

IMAGE QUALITY ASSESSMENT BY SALIENCY MAPS

Edoardo Ardizzone and Alessandro Bruno

*Dipartimento di Ingegneria Chimica, Gestionale, Informatica e Meccanica, Università degli studi di Palermo,
Viale delle scienze edificio 6, Palermo, Italy*

Keywords: Image Quality Assessment, Visual Saliency, Saliency Map, Human Visual System, Perceptual Quality.

Abstract: Image Quality Assessment (IQA) is an interesting challenge for image processing applications. The goal of IQA is to replace human judgement of perceived image quality with a machine evaluation. A large number of methods have been proposed to evaluate the quality of an image which may be corrupted by noise, distorted during acquisition, transmission, compression, etc. Many methods, in some cases, do not agree with human judgment because they are not correlated with human visual perception. In the last years the most modern IQA models and metrics considered visual saliency as a fundamental issue. The aim of visual saliency is to produce a saliency map that replicates the human visual system (HVS) behaviour in visual attention process. In this paper we show the relationship between different kind of visual saliency maps and IQA measures. We particularly perform a lot of comparisons between Saliency-Based IQA Measures and traditional Objective IQA Measure. In Saliency scientific literature there are many different approaches for saliency maps, we want to investigate which is best one for IQA metrics.

1 INTRODUCTION

Digital Images can be distorted during acquisition, compression, transmission, restoration, processing, etc. Image Quality Assessment aims to replace human judgment of perceived image quality with machine-based evaluation. Traditional criteria (Wang, 2006) perform measures based on the differences between reference and distorted image, this measures are not correlated with human visual perception. The “perfect” IQA method is subjective evaluation, because it performs results directly from human visual system. Unfortunately this kind of method is very expensive and time consuming (a lot of time and observers are requested for good performances). IQA methods can be subdivided in two main categories: subjective (Recommendation, 2002) and objective (Wang, 2006) methods. The first class is very expensive and cannot easily performed in real time systems.

The goal of second class methods is to perform a statistical measure of image quality perceived by human being. Objective Methods can be further categorized in the following groups: full-reference, no-reference and reduced-reference. Full-reference means that the original (distortion free) image and distorted image are known; No-reference means that

original image is unknown; Reduced-reference means that the original image is partially available. In this paper we focus our attention on Full-Reference objective methods.

The limit of Full-Reference methods is that the results of their metrics are often far from subjective human evaluation. In the last years IQA methods consider visual attention system to be included in IQA metrics (Ma, 2008, a) (Ma, 2008, b) (Ma, 2010). Visual Saliency aims to replicate the HVS (Human Visual System) attention process through saliency maps. that describe the most focused areas for a human observer. In state of the art there are many different approaches for saliency map computation. We performed a lot of experimental comparisons between IQA measures and Saliency based IQA measures. We studied the influence of different visual saliency approaches on IQA measure performance.

The paper is organised as follows: in section 2 we discuss some State of the art methods about IQA Saliency; in section 3 we describe some IQA metrics; in section 4 we show our experimental results; in section 5 conclusions and future works.

2 STATE OF THE ART

In this section we discuss about some state of the art approaches on Full Reference IQA and Saliency Map Detection.

2.1 Image Quality Assessment

Traditional full reference IQA criteria adopted pixel-wise distances such as PSNR (Peak Signal to Noise Ratio) and MSE (Mean Squared Error). In (Acvibas, 2002) authors showed that these distances are far from quality perceived results.

In (Damera-Venkata, 2000) authors proposed NQM (Noise Quality Measure) and DM (Distortion Measure) that outperform PSNR, but the problem is to how define unique quality metric based on NQM and DM.

Wang et al. in (Wang, 2006) and in (Wang, 2004) proposed MSSIM (Mean Structural Similarity), which analyses the degradation of structural information, that is based only on local correlation property of an image and, for this reason, is not enough precise for IQA applications. Another IQA called VIF (Visual Information Fidelity), is proposed by Sheikh (Sheikh, 2006) is based on mutual information between input and output of HVS (Human vision system). Another interesting IQA measure, MS-SSIM (Multi-Scale Structural Similarity), was proposed by Wang et al. (Wang, 2003) which showed that the perceived quality of an image is heavily dependent upon the scale of observation.

MSSIM and VIF measures are based only on local features, so the global information has lost. The most modern IQA methods consider the process of visual attention as a fundamental aspect for a better IQA.

In (Ma, 2008) authors included saliency features to compute PSNR, MSSIM and VIF measures. These new measures, called SPSNR, SMSSIM and SVIF get better performances than the corresponding original versions. The authors of (Moorthy, 2009) explored visual attention and visual perception for spatial pooling strategies in SSIM metrics. In this paper we explored the relationships between different saliency approaches (Itti,1998) (Harel, 2007) (Ma L., 2010) and the corresponding saliency based metrics against traditional criteria.

2.2 Saliency Maps

Saliency or Visual Saliency is the image processing field that deals with identifying the most important

regions of an image from a perceptual point of view (Frintrop, 2010). In first three seconds a human observer fixates some particular points inside an image and tends to group them into visual significant areas.

A saliency map is effective if precision and recall measures with respect to human fixation points are high. In scientific literature there are different approaches for saliency (Marchesotti, 2009). For a interesting overview about saliency see (Marchesotti, 2009). In our paper we compared three saliency maps to test some IQA metrics: Itti Koch method (Itti,1998) Harel method (Harel,2007) and Ma method (Ma L., 2010). We selected these three methods because they are based on different approaches. Itti Koch model for Saliency detection adopted multi-scale analysis of the image. Multiscale image features are combined into a single topographical saliency map. A dynamical neural network selects attended locations in order of decreasing saliency. This is a bottom-up, stimulus-driven approach. Harel (Harel, 2007) saliency approach is based on a biologically plausible model, it consists of two steps: activation maps on certain feature channels and normalization which highlights conspicuity. This is a bottom-up, stimulus driven saliency model. The approach of Ma method is based on an optimization model: Ant Colony Optimization. From now on we refer to this methods with ITTI (Itti,1998) GBVS (Harel, 2007) and ACO (Ma L., 2010).

3 EVALUATION

In this section we analyze the IQA metrics we used for our experimental set: PSNR, MSSIM, VIF, SPSNR, SMSSIM, SVIF. The first three (PSNR, MSSIM, VIF) are objective measures, the others (SPSNR, SMSSIM, SVIF) are weighted by saliency map values. All the measures analyzed grow with the perceived image quality.

Measures (normalized in the same range [0,1]) are compared with different test conditions in terms of distortion, compression, noise type and localization (global noise, local noise). As suggested by (Ma, 2008), we tested distorted images with the following noises: gaussian, poisson, speckle, salt & pepper. We also considered three possible spatial noise distribution: global noise, noise added only in salient region and noise added only in the not salient region. We consider the gap between two corresponding metrics (PSNR vs SPSNR, VIF vs SVIF, MSSIM vs SMSSIM), as it follows:

$$\text{GapM} = (\text{SB_Met} - \text{NSB_Met}) \quad (1)$$

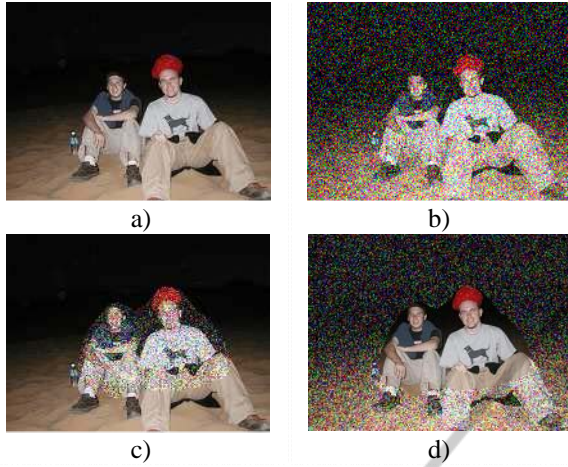


Figure 1: The reference image without distortion a). Global Gaussian noise b). Gaussian noise in salient regions c). Gaussian noise in not salient regions d). Image taken from Torralba Database (Torralba Database).

where SB_Met is the saliency based metric value, NSB_Met is the not saliency based metric value. In our experiments all tests have been done using the ITTI, GBVS and ACO saliency maps. More precisely in our experiments we compared the IQA metrics and the corresponding saliency based ones. In fig. 2,3,4 we show how GapM changes in function of several kind of noise and with various saliency map used.

4 EXPERIMENTAL RESULTS

In this section we show and discuss our experimental results. We subdivided our tests into two parts. In first part we show IQA metrics values with several conditions of noise and visual saliency. In the second part we show the dispersion diagrams of IQA metrics with respect to a human subjective evaluation.

4.1 IQA & GapM

In our tests we used a Database (Live Database) which is made of original and distorted images, and their subjective evaluations. In eq. 1 we defined GapM . In rest of the paper we will refer to:

$$\begin{aligned} \text{GapM}_1 &= \text{SPSNR} - \text{PSNR}; \\ \text{GapM}_2 &= \text{SMSSIM} - \text{MSSIM}; \\ \text{GapM}_3 &= \text{SVIF} - \text{VIF}; \end{aligned}$$

We selected from (Live Database) 100 images with the corresponding corrupted ones by the following

types of noise: Gaussian; Poisson; Salt & pepper, Speckle. We furthermore created from the reference images (without distortion or noise) two noisy version (with noise located only in salient regions or in not salient regions). As described in tab.1, for each corrupted image we computed $\text{GapM}_{(i=1...3)}$.

Table 1: Report example for $\text{GapM}_{(i)}$.

| $\text{GapM}_{(i=1...3)}$ | Gaussian | Poisson | Salt & pepper | speckle |
|---------------------------|----------|---------|---------------|---------|
| SPSNR-PSNR | 0,161 | -1.125 | 0,158 | -0,214 |
| SMSSIM-MSSIM | 0,189 | -0,001 | 0,202 | -0,042 |
| SVIF-VIF | 0,041 | 0,015 | 0,032 | 0,060 |

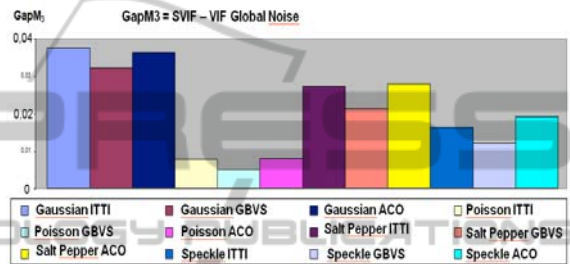


Figure 2: GapM_3 –Global Noise.

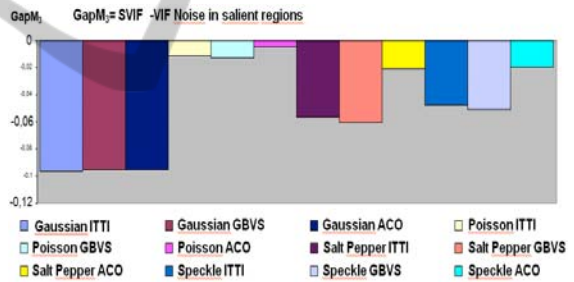


Figure 3: GapM_3 –Noise in salient regions.

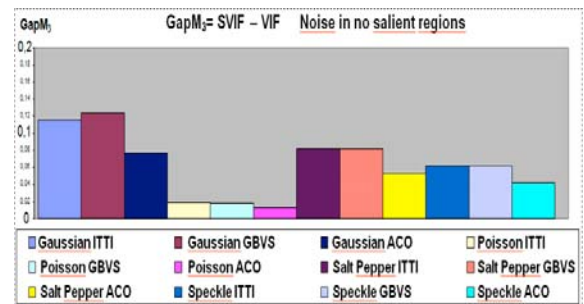


Figure 4: GapM_3 –Global Noise in no salient regions.

In figures 2, 3, 4, for example, we show the mean values of $\text{GapM}_{(3)}$ for all the possible combinations of saliency maps, spatial distributions of noise and kind of noise. In our experiments we noted that $\text{GapM} > 0$ in case of global noise. This tell us that

the saliency based metric perform a better quality image perceived than the other metric. When we tested an image with noise only in the not salient regions SB_Met showed, on the average, higher value with respect of NSB_Met because the salient region appeared with a good grade of perceived quality. On the contrary, if we tested an image with noise only in the salient region, we observed that NSB_Met showed, on the average, higher value with respect of SB_Met, the reason why is: salient region should appear with low grade of perceived quality.

4.2 Validation Test

In this section we show how much the three saliency methods can improve IQA methods. We used the same validation test scheme of (Ma, 2008) for IQA. MOS (mean opinion score) provides an indication of the perceived image quality and DMOS (Ma, 2008) is the difference Mean Opinion Score for an image:

$$DMOS = MOS_{reference} - MOS_{distorted} \quad (2)$$

Where $MOS_{reference}$ is MOS of the reference image, and $MOS_{distorted}$ is MOS of the distorted image. From LIVE database (Live Database) we analyzed:

- 273 images with JPEG2000 compression;
- 200 images with JPEG compression;
- 174 images with white noise in RGB components;
- 174 images with Gaussian Blur;
- 174 images with transmission error in JPEG2000 bitstream using fast-fading Rayleigh (779 distorted images).

As suggested by (Ma, 2008), we plot DMOS vs PSNR, SPSNR, MSSIM, SMSSIM, VIF, SVIF for all the distorted images. We measured IQA metrics within the luma component of the YCbCr model. We also repeated experiments within CieLab model. In fig. 5-6 the SVIF scatter plots for GBVS and ACO saliency maps, that always showed the smaller standard deviation between scatter points and regressive curve. For a given IQA metric it is possible to predict DMOS from the corresponding regressive curve with a five-parametr logistic function (Gottschalk, 2005). The accuracy precision for DMOS is evaluated through Correlation Coefficients.

4.2.1 Correlation Coefficients

In our experiments we evaluated the saliency methods contributions for IQA through correlation coefficients to perform prediction accuracy for DMOS. In detail, we exploited:

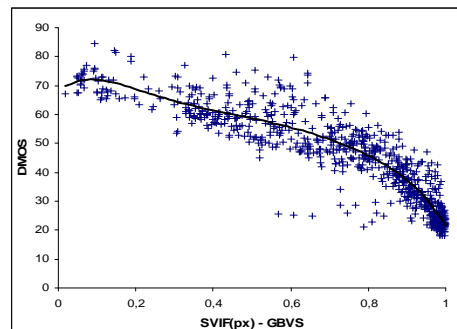


Figure 5: Scatter plot SVIF with GBVS saliency.

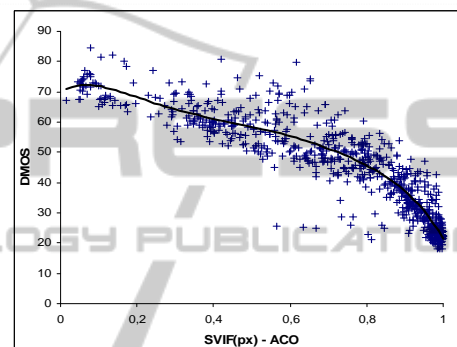


Figure 6: Scatter plot SVIF with ACO saliency.

- CC (Pearson Linear Correlation Coefficient) (Nagelkerke,1991) has value in $[-1,1]$.
- R^2 (Coefficient of Determination) (Nagelkerke,1991) has value in $[0,1]$.
- KRCC (Kendall Rank Correlation Coefficient) (Prokhorov,2001)), has value in $[-1,1]$.
- SROCC (Spearman Rank Order Correlation Coefficient) (Brunnstro,2009)has value in $[-1,1]$.
- MAE (Mean Absolute Error)(Wilmott,2005).
- RMS (Root Mean Square Predict Error) (Wilmott,2005)

In RMS and MAE lower values mean better accuracy, in KRCC, SROCC, CC and R^2 higher values mean better accuracy.

We used these coefficients to measure how DMOS predicted values approximate human subjective DMOS score. A better value tell us which is the best IQA metric. In our experiments we saw that SVIF metric based on ACO saliency maps outperforms the others (tab.2).

SVIF with YCbCr color model and ACO saliency map perform the highest values for CC, R2, KRCC, SROCC, and the lowest values for MAE and RMS. We also tested all the IQA indexes using CIElab color model, but we did not find a metric that absolutely outperformed the others.

Table 2: Mean Values of correlation coefficients (YCbCr Space).

| INDEX | CC | R ² | MAE | RMS | KRCC | SROCC |
|---------------|--------|----------------|--------|--------|--------|--------|
| PSNR | 0,7980 | 0,6368 | 7,7503 | 9,7064 | 0,5883 | 0,7917 |
| MSSIM | 0,9071 | 0,8228 | 5,1338 | 6,7813 | 0,7188 | 0,9004 |
| VIF | 0,9227 | 0,8513 | 4,4954 | 6,2133 | 0,7574 | 0,9241 |
| SPSNR (ITTI) | 0,7893 | 0,6230 | 7,8856 | 9,8883 | 0,5822 | 0,7835 |
| SMSSIM (ITTI) | 0,9131 | 0,8337 | 5,0010 | 6,5761 | 0,7330 | 0,9100 |
| SVIF (ITTI) | 0,9216 | 0,8494 | 4,4692 | 6,2549 | 0,7574 | 0,9236 |
| SPSNR (GBVS) | 0,7964 | 0,6343 | 7,7239 | 9,7391 | 0,5885 | 0,7907 |
| SMSSIM (GBVS) | 0,9120 | 0,8318 | 5,0067 | 6,6060 | 0,7287 | 0,9073 |
| SVIF(GBVS) | 0,9214 | 0,8490 | 4,4496 | 6,2647 | 0,7578 | 0,9236 |
| SPSNR(ACO) | 0,8126 | 0,6603 | 7,4350 | 9,3867 | 0,6058 | 0,8077 |
| SMSSIM(ACO) | 0,9128 | 0,8332 | 5,0408 | 6,6088 | 0,7406 | 0,9149 |
| SVIF(ACO) | 0,9236 | 0,8530 | 4,3895 | 6,1824 | 0,7605 | 0,9251 |

5 CONCLUSIONS

In this work we presented a strong experimentation about the comparison between traditional IQA metrics and visual saliency based ones. All the test confirmed that IQA saliency based methods outperform traditional criteria. We also noted that ACO saliency maps give a stronger support with respect to the other saliency approaches, ITTI and GBVS, especially in case of SVIF metric. Furthermore we pointed out that the SVIF using ACO saliency map also had the largest GapM of several noise distributions. It could be very interesting to stress this kind of experiments with a lot of more saliency approaches. In this way we will establish which could be the best Visual Saliency for Image Quality Assessment.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Alessandro Piero Filippone for helping us in the implementation and experimental phases.

REFERENCES

Z. Wang and A.C. Bovik, 2006. Modern Image Quality Assessment. *New York: Morgan & Claypool*.
 Recommendation, I, 2002. Methodology for the subjective Assessment of the Quality of television pictures. *ITU-R Rec. BT. 500-11*.
 Ma, Q. And Zhang, L., 2008. Saliency-based image quality assessment criterion. *Advanced Intelligent Computing Theories and Applications With Aspects of Theoretical and Methodological Issues*.

Ma, Q. And Zhang, L., 2008. Image Quality Assessment with visual Attention. *19th ICPR*.
 Ma, L. and Li, S. And Ngan, K.N., 2010. Visual Horizontal Effect for Image Quality Assessment. *Signal Processing Letters, IEEE*.
 I. Acvibas, B. Sankur, and K. Sayood, 2002. Statistical Evaluation of image quality measures. *J. Eletron. Imag.*
 N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans and A.C. Bovik, 2000. Image quality assessment based on a degradation model. *IEEE Trans. Image Process.*
 Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image Quality Assessment: from Error visibility to structural similarity. *IEEE Trans. on Image Processing*.
 Sheikh, H. R. , Bovik, A. C., 2006. Image Information and Visual Quality. *IEEE Trans. on Image Processing*.
 Wang, Z., Simoncelli, E., and Bovik, A. C., 2003. Multi-scale structural similarity for image quality assessment. *Proc IEEE Asilomar Conf. on Signals, Systems, and Computers, (Asilomar)*.
 Moorthy, A. K. and Bovik, A. C. , 2009. Perceptually significant spatial pooling techniques for image quality assessment. *Proc. Electronic Imaging*.
 Itti, L. and Koch, C. And Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*.
 Harel, J. and Koch, C and Perona, P., 2007. Grahp-based visual saliency. *Advances in neural information processing systems*.
 Ma, L. and Tian, J. And Yu, W., 2010. Visual Saliency Detection in Image using Ant Colony Optimisation and local phase coherence. *Electronic Letters*.
 Frintrop, S. and Rome, E. And Christensen, H.I., 2010. Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception (TAP)*.
 Marchesotti, L. and Cifarelli, C. And Csurka, G., 2009. A framework for visual saliency detection with applications to image thumbnailing. *12th ICCV*.
 LIVE Database <http://live.ece.utexas.edu/research/quality>
 Torralba Database [http://people.csail.mit.edu/tjudd/ Where PeopleLook/](http://people.csail.mit.edu/tjudd/WherePeopleLook/)
 Nagelkerke, N.J.D., 1991. A note on a general definition of the coefficient of determination. *Biometrika, (78)*.
 Prokhorov, A.V., 2001. Kendall coefficient of rank correlation. Hazewinkel, Michiel, *Encyclopaedia of Mathematics, Springer*.
 Brunnstrom, K. and Hands, D. and Speranza, F. and Webster, A., 2009. VQeg validation and ITU standardization of objective perceptual video quality metrics. *Signal Processing Magazine, IEEE*.
 Wilmott, C.J. and Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*.
 Paul G. Gottschalk, John R. Dunn, 2005. The five-parameter logistic: A characterization and comparison with the four-parameter logistic. *Analytical Bichemistry*.