Subjective Assessment of Face Photographs by Best-Worst Method

How Contracdictive Factors are Evaluated When We See Images

Kyosuke Shimada, Hirohisa Takeshita and Seiichi Gohshi
Department of Informatics
Kogakuin University
Tokyo, Japan
E-mail: em15009@ns.kogakuin.ac.jp, takeshita1312@gmail.com, gohshi@cc.kogakuin.ac.jp

***Abstract*— Human faces are commonly photographed. We know from experience that we are attracted to faces when we appreciate personal photographs. Therefore, it is clear that personal photographs have a different type of interest from landscape photographs in subjective assessment. Finding interesting aspects of such photographs can improve the conditions of taking photographs. These conditions provide a significant guideline for developing digital cameras that capture high-quality photographs. Investigation of interesting factors requires image assessment considering noise and resolution because those are the most important factors of image quality determining in the assessment. This study reports important factors of image quality assessment based on the investigation of processes assessed by resolution and noise on human face photographs.**

Keywords- Subjective assessment; Human face; Noise; Resolution.

I. INTRODUCTION

A. Background

Imaging technology has enabled video equipment including digital cameras to capture high-resolution images in inferior lighting conditions and printed photographs of such images have been high quality. Luminance of images is determined by the sensitivity of the complementary metal-oxide semiconductor (CMOS) image sensor [1] and lighting conditions. Resolution depends on the number of image sensors. The International Standards Organization (ISO) [2] number is an international standard for sensitivity of image sensors. Recent digital cameras can select a wide range of sensitivities from ISO 100 to ISO 6400 for taking images. The image quality is changed by setting the ISO sensitivity and the F-number in the same lighting condition. The F-number affects the luminance of an image. The image quality differences due to setting fluctuations appear on images as resolution and noise. It is difficult to evaluate images taken with changing F-number and ISO sensitivity because the difference in the photographic quality is too small.

The conditions for high quality are high-resolution and low-noise. Various noise reduction methods have been used to reduce noise. However, noise and resolution are contradictory factors because resolution decreases with reduction in noise.

Noise is an essential issue for images because it degrades image quality. Even though sensitivity of imaging devices has dramatically improved in recent years, all images taken with imaging devices contain certain levels of noise. Resolution, which is a contradicting factor with relation to noise, is also an important element when assessing the image quality. Noise and resolution are composed with high frequency elements and it is impossible to divide them with digital signal processing. Due to this reason, we have to assess image quality in accordance with the balance of noise and resolution.

There are two methods to assess image quality. They are objective assessments and subjective assessments. Although the objective assessment is reproducible and reliable, it is difficult to apply it to images with noise and high resolution. In contrast the subjective assessment is a practical method to assess image quality. However, the subjective assessment has not yet been applied to assess the image quality that has contradictory factors, noise and resolution. The subjective assessment comes with advantage of reproducibility. It can be guaranteed with statistical analysis. We have proposed the Best-Worst Method (BWM) to assess images that have high resolution and noise [6]. In this paper we propose BWM to assess human face photographs that have noise and high resolution. When we see an image, our attentions are automatically drawn to human faces. In this paper the assessment results based on our natural instinct are discussed.

B. Previous work

Digital cameras became a commodity and our daily lives are filled with digital images. In this environment, image assessment became an interested research field [3][4][5]. However, any study focusing on evaluation of printed photographs has yet to be reported. One study indicated that observers paid more attention to image quality marks [3] but the details of the evaluation were not shown. Therefore, image quality assessment for printed photographs must develop an evaluation method and investigate evaluation tendencies.

In our previous study, subjective assessment BWM in still life photographs achieved a good result [6]. BWM can evaluate images of slight quality differences, but it is impossible to use typical subjective assessment methods on such images. As an example of typical methods, the

 ISBN: 978-1-61208-464-0

normalization ranking method [7] is not reproducible because the quality difference of the images is too small. At another example, the Double-Stimulus Impairment Scale (DSIS) method [8], was used to evaluate slightly different images, but it is not a method for evaluating several images. Printed photographs of images taken with digital cameras are often compared to another. Therefore, the evaluation method requires considering the conditions, such as several images and slight differences. BWM in previous study performed well under these conditions.

The results of our previous study also showed the effectiveness of BWM. Furthermore, the surveys found specific evaluation areas regarding noise and resolution. Noise was evaluated in a flat dark area. Resolution was evaluated in an area of fine pattern. In addition it is possible to get reproducibility results when evaluable noise and resolution area exist separately in the same image.

In many cases people take personal images. Humans focus on the faces in these images. This property affects the processes of the image quality evaluation. However, the processes of such images have not been elucidated. The photographs of faces have the same properties as the photographs of still-life. In general, noise of images increases with ISO sensitivity increment and resolution increases with image sensors. In other words noise and resolution are the most important factors when assessing the quality of taken images. Evaluation of the image quality factors and the evaluated areas vary with the amount of noise. Therefore, subjective assessment must consider three things when human face images are evaluated. The first one is to use images taken with varied ISO value. The second one is the investigative processes when evaluating noise and resolution. The last one is the usage of an experimental method such as BWM for evaluating images differences.

This study investigates evaluation processes in human face images. Then subjective assessment using BWM will produce reproducible results for evaluation processes.

This paper is organized, as follows. In Section 2, the subjective assessment materials, the evaluation method and the experiment procedures are explained. In Section 3, the experiment results are presented. In Section 4, the results are analyzed and discussed. In Section 5, we conclude our report.

II. Subjective Assessment Method

This section explains necesary elements of subjective asessment. The elements are experimental photographs, evaluation method, and experimental procedure.

A. Photographs used

The photographs for assessment are shown in Figures 1-4. These photographs were named Entrance, Side by Side, Two Rows and Group, respectively. The size of all photographs is 297 $\times$210 mm (almost letter size) and they are printed with high resolution images of 4.048 $\times$ 3,048 pixels. These images were taken with varied ISO values. The ISO values are ISO 100, 200, 400, 800, 1600, 3200, and 6400. The camera was operated in full auto-mode.

Figure 1. Entrance

Figure 2. Side by Side

Figure 3. Two Rows

Figure 4. Group

B. BWM

Firstly, observers select the highest and lowest quality images. These images are then disqualified. Then, the process is repeated on the remaining images until there is only one photograph left. Finally, the observers give higher and lower ranks to disqualified photographs, and give middle rank to the last selected photograph.

C. Experimantal procedure

Observers for this study are 21 men and women of 20 years of age. The number of observers was set in accordance with International Telecommunication Union recommendation ITU-R BT.500-13 [8]. The observers evaluate the experimental photographs using the BWM and assign the quality ranks. The ranks are used to get evaluation scores. Range of the scores is from 7 to 1 because the photographs of seven different ISO values are evaluated at one time. The larger scores mean higher quality. The ISO 100 photograph of Figure 3 was excluded from the experimental photographs because it appeared to be severely degraded. Therefore, photographs of Figure 3 have the scores from 6 to 1 according to the image quality.

Observers received a training session because this experiment targets people without any knowledge of images. The training session was used to explain noise and resolution. A still-life photograph that was not evaluated was used for descriptive purpose. Items of the description are determinable noise or resolution areas and strength of the factors. The observers were trained while looking at the still-life photograph. The determinable noise areas and the non-edge flat areas. In the description of the noise, the observers were explained to recognize the presence of noise in various areas of the photograph. In the description of the resolution the observer focused on fine-edged areas after that they were told that non-blurry areas were high resolution.

This experiment focuses on evaluation process of face images. After the experiments observers were questioned about the areas in each image that were interesting in the evaluation aspects. There were two main areas of interest. The first one was the observer's ability to pick out the areas of each image, which were interesting in terms of evaluating the noise and the resolution. The other one was which areas in each photo the observers chose to focus on first.

III. Results of Experiment

Subjective assessment results are shown in Figures 5-8. Diamond points on the scale are the average of the scores and the bars are deviations. The horizontal axis shows the ISO values. The values of standard deviations are represented on the right-hand side of the diamond points. These figures indicate the quality differences in each image. In the Figures 5-8, the quality differences can be guaranteed statistically if there is not overlap of the bar between images. The statistical difference is explained by the probability density function regarding the normal distribution. The state without overlap of the bars represents the sufficiently small probability that the evaluation scores of each image are too similar. The probability is calculated by the probability density function. If the probability is sufficiently small, image quality difference is sufficiently significant. In Figures 5-8, the images without the lowest ISO value indicated quality difference because there were no overlaps of the bars between them. Therefore, this experiment was in obtaining the quality differences.

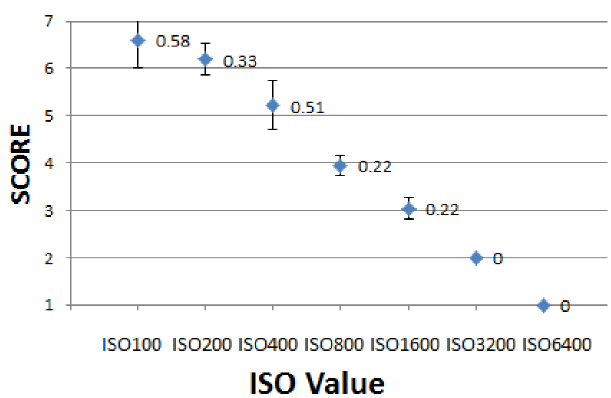

Figure 5. Result of Figure 1

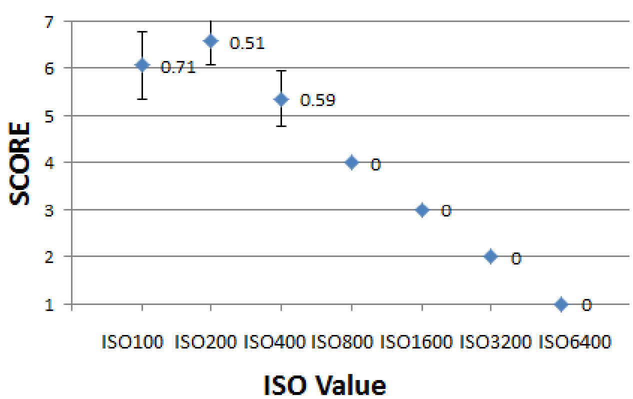

Figure 6. Result of Figure 2

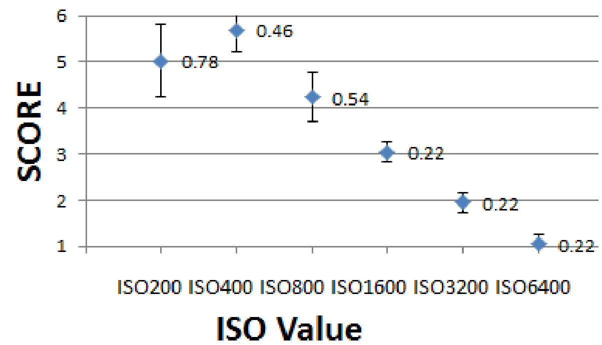

Figure 7. Result of Figure 3

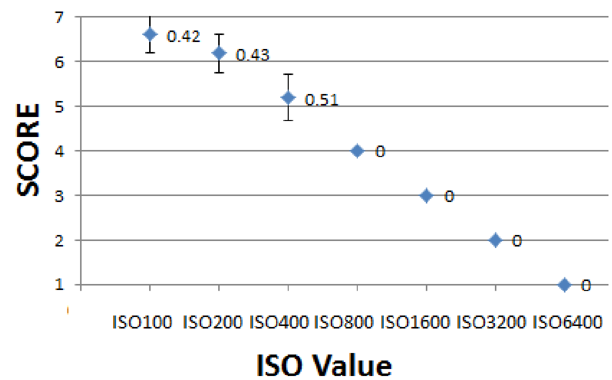

Figure 8. Result of Figure 4

TABLE I. Interesting Areas in terms of Noise and Resolution

Name	*Noise*	*Resolution*
Figure 1	Wall of building, Cheek	Hair, Outline of face
Figure 2	Cheek, Forehead	Hair, Eye
Figure 3	Cheek, Forehead	Hair, Eye
Figure 4	Curtain, Cheek	Hair, Outline of face

TABLE II. Number of People Interested in the Face areas

Name	*Noise(Unit: people)*	*Resolution(unit: people)*
Figure 1	12	6
Figure 2	21	14
Figure 3	21	19
Figure 4	10	9

The results of questions regarding interesting areas are shown in Table 1 and 2. Table 1 represents the interesting areas for evaluation of noise and resolution. From the areas, specific features regarding evaluation of noise and resolution were acquired. The features of noise evaluation were that the areas which contained flat component. The resolution is evaluated fine-edges in each image. The flat areas are walls of building and cheeks. The fine-edged areas are hair, eye and outline of face. A few observers evaluated the areas

 ISBN: 978-1-61208-464-0

other than those listed in Table 1. These areas also had the same features.

Table 2 shows number of the observers evaluating the face areas. A face has hair, cheeks, eyes, etc. The numbers were counted independently. The presence or absence of evaluation in the face areas is shown in Table 2. All observers evaluated noise in the face areas of Figure 2 and 3. However, Figures 1 and 4 were not evaluated in certain area. Therefore, in Figure 2 and 3 existence of the face evaluation was acquired. Although the experimental images had many differences of subject conditions, only Figures 2 and 3 have the common feature. The feature is the larger face than in Figures 1 and 4.

The results regarding the first choice evaluating areas are shown as following. All observers selected the face areas for evaluating at first in Figures 1-4. This result was the same regardless of the areas chosen for evaluating the quality.

The results regarding evaluation noise and resolution were acquired from the questions after the assessments. The results showed that noise becomes visible in the photographs over ISO 800. It was verified such evaluation according to the results that 20 observers evaluated noise in the images.

IV. DISCUSSION

The results are discussed to indicate effectiveness of BWM in this section. In addition they are analyzed to perfect the evaluation processes.

A. *Effectiveness of BWM*

The effectiveness of the BWM is shown by accurate and reproducible results. In Figures 5-8, the evaluation score decreases with increase of ISO value. It means that these experimental results are accurate. The accurate results prove that each photograph is given correct evaluation score. In order to verifying this method, we have to compare the experimental results with general image quality evaluations. In general, the images taken with high ISO value contain high levels of noise. Because the images with high levels of noise are low quality, the images of high ISO values should be evaluated at low scores. In Figures 5-8, these experimental results are consistent with the general image quality evaluations. Therefore, experiment of this study obtained accurate results.

The reproducibility is the most important things in subjective assessments. In Figure 5-8, each mean of scores are separated beyond the deviations. It means that the subjective assessment results have high reproducibility. The images were taken at the same place with camera in auto-mode. Only ISO values are different. Although some images looked similar, observers were able to assess them with reproducible results. It means that BWM can be applicable to the evaluation of images with slight different image quality. The similar assessments for still life with BWM were successful. The difference between still life photographs and personal one is just one thing whether the people were present in them. That difference is the interesting point when it comes to the observers watching the images. If there are men in the images, observers watch human faces. The first areas of focus strongly affect the assessment results. This is according to our experience.

B. *Variations of evaluation areas by face size*

Figures 2 and 3 were evaluated in face areas. These figures had larger face areas than Figures 1 and 4. Therefore, this section will show numerical values of face size to indicate a criterion of evaluation processes. In addition variation of the processes due to changing face size will be discussed.

The size of face area was calculated to acquire numerical criterion in each figure. The size means the ratio of the face area to the entire image. The areas that were used for calculating are shown as square areas in Figure 9. Width of the squares was decided according to the contour lines of cheeks. Height of the squares was selected by distance from the forehead to the chin. These face areas were named Area A, B, C, and D. Number of pixels of the each area was used for calculating the ratio. The calculating results of face size are shown in Table 3. The ratio of Area A means that Figure 1 contains the face areas of 1.0% size. The values of centimeter in Table 3 represent the size of face area in printed photographs. In case of Figure 1, actual size of 1.0% means area of 6.237 cm^2 because the area of photograph in this experiment is 623.7 cm^2.

By these results, existences of face evaluation were shown in areas of 1.5% face size or more. Only Figures 2 and 3 contained such face size. In experimental results, the figures were evaluated in face areas. Therefore, the numeric values are consistent with the experimental results. It means that the numeric value of necessary face size for evaluation is 1.5% or more.

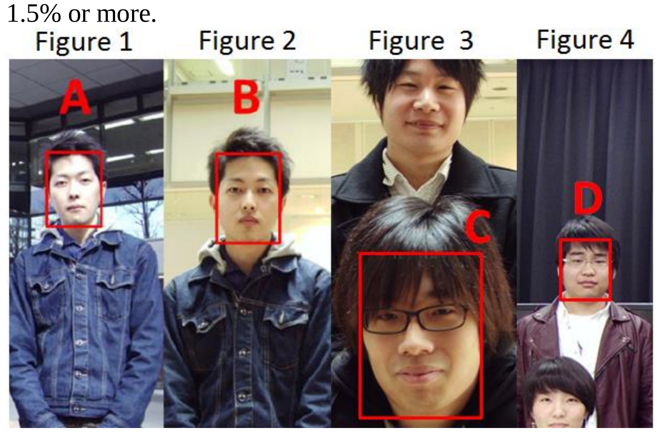

Figure 9. Cut face areas

TABLE III. RATIO OF FACE AREA

Name	*Ratio (%)*	*Area (cm2)*
Printed Photograph (297 ×210 mm)	100	623.700
Figure 1: Area A	1.0	6.237
Figure 2: Area B	1.5	9.356
Figure 3: Area C	4.4	27.443
Figure 4: Area D	0.4	2.495

 ISBN: 978-1-61208-464-0

In other words the observers evaluate face areas if an image contains face size of 1.5% or more. The face size represents area of 9.356 cm^2 in printed photographs. It means that face area of 9.356 cm^2 in photographs are required for face evaluating. The size of 9.456 cm^2 is area of approximately 3.6 cm × 2.6 cm.

C. Image quality factor by changing physical amount of noise

The experimental results showed that noise areas are evaluated in images above ISO 800. This section confirms the accuracy of these results by calculating a physical amount of noise in the evaluation area.

In general, calculating noise requires flat areas of luminance, but it is difficult to find such areas in image. In this study, areas containing ramp of luminance were selected to compare noise amount with evaluation results. Actually, calculated results matched subjective scores. Figure 10 shows selected areas for noise calculation as the areas in the frames. These areas are 100 × 100 pixels and were selected based on experimental results.

Noise amounts were calculated using (1) and (2) in each ISO value images. The noise amount was required as the decibel (dB) value that each ISO value compared to ISO 100. The large dB value means that the image contains much noise. The equation to get the values is shown as following.

$$\mathrm{MSE_{ISO}} = \frac{1}{N^2}\sum\sum(\mathrm{Y_{ISO}} - \mathrm{AVE_{ISO}})^2 \qquad (1)$$

$$\mathrm{ANS_{ISO}} = 10\log\frac{\mathrm{MSE_{ISO}}}{\mathrm{MSE_{100}}} \qquad (2)$$

There are two calculation processes. They are mean square error (MSE) and logarithm. The equation of MSE is shown in (1). MSE represents the differences between pixel values and mean of the values. The value of MSE is acquired in each ISO value. The dB value is calculated by MSE. The equation is shown in (2).

The calculated noise results amounts are shown in Figure 11-14. The horizontal and vertical axes represent the amount of noise and the means of evaluation scores respectively. ISO value is shown at the top of circle plots. The existence of noise evaluations in each image is shown by the calculated noise results amounts. Confirmation of the existence needs to refer to the relation between the evaluation scores and the noise amounts. In the calculated results if the amount of noise increases with decreasing evaluation score of the image, it indicates such images are evaluated by noise. The reason is that the amount of noise in the evaluation areas affects the evaluation scores. According to Figures 11-14, noise is major factor to determine the image quality over ISO 800. This result matches the observers comments after the assessments of Figures 1-4. In Figure 12, the score of ISO 200 is higher than that of ISO 100. This is caused by the resolution. It means that noise levels of these photos were assessed similar, and that the resolution of the photo of ISO 200 was evaluated higher than that of ISO 100.

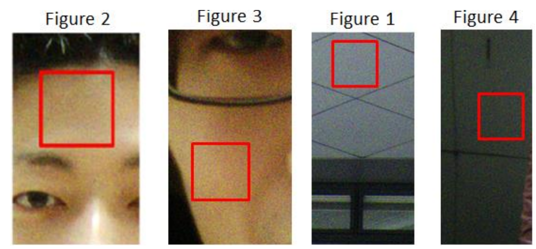

Figure 10. Cut area to calculate noise level

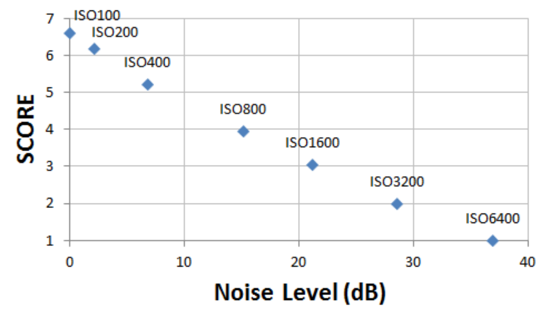

Figure 11. Results of noise calculation: Figure 1

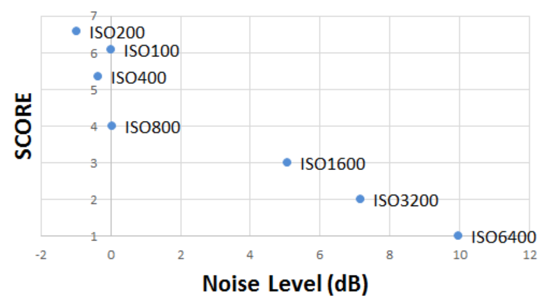

Figure 12. Results of noise calculation: Figure 2

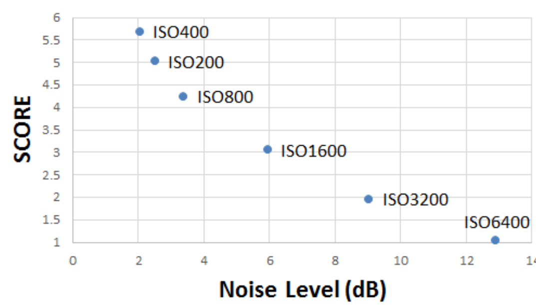

Figure 13. Results of noise calculation: Figure 3

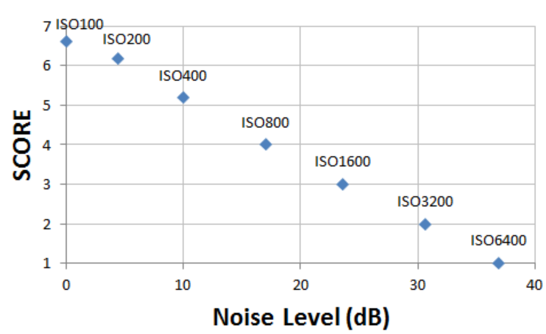

Figure 14. Results of noise calculation: Figure 4

 ISBN: 978-1-61208-464-0

Overall results of noise amount are discussed. The scatter points of ISO 800 or greater showed decreasing evaluation scores in accordance with increasing amounts of noise in all images. Furthermore, survey results regarding image quality factors required that ISO 800 or greater images are evaluated by noise. Therefore, these calculated results indicated that there is noise evaluation in ISO 800 or greater images.

D. Overall evaluation by noise and resolution

Resolution and noise are high frequency components that appear as texture or noise on images. It is extremely difficult to separate these components using signal processing. However, human vision can evaluate such images by separating them. The experiment of this study considered noise and resolution as the factors of image quality and investigated the evaluation processes of these components by human subjectivity. This section discusses how to evaluate noise and resolution in image quality assessment of human vision.

Contradictory factors (noise/resolution) are evaluated in photographs with different ISO sensitivity value. The overall evaluation process is as follows. First, the observers try to evaluate noise in face area. If they cannot evaluate the face area, they evaluate noise in area other than face. Second, when they cannot evaluate noise, resolution of face area is evaluated. In the same way as the noise, if evaluation of resolution of face is impossible, the observers evaluate areas other than the face. In other words, if they cannot evaluate the face area, the evaluation is done in the area around the face. At evaluating time, noise is preferentially evaluated.

V. CONCLUSION

In this paper the subjective assessment for noise and resolution in photographs with human faces is proposed. It is our nature that we try to find human faces when we see an image. Because of this, observers try to evaluate the image quality in human face areas. However, it is not easy to evaluate the contradictive factors, noise and resolution in face areas. BWM was applied for observers to recognize the difference in noise levels and resolution. The BWM assessment results are theoretically analyzed and the statistical differences are obtained. BWM is effective when evaluating the contradictive factors and minor level differences. The following facts are also shown in our experiments. Face areas in an image are initially evaluated. If the observers are not able to find image quality differences in the face areas, they shift their attention to the background in order to evaluate noise. In the evaluation, noise preceded the resolution. Furthermore this study indicated that the certain size of face area is required for the face evaluation. The optimal size is 1.5% or more of the entire image.

The results in this paper could contribute to the future digital camera solutions. In the future it will become necessary to evaluate images with changing photography.

REFERENCES

[1] D. Das, H. J. Mills, and S. Collins, “A wide dynamic range CMOS image sensor with the optimum photoresponse per pixel”, IEEE International Symposium on Circuits and Systems (ISCAS), Rio de Janeiro, Brazil, May 15-18 2011, pp.1560-1563.

[2] ISO 5800:197, “Photography – Colour negative films for still photography – Determination of ISO speed”, November. 1984.

[3] A. E. Boberg, “Subjective vs Objective Image Quality”, XVIIth ISPRS Congress Technical Commission 1: Primary Data Acquisition, August, 1992, pp.73-78.

[4] Y. He, Y. Xuan, W. Chen, and X. Fu, “Subjective Image Quality Assessment: a Method Based on Signal Detection Theory”, IEEE International Conference on Systems, Man and Cybernetics, 2009(SMC 2009), San Antonio, USA, Oct. 2009, pp.4915-4919.

[5] R. K. Mantiuk, A. Tomaszewska, and R. Mantiuk, “Comparison of Four Subjective Methods for Image Quality Assessment”, COMPUTER GRAPHICS FORUM Volume 31, August. 2012 pp.2478-2491.

[6] H. Takeshita and S. Gohshi, “Subjective Assessment for Digital Images”, IEEJ International Workshop on Image Electronics and Visual Computing 2014 (IEVC2014), October 7-10 2014, 3C-3.

[7] T. Fukuda and R. Fukuda, “Guide of Human Engineering: How to scientific sensibility”, 2009, pp.73-123.

[8] ITU-R Recommendation BT.500-13,“Methodology for the subjective assessment of the quality of television pictures”, Jan.2012.

[9] Digital Image Processing Compilation Committee, “Digital Image Processing”, CG-ARTS Society, 2004, pp.79-80.

 ISBN: 978-1-61208-464-0