated through subjective evaluation experiments. Fig. 1 gives a representative example of one original image and its distorted counterparts, as well as their DMOS values in the LIVE database.

Although haze images look like distorted (e.g. blurred image), they are directly taken by cameras and that they are not products of any degradation process. Consequently, in Table I, there is no information about original or degraded images for the HID2016 database. It is not difficult to understand that image quality evaluation on haze images is under the category of blind image quality assessment. For this case, we only need to collect haze images and directly perform subjective experiments on them. No original images with perfect visual quality exist, nor are they needed in our situation.

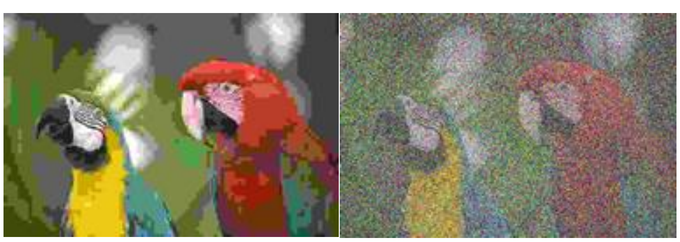
*This work was supported in part by Guangdong Province Ordinary University Characteristic Innovation Project (2015KTSCX122) and Shenzhen Science and Technology Plan of Basic Research Project (JCYJ20140509172609166). Y. Chu, Z.Y. Chen and Y. Fu are with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, Guangdong 518060, China. H.Y. Yu is with the Department of Electrical and Computer Engineering, University of Massachusetts Lowell, Lowell, MA 01854, USA. Corresponding author: Y. Chu (e-mail: chuying@szu.edu.cn).

TABLE I. BASIC INFORMATION IN DIFFERENT SUBJECTIVE IMAGE QUALITY EVALUATION DATABASES.

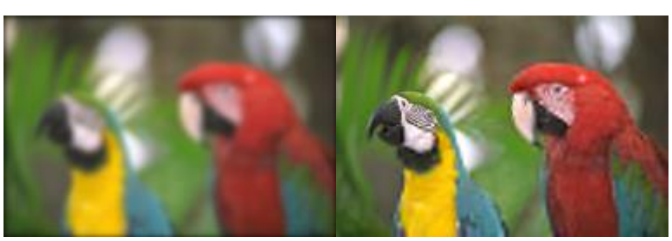
Database	Number of original image	Resolution	Original image type	Number of distortion type	Number of degradation degree	Number of distorted images	Subjective scores	Subjective rating range
LIVE[11]	29	various	24 bit RGB	5	4~5	779	DMOS	[-2.6400, 111.7747]
CSIQ[12]	30	512×512	24 bit RGB	6	4~5	866	DMOS	[0, 1]
TID2008[13]	25	512×384	24 bit RGB	17	4	1700	MOS	[0, 7.7143]
TID2013[14]	25	512×384	24 bit RGB	24	5	3000	MOS	[0.2424, 7.7143]
A57[15]	3	512×512	8 bit Gray	6	3	54	DMOS	[0.0890, 1]
MICT[16]	14	768×512	24 bit RGB	2	6	168	MOS	[1, 4.8800]
IVC[17]	10	512×512	24 bit RGB	5	5	185	MOS	[1, 4.8846]
IVC Art[18]	8	512×512	24 bit RGB	3	5	120	MOS	[0.9500, 4.7500]
HID2016	Not applicable	various	Not applicable	Not applicable	Not applicable	301	MOS	[1, 5]



(a) Original image, DMOS=0 (b) JPEG 2000, DMOS=74.9583



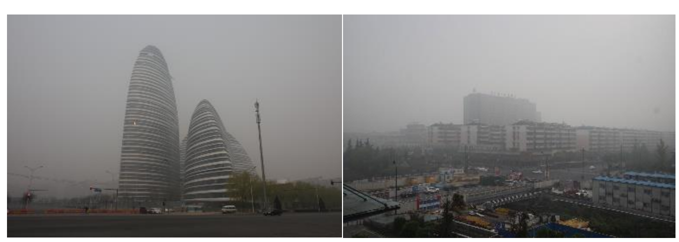
(c) JPEG, DMOS=87.7808 (d) White noise, DMOS=99.6615



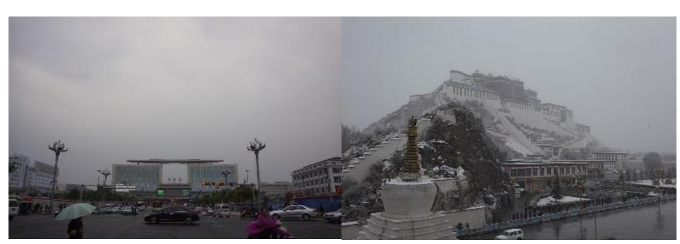
(e) Gaussian blur, DMOS=79.7590 (f) Fast fading, DMOS=58.4194

Fig. 1. Example of original and degraded images in the LIVE database.

In 2014, a photographic exhibition named "Breathing in China" was held by a Chinese website called April Winds [19]. Photos for a constant scene in different dates were recorded by photographers in some main cities in China. Specifically speaking, they were recorded in quite different weathers with various levels of haze. We contacted one of the principals of the website and some of the photographers. In the end, the raw haze images of eight cities were provided to us. Representative haze images are shown in Fig. 2, and the numbers of pictures in each city are listed in Table II.



(a) Beijing, MOS=1.625 (b) Hangzhou, MOS=1.5



(c) Kunming, MOS=3.125 (d) Lasa, MOS=2.625

TABLE II. BASIC INFORMATION OF HAZE IMAGES TAKEN IN DIFFERENT CITIES IN THE HID 2016 DATABASE.

City	Beijing	Hangzhou	Kunming	Lasa	Shijiazhuang	Taiyuan	Tianjin	Wuhan	Total
Number of photos taken	41	40	41	40	41	34	41	30	308
Number of photos tested	41	40	41	38	38	33	41	29	301
Subjective rating range	[1.625, 5]	[1.5, 4.625]	[3.125, 4.813]	[2.625, 4.875]	[1.063, 2.25]	[2.25, 4.188]	[1, 4.813]	[1.563, 4.438]	[1, 5]

(e) Shijiazhuang, MOS=1.0625 (f) Taiyuan, MOS=2.25

(g) Tianjin, MOS=1 (h) Wuhan, MOS=1.5625

Fig. 2. Representative haze images in the HID2016 database.

2.B. Environmental Index

The contribution of HID2016 is that we also provide environmental indexes for each image, such as air quality index (AQI), fine particulate matter (PM2.5) index and inhalable particles (PM10) index, to explore the relationship between haze image quality and air quality. Taking Beijing as an example, the AQI, PM2.5 and PM10 values of 41 photos captured in 41 continuous days are shown in Fig. 3.

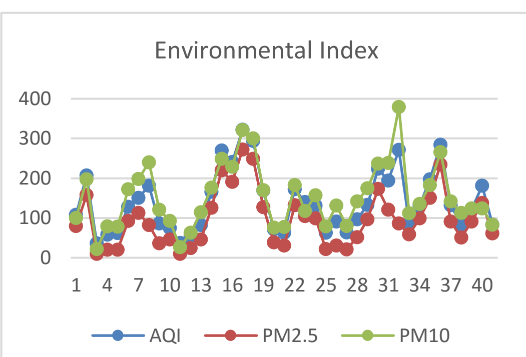


Fig. 3. Environmental indexes for pictures taken in Beijing. The lower the index values are, the better the air quality is.

In fact we downloaded historical weather data [20] for each of the picture in the database and performed some analysis on them. Detailed information about their relationship will be provided in section IV.

III. SUBJECTIVE EXPERIMENTS

To build a haze image database for blind IQA, we need to develop subjective experiments on image quality. Both the International Telecommunication Union-Radiocommunication (ITU-R) [21] and European Broadcasting Union (EBU) [22] released protocols on subjective assessment based on a psychology testing mechanism. In this paper we use the ITU-R_BT.500-11 protocol to build up the haze image subjective quality evaluation database.

3.A. Setup

The laptop model used for subjective testing is Sony VAIO VPCEG-211T with a preset resolution of 1366×786 and 100% luminance. The distance between subjects and the display screen was twice as the height of the laptop screen.

Due to photography angle deviation or image damage, seven captured photos are excluded from our experiments. The final total number of tested pictures is 301, and the number of tested image for each city is listed in Table II. Because the photos were taken by different photographers, the camera specifications and models are quite different. Some of the pictures were saved in raw format, and some others were not. Besides, the resolutions of the photos were not unified. To display them properly, all the photos were saved as bitmap format and were resized to fill the full screen of the laptop.

3.B. Subjects

There are sixteen subjects involved in the study. Among them, twelve testers are female and the other four are male. All of them are undergraduate students in Shenzhen University, and none of them is specialist in the image processing field. The testers range in age from 20 to 30 and their visual acuity and color vision are normal.

3.C. Test Method

Four test methods are recommended by the ITU-R_BT.500-11 protocol. Here we use the single-stimulus (SS) method. The testing process includes: 1) Provide test instructions to the subjects; 2) Show a single test image to them (no reference image provided); 3) Ask the subject to evaluate the test image; 4) Repeat steps 2 and 3. It is suggested by the protocol that the testing process should last no more than half an hour, and the testing images should be displayed in a random order. When the whole test procedure is over, the mean opinion score of each image is calculated.

Figure 4 shows the display process of the SS method, where T1 and T2 equal to ten seconds, and T3 lasts 3 seconds. No picture but gray background is shown during T1 and T3 periods, and a haze test image is shown during T2 period. Then, an evaluation score is given by the tester during T2 and T3 period.

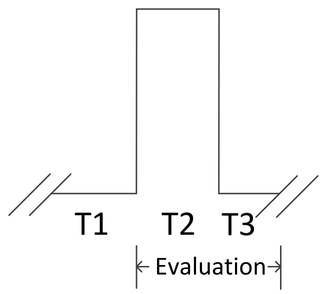


Fig. 4. Flowchart of the SS subjective test.

In the ITU-R_BT.500-11 protocol, there are five grades of evaluation indicator. Aiming at evaluating our special objects, i.e., haze images, new judgement criteria are added and listed in Table III.

TABLE III. ITU-R QUALITY, IMPAIRMENT SCALES AND SUPPLEMENTARY SCALES FOR HAZE IMAGE.

Scale	Quality	Impairment	Judgement
5	excellent	imperceptible	no haze
4	good	perceptible, but not annoying	almost no haze
3	fair	slightly annoying	slight haze
2	poor	annoying	moderate haze
1	bad	very annoying	heavy haze

IV. PRELIMINARY RESULTS AND ANALYSIS

As mentioned in Section 2.B, we are interested in exploring the relationship between subjective score and environmental index for haze images. As a preliminary study, here we simply use the degree of fitting between MOS values and AQI/PM2.5/PM10 values as a comparison indicator.

A three-order polynomial function is adopted as the fitting function. The calculated R-square coefficient is in range [0, 1]. The larger the value is, the better the fitting degree is. Taking Beijing as an example, the fitting curve for MOS and AQI index is illustrated in Fig. 5, and the comparison results of fitting degree for all the cities are listed in Table IV.

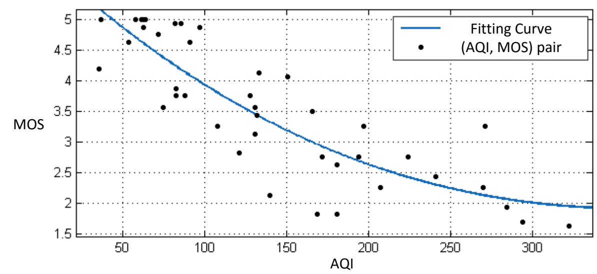


Fig. 5. Scatter plot of MOS vs. AQI for Beijing, using three-order polynomial function as fitting curve.

TABLE IV. COMPARISON RESULTS OF FITTING DEGREE BETWEEN SUBJECTIVE SCORE AND ENVIRONMENTAL INDEX.

City	AQI	Rank AQI	PM2.5	Rank PM2.5	PM10	Rank PM10
Beijing	0.7224	1	**0.7819**	1	0.5282	1
Hangzhou	0.4749	2	**0.4912**	4	0.2930	4
Kunming	0.3036	4	**0.4191**	5	0.2160	6
Lasa	0.0575	8	0.0481	8	**0.0887**	8
Shijiazhuang	0.1574	7	**0.1769**	7	0.0975	7
Taiyuan	0.2728	5	**0.5187**	2	0.2489	5
Tianjin	0.4358	3	**0.4972**	3	0.3498	2
Wuhan	0.2723	6	**0.3380**	6	0.2995	3

From the results in Fig. 5 and Table IV, we observe several facts in order. First, the MOS indexes in the HID2016 database keep a roughly negative correlation with the AQI index. Although it is not shown here, this accordance commonly exist for PM2.5 and PM10 indexes of Beijing, and the conclusion fits several other cities in the database. Second, comparing to AQI and PM10, the relationship between MOS and PM2.5 values is easily fitted by three-order polynomial functions, except for Lasa. The comparison results among AQI, PM2.5 and PM10 are shown in Fig. 6. Third, the fitting degree of Beijing is constantly better than other cities, and those of Lasa and Shijiazhuang are stably in the bottom. As for the performance difference, one possible reason is that the subjective rating range for Beijing is broader than those for Shijiazhuang and Lasa, as shown in Table. 2. The limitation of the sample range results the difficulty to find regularity from the data. To observe intuitively, in Fig. 7 we illustrate small version of the haze images in these three cities during forty continuous days.

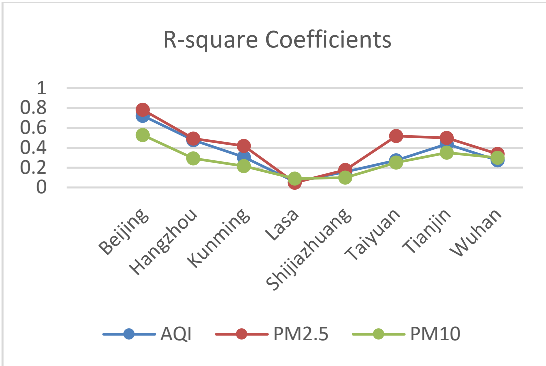


Fig. 6. Comparison results of R-square coefficients between subjective evaluation score and environmental indexes.

From Fig. 7 we can see that the air quality in Beijing in the shooting period varied dramatically, and the air quality in Shijiazhuang and Lasa remained at a very bad or good level, respectively. These phenomena are quite consistent with the changing tendency of the subjective evaluation results.

(a) Beijing

(b) Shijiazhuang

(c) Lasa

Fig. 7. Comparison of haze images in different cities [19].

Besides the sample range, another possible reason for the poor performance in Lasa might be the fact that the shooting position and focal length of the camera changed relatively greatly than other cities. As shown in Fig. 7 (c), the shooting position in the fifth row, fourth column is quite different from others. Although we already removed this picture from the subjective evaluation experiment, the similar problem exists for the rest of this group of pictures. By watching carefully, it is easy to find that the size of the foreground (i.e. the mountain) changed occasionally in different recording dates.

V. Conclusions

In this paper we built a haze image database and performed preliminary studies. First, we collected haze images to evaluate their image quality. Second, subjective evaluation tests were developed to provide benchmark for objective blind IQA on haze images. Finally, we analyzed the experimental data and revealed some implicit relationships between subjective scores and environmental indicators.

It should be pointed out that the current study is quite preliminary and there are a lot of further work need to be done in the future. For example, the performance of different blind IQA metrics needs to be evaluated in the HID2016 database. Furthermore, the correlation between subjective evaluation results and environmental indexes should be thoroughly studied. We will keep improving the database so that in the near future a stable version could be provided for the public to download.

Acknowledgement

We thank the April Winds website who held the photographic exhibition “Breathing in China” in 2014, and thank the photographers who took the haze pictures, especially those who authorized us to use the haze photos to build the HID2016 database.

References

[1] M. Ding, and R. Tong, "Efficient dark channel based image dehazing using quadtrees," Science China Information Sciences, vol. 56, pp. 1-9, 2012.
[2] Q. Liu, M. Chen, and D. Zhou, "Single image haze removal via depth-based contrast stretching transform," Science China Information Sciences, vol. 58, pp. 1-17, 2014.
[3] Y. H. Shiau, P. Y. Chen, H. Y. Yang, C. H. Chen, and S. S. Wang, "Weighted haze removal method with halo prevention," Journal of Visual Communication and Image Representation, vol. 25, pp. 445-453, 2014.
[4] Z. Wang, and Y. Feng, "Fast single haze image enhancement," Computers & Electrical Engineering, vol. 40, pp. 785-795, 2014.
[5] J.-B. Wang, N. He, L.-L. Zhang, and K. Lu, "Single image dehazing with a physical model and dark channel prior," Neurocomputing, vol. 149, Part B, pp. 718-728, 2015.
[6] D. Wu, Q. Zhu, J. Wang, Y. Xie, and L. Wang, "Image Haze Removal Status, Challenges and Prospects," Proc. of 2014 4th IEEE International Conference on Information Science and Technology, pp. 492-497, 2014.
[7] S. Fang, J. Yang, J. Zhan, H. Yuan, and R. Rao, "Image Quality Assessment on Image Haze Removal," Proc. of Chinese Control and Decision Conference, pp. 610-614, 2011.
[8] J. Li, H. Zhang, D. Yuan, and M. Sun, "Single image dehazing using the change of detail prior," Neurocomputing, vol. 156, pp. 1-11, 2015.

[9] F. Guo, J. Tang, and Z.-X. Cai, "Image dehazing based on haziness analysis," International Journal of Automation and Computing, vol. 11, pp. 78-86, 2014.
[10] N. Hautiere, J.-P. Tarel, D. Aubert, and E. Dumont, "Blind contrast enhancement assessment by gradient ratioing at visible edges," Image Analysis and Stereology, vol. 27, pp. 87-95, 2008.
[11] H.R.Sheikh, Z.Wang, L.Cormack,et al., "LIVE image quality assessment database release 2," http://live.ece.utexas.edu/research/quality.
[12] E.C. Larson, and D.M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," Journal of Electronic Imaging, vol. 19, 2010.
[13] N. Ponomarenko, V. Lukin, A. Zelensky, et al., "TID2008 - a database for evaluation of full-reference visual quality assessment metrics," Advances of Modern Radioelectronics, vol. 10, pp.30-45, 2009.
[14] N. Ponomarenko, O. Ieremeiev, V. Lukin, et al., "Color image database TID2013: Peculiarities and preliminary results," 2013 4th European Workshop on Visual Information Processing, pp. 106-111, 2013.
[15] D.M. Chandler, and S.S. Hemami. "A57 database," http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html.
[16] Z.M.P. Sazzad, Y. Kawayoke, and Y. Horita. "Image quality evaluation database," http://mict.eng.u-toyama.ac.jp/database_toyama/.
[17] A. Ninassi, P.L. Callet, and F.P. Autrusseau, "No Reference image quality metric using perceptual data hiding," SPIE Human Vision and Electronic Imaging, 2006.
[18] C. Strauss, F. Pasteau, F. Autrusseau, et al., "Subjective and objective quality evaluation of lar coded art images," 2009 IEEE International Conference on Multimedia and Expo, pp. 674-677, 2009.
[19] Blog of April winds, "Breathing in China", 2014, http://admin.blog.siyuefeng.com/blogArticle/show/9914#.
[20] Weather forecast website, http://www.tianqihoubao.com/.
[21] ITU-R, "Recommendation ITU-R_BT.500-11, Methodology for the subjective assessment of the quality of the television pictures," 2002.
[22] EBU, "Subjective quality of internet video codes-phase 2 evaluation using SAMVIQ ".