VOL. 10, NO. 17, SEPTEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

AN EVALUATION OF THE EFFECT OF IMAGE DOWN-SAMPLING ON PERFORMANCE INDICATORS OF IQA ALGORITHMS

K.R. Joy[1] and E. Gopalakrishna Sarma[2]
[1]Electronics & Communication Engineering, Karpagam University, Coimbatore, India
[2] Sree Buddha College of Engineering for Women, Pathanamthitta, India
E-Mail: joykarotte@gmail.com

ABSTRACT

Image Quality Assessment (IQA) tools are becoming more and more indispensible in today's digital image processing applications. Researchers are coming up with new approaches and methodologies as well as modifications to the existing methodologies for the overall improvement of the performance of IQA algorithms. Many currently available IQA algorithms deploy automatic down-sampling the input images while estimating the image quality. In this paper we analyze the effect of image down-sampling on the estimated values of quality and on the performance indicators of full reference IQA algorithms. Results show that the estimated quality and the performance indicators have a strong dependency on the down-sampling factor.

Keywords: image quality assessment, IQA, FR-IQA, image down-sampling, IQA algorithms.

INTRODUCTION

Image Quality Assessment (IQA) has become an important area of research in digital image processing. IQA is the process and methodology of assessing the quality of an image for the intended purpose. As most of the images are ultimately viewed by human beings, the best method to judge the quality of an image is by visually evaluating it by human observers. But, subjective evaluation of image quality is not only time consuming, but also expensive and not practical in real-time or embedded applications. Therefore, it is well accepted that objective assessment methods can do this task very conveniently. However, such objective algorithms must incorporate features of human visual system (HVS) [1], [2], [3] for a realistic assessment of quality. Applications of IQA include medical imaging, security and surveillance, remote sensing, astronomy, multimedia communications, entertainment etc. IQA is also being used for monitoring the image quality for controlling quality of processing systems, for benchmarking image processing systems, for optimizing algorithms and parameter settings for image processing systems etc [4]. Currently, a number of objective IQA algorithms are available in the literature which can estimate the quality of a digital image with varying degree of accuracy and complexity [3], [5].

Classification of IQA Algorithms

IQA algorithms can be broadly classified in to three major categories depending upon whether a reference image is available or not. They are no-reference IQA (NR-IQA or Blind IQA), full reference IQA (FR-IQA) and reduced reference IQA (RR-IQA) [3], [6]. NR-IQA refers to quality assessment of an image without a reference image. Our visual system can easily distinguish high-quality images from low-quality images with little effort and without seeing the original image [7], [8]. In the same way, NR-IQA assesses the quality of an image without any reference image. There are three basic approaches towards NR-IQA based on how the objective algorithm derives the quality score. They are: 1. Distortion-Specific approach, 2. Feature extraction and learning based approach: and 3. Natural Scene Statistics based approach (NSS). FR-IQA compares the distorted image with a reference image for the assessment of quality. Since this method has the complete information about the reference image, the results of FR-IQA are supposed to be superior to other IQA algorithms. Some of the approaches for FR-IQA are those based on image fidelity and pixel wise differences, those based on human visual system (HVS), those based on image structures, those based on information content, those based on image statistics and machine learning etc. In Reduced Reference IQA model, the quality of the distorted image is assessed with partial information from the reference image [7], [9]. The partial information is the features extracted from the reference image. It is a compromise between FR and NR approaches to IQA in terms of quality prediction accuracy and the amount of information required for describing the reference image. In the case of FR-IQA, the reference image is always required to estimate the quality of the distorted image, but the results are reliable. NR-IQA does not require a reference image, but results are not proven to be consistently reliable in performance. In this paper, we focus on FR-IQA algorithms.

Existing Work

The earliest approaches for FR-IQA were based on computation of pixel-wise differences between the distorted image and the reference image such as Mean Squared Error (MSE) and Peak Signal to Noise Ratio (PSNR) [10], [11]. These methods are known for their simplicity and speed of execution. The main disadvantage of these algorithms is their poor correlation with perceived quality. Over time, FR-IQA has evolved to more advanced procedures like SSIM [12], FSIM [13], ESSIM [14], GMSD [15] etc. The quality prediction capabilities of the FR-IQA algorithms have improved significantly during recent years. In the same way, the correlation of the estimated quality index with the perceived quality has also improved.

Down-sampling the input images has been suggested by many authors while performing objective FR-IQA in order to take in to account the viewing distance of the image [16], [17]. The down-sampling factor was generally decided based on the height or width of the display device and the viewing distance. A study on the effect of down sampling was conducted by Zhou Wang et al in 2003 in connection with Multi Scale Structural Similarity (MS-SSIM) algorithm [18]. But its scope was limited to SSIM algorithm on JPEG/JPEG2000 images of LIVE database [19] and did not cover the effect on estimated quality or runtime.

Proposed Work

Many currently available FR-IQA algorithms deploy automatic down-sampling of the input images during the quality assessment process. The down-sampling factor in such algorithms is determined by the input image size as illustrated in section 2. Due to the recent advancements in digital image processing technology, the sizes of the digital images used in most of the applications are increasing at an exponential scale. The manner in which an image is viewed by human observers has also undergone changes. Many applications require fine details of the image. In such situations, assessing the quality on the down-sampled image may give incorrect results. Further, down-sampling process may introduce runtime overheads which will increase the runtime of the IQA algorithms for large images [20]. In this paper, we analyze the effect of down-sampling on the estimated quality, the runtime of the algorithms and the performance indicators such as Spearman Rank Order Correlation Coefficient (SROCC), Kendall Rank Order Correlation Coefficient (KROCC), Pearson Linear Correlation Coefficient (PLCC) and Root Mean Square Error (RMSE) [21], [22]. It has been observed that down-sampling the input images affect all the above parameters. The extent to which these parameters are affected depends upon the down-sampling factor ***f***. The rest of this article is organized as follows. Section 2 describes the down-sampling process as deployed in most of the FR-IQA algorithms. Section 3 describes the details of the experiments, section 4 gives the results and discussions and section 5 gives the conclusion.

DOWNSAMPLING PROCESS

By down-sampling, the sizes of the input images are scaled down. A typical MATLAB code for down-sampling the input images are shown below:

```
% Down-sampling input images
% im1 is the reference image
% im2 is the test image
[M,N]=size(im1);
f = max(1,round(min(M,N)/256));
if(f>1)
  lpf = ones(f,f);
  lpf = lpf/sum(lpf(:));
  im1=imfilter(im1,lpf,'symmetric','same');
  im2=imfilter(im2,lpf,'symmetric','same');
  im1 = im1(1:f:end,1:f:end);
  im2 = im2(1:f:end,1:f:end);
end
```

By executing the above code, the size of the input images would be reduced by a factor ***f***, which depends upon the original image size. For example, the value of ***f*** is 14 for an input image of size 5202*3465 pixels and 2 for an image of 512*512 pixels.

DETAILS OF EXPERIMENTS

The experiments were conducted on a personal computer with Intel Core i5-2340M processor, 4 GB RAM and Windows 7 operating system. The following aspects are studied during the experiments:

- Variation of estimated quality as a function of down-sampling factor ***f***
- Variation of run time as a function of down-sampling factor
- Variation of performance indicators such as SROCC, KROCC, PLCC and RMSE for selected algorithms and image data bases.

For convenience, we have selected two popular IQA algorithms which deploy automatic image down-sampling for this study as representative algorithms and the results are described in this paper. They are the Structural Similarity Index (SSIM) [12] and Feature Similarity Index (FSIM) [13]. In fact the behavior of other algorithms is also similar, but the results are not discussed in this paper. For measuring the run time, the algorithm was executed a number of times and the average run time was taken. The MATLAB source codes and default parameters set by the authors of the respective algorithms were used in the experiment. We used images from various sources. For small images, images from public data bases such as LIVE2 [19], TID 2008 [23] and CSIQ [24] were used. For large images, images were captured using a digital camera for this study. The IQA algorithms were executed for various values for the down-sampling factor ***f***. Instead of automatic down-sampling, manual down-sampling was used. The experiment was repeated for images of different sizes and distortion levels.

RESULTS AND DISCUSSIONS

Variation of Estimated Quality with *f*

The variation of estimated quality Vs down-sampling factor shows similar trends for both the algorithms. For small values of ***f***, the estimated quality increases in a near linear manner with ***f***. As ***f*** increases further, the increase in the estimated quality follows a non-linear relation. For very large values of ***f***, the estimated

VOL. 10, NO. 17, SEPTEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

quality saturates to its maximum value. This is illustrated in Figures-1 (a) and 1 (b). The curves can be divided in to three regions. (1) A near linear region where the estimated quality increases almost linearly with ***f***. This is represented by OP in the graph. In this region, as ***f*** increases, the information content of both reference and distorted images reduces and the difference between the images becomes less prominent. As a result, the estimated quality increases. For large images, this region spans around ***f*** = 1 to ***f*** = 8. (2) Non-linear region PQ where the estimated quality index varies in a nonlinear manner with ***f***. (3) A saturation region QR where there is no noticeable change in the estimated quality with increase in ***f***. This phenomenon can be explained as follows. As the down-sampling factor ***f*** increases, the information content of the down-sampled images decreases significantly. As ***f*** increases further, the information loss becomes very high, such that the reference image and the distorted image become almost similar. At this stage, the estimated quality index approaches unity. It is important to note that any meaningful comparison of quality is possible only when the down-sampling factor ***f*** is in the linear portion of the curve.

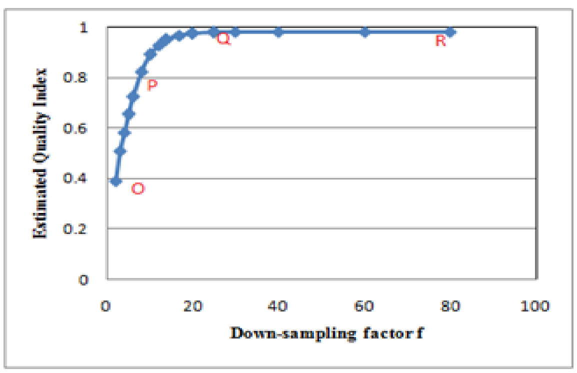

Figure-1(a). Variation of estimated quality with down sampling factor for an RGB image of size 5202*3465 pixels using FSIM algorithm.

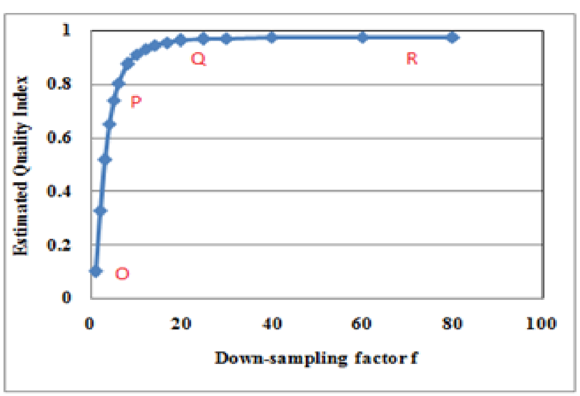

Figure-1(b). Variation of estimated quality with down sampling factor for an RGB image of size 5202*3465 pixels using SSIM algorithm.

Variation of Runtime with *f*

The runtime of the algorithms initially shows a decrease with increasing values of ***f***. It reaches a minimum value and then increases as ***f*** increases. This variation is graphically illustrated in Figure-2(a) and 2(b) for SSIM and FSIM respectively. The initial decrease in runtime is due to the reduction of the image size and hence the reduction in the number of steps in computing the quality index. This process continues until the runtime reaches its minimum value. However, as the down-sampling factor ***f*** increases further, the down-sampling process itself becomes an overhead and takes a significant time for its execution. This overrides the benefits of the reduced computational steps due to the decrease in the image size. As a result, the overall runtime increases. This is illustrated in Table-1 and graphically in Figure-3.

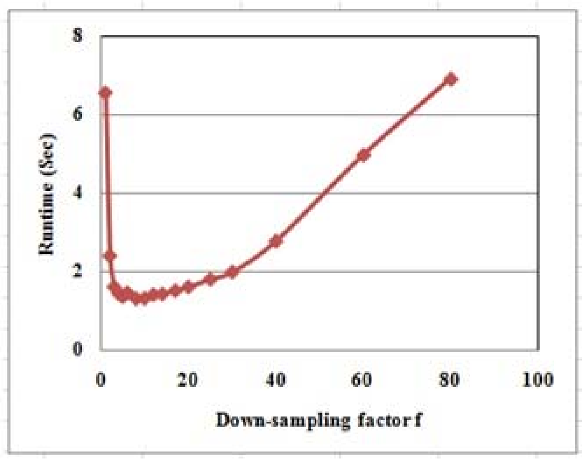

Figure-2(a). Variation of estimated quality with down sampling factor for an RGB image of size 5202*3465 using SSIM algorithm.

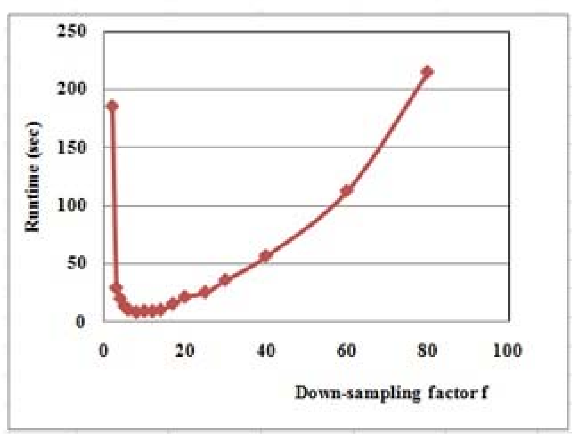

Figure-2(b). Variation of average runtime with down sampling factor for an RGB image of size 5202*3465 using FSIM algorithm.

VOL. 10, NO. 17, SEPTEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

For small values of ***f***, the time for quality estimation is more significant where as the time for down sampling is negligible. As ***f*** increases, the time for quality estimation decreases and time for down-sampling increases. For large values of ***f***, the major portion of the runtime is utilized for down-sampling. It is important to note that in order to optimize the runtime efficiency of the algorithms, the down-sampling factor must be selected in such a way that it operates at the minimum runtime.

Table-1. Components of total runtime as a function of down-sampling factor f. FSIM algorithm used on RGB image of 2601 * 1733 pixels.

Down-sampling Factor f	Down-sampling Time (Sec)	Quality Estimation Time (Sec)	Total Time
2	0.249	16.662	16.910
3	0.257	6.329	6.586
4	0.298	3.318	3.617
6	0.443	1.345	1.788
8	0.677	0.864	1.541
10	0.963	0.468	1.431
12	1.360	0.341	1.701
16	2.284	0.244	2.528
20	3.393	0.200	3.593
24	4.815	0.176	4.990
30	7.498	0.083	7.581

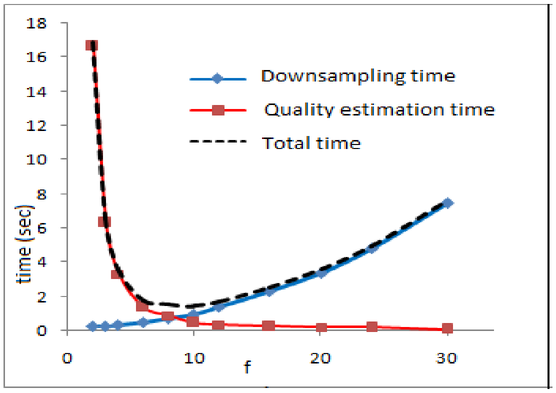

Figure-3. Components of total runtime as a function of down-sampling factor f. FSIM algorithm used on RGB image of 2601 * 1733 pixels.

Variation of Correlation Parameters

In this section we discuss the variation of correlation parameters such as Spearman Rank Order Correlation Coefficient (SROCC), the Kendall Rank Order Correlation Coefficient (KROCC), the Pearson Linear Correlation Coefficient (PLCC) and the Root Mean Square Error (RMSE) with the down-sampling factor. SROCC and KROCC measure the prediction monotonicity. These parameters depend on the rank of the data points. The relative distances between the data points are not considered in the computation of these parameters. For the other two parameters (PLCC and RMSE), we have to do a regression analysis and perform a nonlinear mapping between the predicted quality indices and the MOS values. An algorithm is considered good, if the parameters SROCC, KROCC and PLCC are high and RMSE is low. Detailed procedures for computing these parameters are described in [25] and [26]. For this study, we used the images in the LIVE2, TID2008 and CSIQ image databases. We used the same IQA algorithms (SSIM and FSIM) for this study also. Since the images in these data bases are relatively smaller in size, we used the down-sampling factor from 1 to 8.

The results of this study are illustrated graphically in Figures-4 to 9. It is interesting to note that all the results show similar trend. The correlation parameters SROCC, KROCC and PLCC have a maximum value for a down-sampling factor around 2. The values of these parameters decrease for other values of ***f***. Similarly the value of RMSE is minimum for down-sampling factor 2.

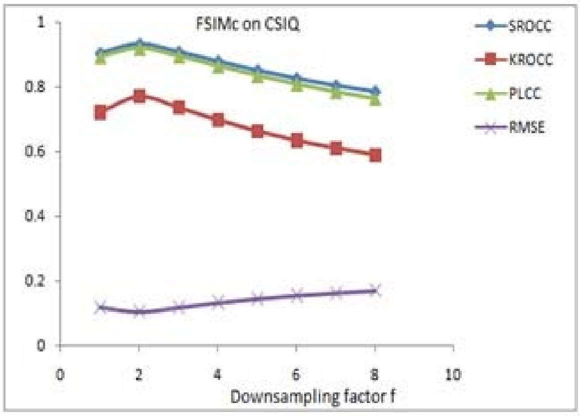

Figure-4. Variation of SROCC, KROCC, PLCC and RMSE for FSIMc on CSIQ image data base.

VOL. 10, NO. 17, SEPTEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

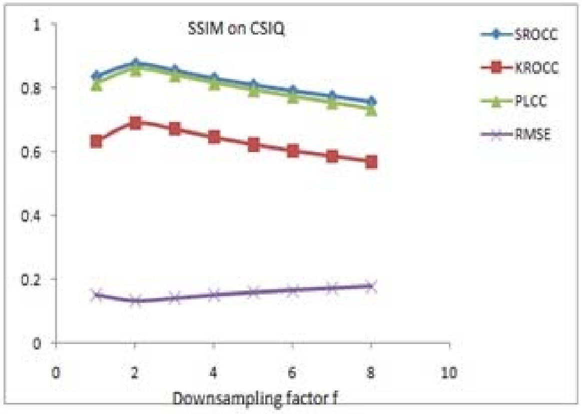

Figure-5. Variation of SROCC, KROCC, PLCC and RMSE for SSIM on CSIQ image data base.

It indicates that if we use the above image data bases for any study, we get the best results when we use a down-sampling factor of 2.

The variations in the values of SROCC, KROCC, PLCC and RMSE can be explained with the help of fig. 10. It illustrates the changes in the SSIM values calculated for two image pairs from the LIVE database. In the beginning (***f***=1), SSIM for image 1 is greater than image 2. After point A, SSIM for image 2 becomes higher.

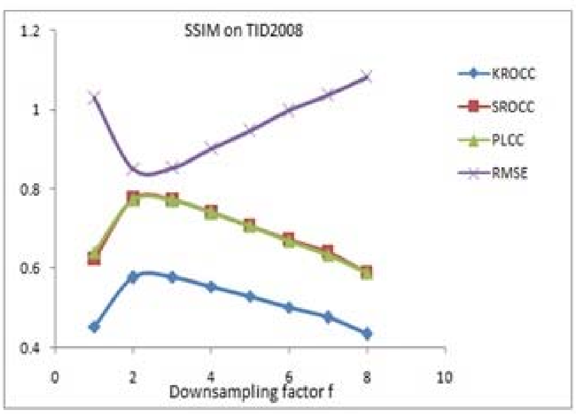

Figure-6. Variation of SROCC, KROCC, PLCC and RMSE for FSIMc on TID2008 image data base.

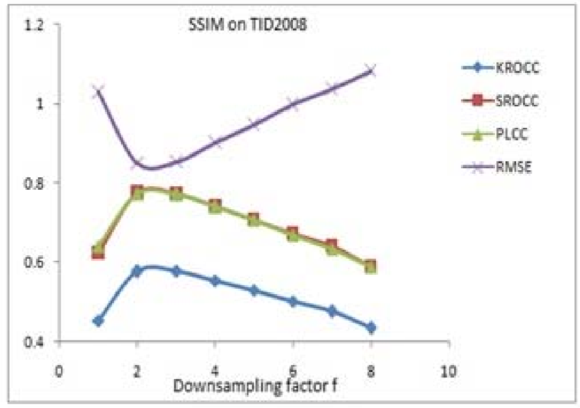

Figure-7. Variation of SROCC, KROCC, PLCC and RMSE for SSIM on TID 2008 image data base.

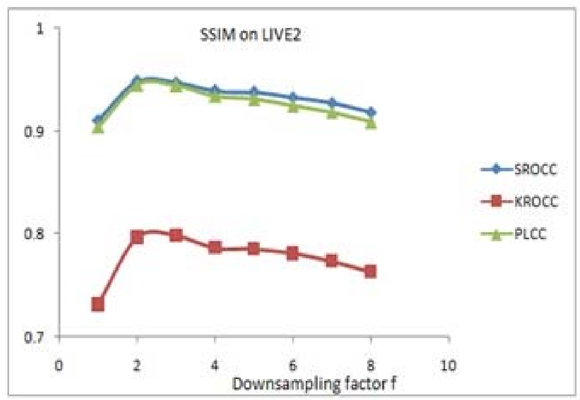

Figure-8. Variation of SROCC, KROCC, PLCC and RMSE for SSIM on LIVE2 image data base.

Again after point C, SSIM for image 1 becomes higher. The changes in the SSIM values of these two images change their relative rank with respect to the MOS values. Since the correlation parameters depend on the relative rank and/or the distances with reference to the MOS values, these parameters are changed as the rank changes or in other words as the down-sampling factor changes. The amount of change in quality index (e.g. SSIM) depends upon the particular image and the down-sampling factor.

VOL. 10, NO. 17, SEPTEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

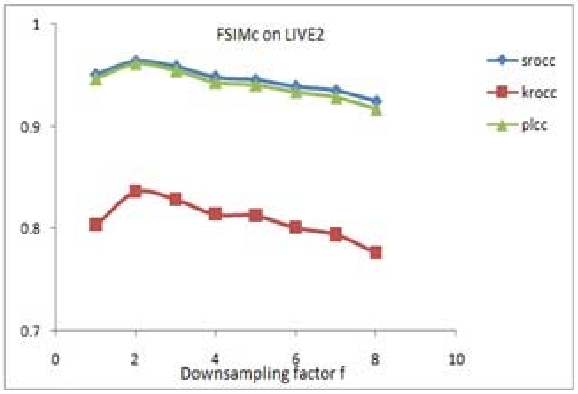

Figure-9. Variation of SROCC, KROCC, PLCC and RMSE for SSIM on LIVE2 image data base.

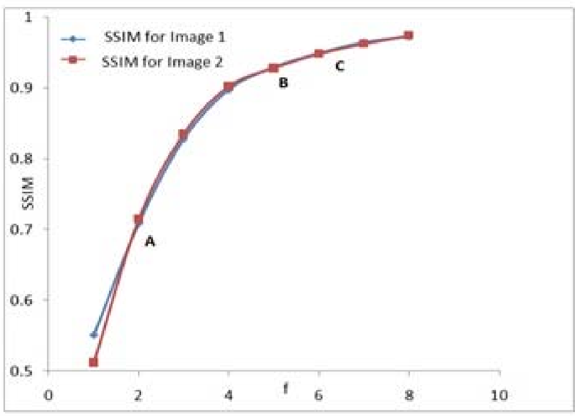

Figure-10. Variation of SSIM values for two different images as a function of ***f***.

CONCLUSIONS

In this paper, we have evaluated the effects of down-sampling input images on the estimated quality and performance indicators of FR-IQA algorithms. We used popular FR-IQA algorithms SSIM and FSIM as well as popular public image data bases for this study apart from custom made large images. It has been observed that the performance indicators vary widely with the down-sampling factor. Therefore, before reaching in to a conclusion about the quality of an image or about the performance of an IQA algorithm, it is important to check the experimental conditions upon which the results are obtained. If the same FR-IQA algorithm is run on the same image pair, but with different down-sampling factors, the estimated quality would be higher when the down-sampling factor is high. Similarly, if we rank a set of images based on their estimated quality, the rank order may be different for the same image at different down-sampling factors. Hence the comparison of images based on their estimated quality may give different results at different down-sampling factors. Since many IQA algorithms deploy automatic image down-sampling, the down-sampling factor becomes very high for large image sizes. Such algorithms may give inflated value (almost equal to unity) for quality which may not be correct. Therefore such algorithms may have to be fine tuned before applying to such large images. One major limitations of this study is the non-availability of public image databases with large images. However, this study gives the IQA research community an insight to understand the dependency of down-sampling factor on the estimated quality and the performance indicators of the algorithms.

REFERENCES

[1] Gao X., Lu W., Tao D. and Li X. 2010. Image quality assessment and human visual system, Visual Communications and Image Processing, International Society for Optics and Photonics, July.

[2] Rafael C. Gonzalez and Richard E. Woods. 2008. Digital Image Processing, 3rd edition, Pearson Education Inc, pp. 39-43.

[3] Chandler D. M. 2013. Seven challenges in image quality assessment: past, present, and future research, ISRN Signal Processing.

[4] Wang Z. 2011. Applications of objective image quality assessment methods- Application Corner, IEEE Signal Processing Magazine, Vol. 28, No. 6, pp. 137-142.

[5] K.R. Joy and E. Gopalakrishna Sarma. 2014. Recent Developments in Image Quality Assessment Algorithms: A Review. J. Theoretical and Applied Information Technology, Vol. 65, No. 1, pp. 192-201.

[6] Liu T. J., Lin Y. C., Lin W. and Kuo C. C. J. 2013. Visual quality assessment: recent developments, coding applications and future trends. APSIPA Transactions on Signal and Information Processing, 2, e4.

[7] Wang Z. and Bovik A. C. 2011. Reduced-and no-reference image quality assessment. IEEE Signal Processing Magazine, Vol. 28, No. 6, pp. 29-40.

[8] Saad, M. A.; Bovik, A. C. & Charrier, C. 2010. A DCT statistics-based blind image quality index. IEEE Signal Processing Letters **17**(6), 583-586.

[9] Li Q. and Wang Z. 2009. Reduced-reference image quality assessment using divisive normalization-based image representation. IEEE J. Selected Topics in Signal Processing, Vol. 3, No. 2, pp. 202-211.

[10] Rafael C. Gonzalez and Richard E. Woods. 2008. Digital Image Processing, 3rd edition, Pearson Education Inc., 354

[11] Wang Z. and Bovik A. C. 2009. Mean squared error: love it or leave it? A new look at signal fidelity measures. IEEE Signal Processing Magazine, Vol. 26, No. 1,pp. 98-117.

[12] Wang Z., Bovik A. C., Sheikh H. R. and Simoncelli E. P. 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. Vol. 13, No. 4, pp. 600-612.

[13] Zhang L., Zhang D. and Mou X. 2011. FSIM: a feature similarity index for image quality assessment. IEEE Trans. Image Process, Vol. 20, No. 8, pp. 2378-2386.

[14] Zhang X., Feng X., Wang W. and Xue W. 2013. Edge strength similarity for image quality assessment. IEEE Signal Processing Letters, Vol. 20, No. 4, pp. 319-322.

[15] Wufeng Xue, Lei Zhang, Xuanqin Mou. and AC Bovik. 2014. Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index, IEEE Trans. IP, Vol. 23, No. 2, pp. 684-695.

[16] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli. The SSIM index for image Quality assessment. Online available at: <http://ece.uwaterloo.ca/_z70wang/research/ssim/>

[17] W. Lin and C.-C. Jay Kuo. 2011. Perceptual visual quality metrics: A survey. J. Visual Communication and Image Representation, Vol. 22, No. 4, pp. 297–312.

[18] Wang Z., Simoncelli E. P. and Bovik A. C. 2004. Multiscale structural similarity for image quality assessment, Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, IEEE, Vol. 2, pp. 1398-1402.

[19] H.R. Sheikh, Z. Wang, L. Cormack and A.C. Bovik: Live Image Quality Assessment Database Release 2. Online at http://live.ece.utexas.edu/research/quality/subjective.htm

[20] K.R. Joy and E. Gopalakrishna Sarma 2014. Impact of Image Size on Performance of IQA Algorithms on Large Images. Adv. in Nat. Appl. Sci., Vol. 8, No. 19, pp. 23-30, 2

[21] Final Report From the Video Quality Experts Group on the validation of Objective Models of Video Quality Assessment, Phase II, VQEG, August 2003. Online: http://www.vqeg.org

[22] Sheikh H. R., Sabir M. F. and Bovik A. C. 2006. A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Process. Vol. **15, No.** 11, pp. 3440-3451.

[23] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. 2009. TID2008 - A database for evaluation of full-reference visual quality assessment metrics, Adv. Modern Radioelectron., Vol. 10, pp. 30-45.

[24] E.C. Larson and D.M. Chandler. 2010. Most apparent distortion: full-reference image quality assessment and the role of strategy. J. Electr. Imaging, vol. 19, pp. 001006:1-21.

[25] Zhang L., Zhang L., Mou X. and Zhang D. A comprehensive evaluation of full reference image quality assessment algorithms. 2012. In 19th IEEE International Conference on Image Processing (ICIP), pp. 1477-1480.

[26] Lin Zhang Lei Zhang: Research on Image Quality Assessment. Online at sse.tongji.edn.cn/linzhang/IQA/IQA.htm