

IMAGE QUALITY ASSESSMENT BASED ON DETAIL DIFFERENCES

Elio D. Di Claudio^(*) and Giovanni Jacovitti^(**)

Department of Information Engineering, Electronics and Telecommunications (DIET)
University of Rome “La Sapienza”, Via Eudossiana, 18, I-00187 Rome, Italy
Phone: ^(*)+39-06-44585838, ^(**)+39-06-44585490; Fax: +39-06-4873300
Email: ^(*)elio.diclaudio@diet.uniroma1.it, ^(**)giovanni.iacovitti@diet.uniroma1.it

ABSTRACT

This paper presents a novel Full Reference method for image quality assessment based on two indices measuring respectively detail loss and spurious detail addition. These indices define a two dimensional (2D) state in a Virtual Cognitive State (VCS) space. The quality estimation is obtained as a 2D function of the VCS, empirically determined via polynomial fitting of DMOS values of training images. The method provides at the same time highly accurate DMOS estimates, and a quantitative account of the causes of quality degradation.

Index Terms— Image quality assessment, gradient tensor, virtual cognitive space, VICOM, detail analysis.

1. INTRODUCTION

Image quality assessment (IQA) methods perform objective estimates of perceptual quality using image analysis techniques. In some applications, it is only required to predict the perceptual quality of a given test image. In these cases, it could suffice to compare selected statistics of the test image to the corresponding ones of perfect photographic images. In other applications, the problem is to verify the impact of a transfer or reproduction process on image quality. This problem is best solved by direct comparison of the observed image with its original version (*Full Reference*, FR IQA methods). In such cases however, it is also desirable to obtain from IQA methods further indications about the causes of image degradation.

To this regard, some state of the art FR IQA methods do analyze multiple features related to quality, so that such indications could be inferred. For instance, the VIF method [1] explicitly models the image distortion as a local mix of blur and noise. Likewise, the FSIM method [2] measures phase congruency and gradient magnitude. Nevertheless, these techniques do not provide explicit analytic supports for the analysis of the distortion, until now. The *Virtual Cognitive Model* (VICOM) is aimed to overcome this limit [3]. To this purpose, the VICOM predicts perceptual quality by modeling the image distortion as a mix of *gradient loss* and *gradient signal to noise ratio* (SNR) *loss* [3]. This feature pair defines a two-

dimensional (2D) *virtual cognitive state* (VCS) space providing distortion classification of images. The perceived quality is predicted by mapping the VCS into a DMOS scale through a 2D function. The scope of the present contribution is to introduce a novel VCS space defined by *detail loss* and *detail addition* features. The appeal of these features is that they are visually recognizable in images affected by a variety of distortions. On the other hand, it is difficult to define them in a *formal* and *effective* way. An early approach was reported in [4]. Similar features were also adopted in [5], [6], [7] employing a rather different technique. Herein, their formal definition and calculus are based on the analysis of the gradient tensor (GT) [8] leading to a formal definition of DL (detail loss) and DA (detail addition) indices as coordinates of the VCS space. A (desired) property of this space is that states corresponding to blurred images and noisy images are close to the coordinate axes whereas the VCS of other distorted images lie in the middle. Starting from VCS values, the overall perceptual quality is mapped onto the scale of subjective Differential Mean Opinion Score (DMOS) through a 2D polynomial function, determined by fitting empirical DMOS values for a set of training images. Accurate prediction is attained using low order 2D modified polynomials. Interestingly enough, even *linear fitting* leads to a prediction accuracy comparable to the one possessed by best methods prone to linearization with parametric logistic curves. This implies that, unlike most existing IQA methods, the method presented here exhibits a distinct *homoscedastic* behavior of fitting error.

The method presented here will be referred to as VICOM G in the sequel, since it makes use of a VICOM-like scheme [3]. The version based on linear fitting will be referred to as VICOM GL.

2. COMPLEX GRADIENT EIGENANALYSIS

With reference to the generic pixel position $\mathbf{p} \equiv (x_1, x_2)$, the test image is denoted by $I(\mathbf{p})$ and the reference image by $\tilde{I}(\mathbf{p})$. The Gaussian smoothed *complex gradients* [2], [3] of the reference and the test images are denoted

by $\tilde{g}(\mathbf{p})$ and $g(\mathbf{p})$.

For each pixel of the reference image $\tilde{I}(\mathbf{p})$, the *Gradient Tensor* (GT) [8] is calculated as

$$\tilde{\mathbf{R}}(\mathbf{p}) = w(\mathbf{p}; \sigma_w) * \begin{bmatrix} \{\text{Re}[\tilde{g}(\mathbf{p})]\}^2 & \text{Re}[\tilde{g}(\mathbf{p})]\text{Im}[\tilde{g}(\mathbf{p})] \\ \text{Re}[\tilde{g}(\mathbf{p})]\text{Im}[\tilde{g}(\mathbf{p})] & \{\text{Im}[\tilde{g}(\mathbf{p})]\}^2 \end{bmatrix}$$

where the operator $*$ denotes 2D convolution and $w(\mathbf{p}; \sigma_w)$ is a local Gaussian window characterized by spread σ_w .

Rotating the gradient by its *dominant orientation* $\theta(\mathbf{p})$ [8] one obtain the *aligned gradient*

$$\tilde{y}(\mathbf{p}) = e^{-j\theta(\mathbf{p})} \tilde{g}(\mathbf{p})$$

whose GT is *diagonal*. Its real part $\text{Re}[\tilde{y}(\mathbf{p})]$ is the *dominant (strongest)* gradient component of the pattern captured by the window, and its imaginary part $\text{Im}[\tilde{y}(\mathbf{p})]$ is the *secondary (weakest)* gradient component, equal to zero for perfect linear patterns [11]. The energy of these orthogonal components are given by their GT eigenvalues $\tilde{\lambda}_1(\mathbf{p}) \geq \tilde{\lambda}_2(\mathbf{p})$.

3. EDGE AND WEAK TEXTURE FIELDS

Partially following [7], it is convenient to limit quality estimation to two relevant subsets of image point, called *edge* and *weak texture* points. They are defined in the reference image $\tilde{I}(\mathbf{p})$ as follows.

- A generic point is classified as an *edge point* if its gradient magnitude is large, its Laplacian small, and the strongest GT component largely prevails on the smallest one. Formally:

$$A1) \quad 0.1y_M < |\tilde{y}(\mathbf{p})| < 0.3y_M, \quad y_M = \max_{\mathbf{p}} \{|\tilde{y}(\mathbf{p})|\}$$

$$A2) \quad |\tilde{l}(\mathbf{p})| < |\tilde{y}(\mathbf{p})| + \varepsilon, \quad \text{where } \tilde{l}(\mathbf{p}) \text{ is the Laplacian of } \tilde{I}(\mathbf{p}) \text{ and } \varepsilon \text{ is a small threshold (set to 1)}$$

$$A3) \quad \tilde{\lambda}_1(\mathbf{p}) > 32\tilde{\lambda}_2(\mathbf{p}).$$

- A generic point is classified as *weak texture point* if its gradient magnitude is small but non-zero. Formally:

$$0.01y_M < |\tilde{y}(\mathbf{p})| \leq 0.1y_M.$$

For notational convenience, the *edge field* is characterized by the membership function $F_e(\mathbf{p})=1$ on edge points and $F_e(\mathbf{p})=0$ elsewhere. The *weak texture field* is characterized by a similar function $F_w(\mathbf{p})$.

4. DETAIL LOSS AND ADDITION ANALYSIS FROM ORTHOGONAL ORIENTATIONS

The detail analysis is performed in the aligned gradient domain by separately accounting for gradient modifications into the test image in the principal and in the secondary GT directions. To this purpose, the following linear distortion model is adopted:

$$y(\mathbf{p}) = b(\mathbf{p})\tilde{y}(\mathbf{p}) + e(\mathbf{p})$$

where the *real valued* (isotropic) *gain* coefficient $b(\mathbf{p})$ accounts for *local gain/attenuation* and the residual $e(\mathbf{p})$ models possible *noise and artifacts*.

This scheme may be viewed as a translation in the complex gradient domain of the wavelet based decomposition concept followed in [1]. Herein, $b(\mathbf{p})$ is estimated by the following least squares (LS) fitting

$$b(\mathbf{p}) = \frac{\text{Re}[\tilde{y}(\mathbf{p})^* y(\mathbf{p})] * w(\mathbf{p}; \sigma_w)}{|\tilde{y}(\mathbf{p})|^2 * w(\mathbf{p}; \sigma_w) + C_1} \quad (1)$$

where $C_1 = 0.1$ is a regularizing constant [3].

Based on this model it is possible to distinguish among spurious details and lost details, defined as follows :

- Spurious details (noise and artifacts) are detected along the *weakest* aligned gradient component, wherever the local distortion energy defined as $P_e(\mathbf{p}) = |\text{Im}[e(\mathbf{p})]|^2 * w(\mathbf{p}; \sigma_w)$ exceeds the original gradient energy $\tilde{\lambda}_2(\mathbf{p})$. Thus, the value of their binary membership function is $F_{sd}(\mathbf{p})=1$ if $P_e(\mathbf{p}) > \tilde{\lambda}_2$ and $F_{sd}(\mathbf{p})=0$ elsewhere.
- Lost details points are detected wherever the local gradient energy $P(\mathbf{p}) = |y(\mathbf{p})|^2 * w(\mathbf{p}; \sigma_w)$ is smaller than the gradient energy of the original image. Thus, the value of their membership function is $F_{ld}(\mathbf{p})=1$ when $P(\mathbf{p}) < \tilde{\lambda}_1(\mathbf{p}) + \tilde{\lambda}_2(\mathbf{p})$ and $F_{ld}(\mathbf{p})=0$ elsewhere.

To show how effective these rules are, in Fig. 1 detail differences are shown for a JPEG coded versions of the *Church and Capitol* image. The spurious detail map is given by the product $F_{sd}(\mathbf{p}) \cdot |y(\mathbf{p})|$. The lost detail map is given by the product $F_{ld}(\mathbf{p}) \cdot |\tilde{y}(\mathbf{p})|$.

5. MEASURING THE VISUAL IMPACT OF SPURIOUS AND LOST DETAILS

In the spirit of [1] and [3], the visual impact of a spurious detail is put in the form of *log-transformed, SNR measure* referred to the principal GT component of the reference image and to the secondary component of the residual:

$$\gamma_{sd}(\mathbf{p}) = \ln \left[1 + \frac{\tilde{\lambda}_1(\mathbf{p})}{C_2 + F_{sd}(\mathbf{p}) \cdot P_e(\mathbf{p})} \right]$$

where C_2 measures the background vision noise level set to $C_2 = 100$ for a 256 level image. Likewise, the impact of detail losses should be calculated as a *log-transformed SNR measure* referred to the principal GT component of the reference image and to the background vision noise only. Actually, it has been verified that the impact $\gamma_{sd}(\mathbf{p})$ is substantially linear with $b(\mathbf{p})$. To account for this empirical evidence, $\gamma_{ld}(\mathbf{p})$ is linearized with respect to $b(\mathbf{p})$ around $b(\mathbf{p}) = 1$, yielding:

$$\gamma_{ld}(\mathbf{p}) = \frac{\tilde{\lambda}_1(\mathbf{p})}{\tilde{\lambda}_1(\mathbf{p}) + C_2} b(\mathbf{p})$$

6. DMOS ESTIMATION

The cumulative impact of spurious details and lost details is measured by pooling these contributions over the image support. In order to compensate for content dependency, these quantities are normalized with respect to the same measures taken for the original images (where background noise only is present, and $b(\mathbf{p}) = 1$).

The *DA* (detail addition) and *DL* (detail loss) indices are defined as follows:

$$DA = 1 - \frac{\sum_{\{\mathbf{p}:F_e(\mathbf{p})=1\}} \gamma_{sd}(\mathbf{p}) + \sum_{\{\mathbf{p}:F_w(\mathbf{p})F_{sd}(\mathbf{p})=1\}} \gamma_{sd}(\mathbf{p})}{\sum_{\{\mathbf{p}:F_e(\mathbf{p})=1\}} \tilde{\gamma}_{sd}(\mathbf{p}) + \sum_{\{\mathbf{p}:F_w(\mathbf{p})F_{sd}(\mathbf{p})=1\}} \tilde{\gamma}_{sd}(\mathbf{p})} \quad (2)$$

$$DL = 1 - \frac{\sum_{\{\mathbf{p}:F_e(\mathbf{p})=1\}} \gamma_{ld}(\mathbf{p}) + \sum_{\{\mathbf{p}:F_w(\mathbf{p})F_{ld}(\mathbf{p})=1\}} \gamma_{ld}(\mathbf{p})}{\sum_{\{\mathbf{p}:F_e(\mathbf{p})=1\}} \tilde{\gamma}_{ld}(\mathbf{p}) + \sum_{\{\mathbf{p}:F_w(\mathbf{p})F_{ld}(\mathbf{p})=1\}} \tilde{\gamma}_{ld}(\mathbf{p})} \quad (3)$$

Finally, the DMOS prediction \hat{D} of the VICOM G is given by the P -th order formula

$$\hat{D} = \sum_{k=0}^P \sum_{m=0}^k a_{km} \left[(h + DL)^\alpha \right]^k \left[(h + DA)^\beta \right]^m \quad (4)$$

where α and β are real positive exponents and h is a small constant ($h = 0.1$) to make sure that $h + DL > 0$. Only Q coefficients a_{km} are allowed to be non-zero. They are obtained by sequentially minimizing the LS error $\left| DMOS - \hat{D} \right|^2$ averaged over the training database [3].

The most simple case corresponds to linear fitting ($P = 1$, $Q = 3$). The VICOM GL estimate is then obtained as

$$\hat{D}_{lin} = a_{00} + a_{10}DL + a_{01}DA$$

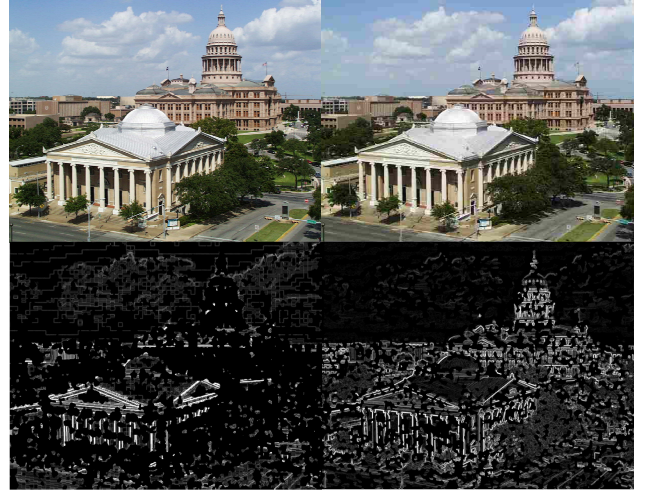


Fig. 1 Top: the original *Church and Capitol* image (left) and its JPEG coded version (right). Bottom: spurious details map (left) and lost details map (right).

7. COMPUTATION STEPS

The calculus of the DL and DA indices is performed through four steps.

- In the first step, complex Gaussian smoothed gradients and the Laplacian are calculated for both the test and the reference images.
- In the second step the local GT analysis is performed applying a Gaussian filter to the squared gradient components [8], yielding spurious and lost detail maps.
- In the third step, the cross products $\text{Re}[\tilde{y}(\mathbf{p})^* y(\mathbf{p})]$ are calculated. Then, the gain $b(\mathbf{p})$ and the SNR values for missing and spurious details are determined.
- Finally, indices DL and DA are obtained by pooling, and the quality is estimated by mapping VCS values into the DMOS scale.

8. THE VICOM G QUALITY CHARTS

In Fig. 2 the VCS space is plotted for the LIVE database [1], [12]. Constant DMOS curves predicted by the VICOM G 2nd order and the VICOM GL are superimposed. These curves are the contour plots of fitting surfaces calculated for LIVE DMOS data as *realigned* in [12]. In these experiments the values of the filter widths were $\sigma = 0.75$ and $\sigma_w = 3\sigma = 2.25$. For the 2nd order VICOM G ($P = 2$, $Q = 4$, $\alpha = 0.45$, $\beta = 0.55$) the values of the non-zero parameters were, $a_{10} = -19.8$, $a_{20} = 107.0$, $a_{11} = -77.9$ and $a_{02} = 102.8$. For the VICOM GL ($P = 1$, $Q = 3$, $\alpha = 1$, $\beta = 1$), the values of other parameters were $a_{00} = -5.5$, $a_{10} = 55.3$ and $a_{01} = 66.3$.

In the VCS space, blurred images cluster in a zone characterized by small spurious detail addition, and noisy images are mostly characterized by small detail loss except for very high noise saturating levels, whereas coded images have intermediate positions. Observe that JPEG images are classified as more noisy than JPEG 2000 images of the same subjective quality.

The scatterplots of estimated DMOS versus empirical DMOS for the two predictors are shown in Figs. 2 and 3. As summarized by the statistical indicators of Table I, the second order estimator outperforms the other quoted ones on the LIVE database, especially because of its very low RMSE. In particular, the prediction accuracy of the linear version is remarkable if compared to quality indices requiring post-linearization [3], [12].

Further experiments were made on a subset of the TID2008 (JPEG2000, JPEG, AWGN, Gaussian blur and JPEG2000 transmission errors) [3], [10] using

The SROCC of various FR IQA predictors trained with the TID2008 subset is listed in Table II.

In these experiments the values of the filter widths were $\sigma = 1.0$ and $\sigma_w = 3\sigma = 3.0$. For the 2nd order VICOM G ($P = 2, Q = 4, \alpha = 0.45, \beta = 0.55$), the values of other non-zero parameters were $a_{00} = 27.2, a_{10} = 80.9, a_{11} = -65.9$ and $a_{02} = 48.5$. For the VICOM GL ($P = 1, Q = 3, \alpha = 1, \beta = 1$), the values of other parameters were $a_{00} = 20.9, a_{10} = 49.0$ and $a_{01} = 36.4$.

For these images the VICOM G predictors exhibit a lower SROCC in comparison with other predictors estimators, such as the FSIM and the recent GMSD [13]. Nevertheless, compared to these predictors, the VICOM scatterplots exhibit more uniform distribution of prediction errors with respect to different distortion errors and to different quality levels, as shown by the plots of Fig. 5 (FSIM and GMSD where post-linearized by a five coefficient logistic function).

8. CONCLUSIVE REMARKS

With respect to conventional methods, the full reference quality assessment method illustrated here presents three main points of interest.

From the operative viewpoint, the method enables to represent the quality of an image as a point in a two dimensional VCS space. VICOM G charts provide not only quantitative estimates of the overall perceptual quality, but also and the relative impact of noise/artifacts

and cancellation of patterns, thus offering a better understanding of image distortion.

From the accuracy viewpoint, the method exhibits very high performance with respect to other competitive methods, and uniform variance of error along the full range of quality levels (*homoscedasticity*) and with respect to different kinds of distortion.

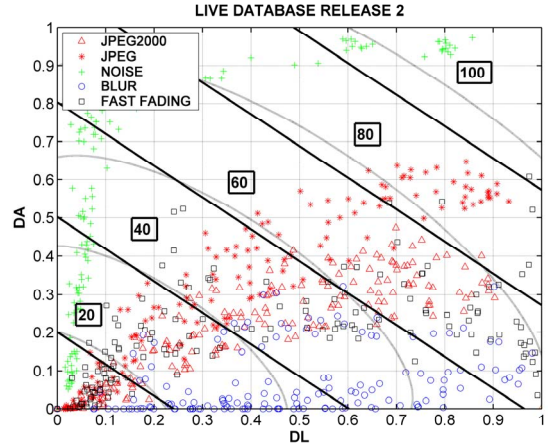


Fig. 2. The VCS plane for the LIVE images. Superimposed are the contour plots of fitting surfaces (curves of equal predicted DMOS) for VICOM GL (black) and the VICOM G 2nd order (gray).

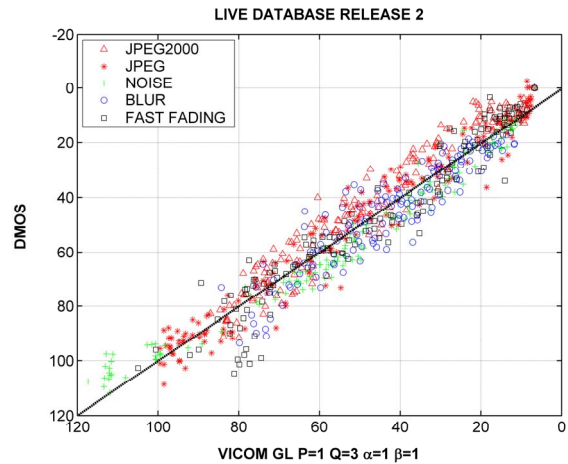


Fig. 3 The VICOM GL scatterplot for the LIVE database. RMSE=7.18, SROCC=0.966.

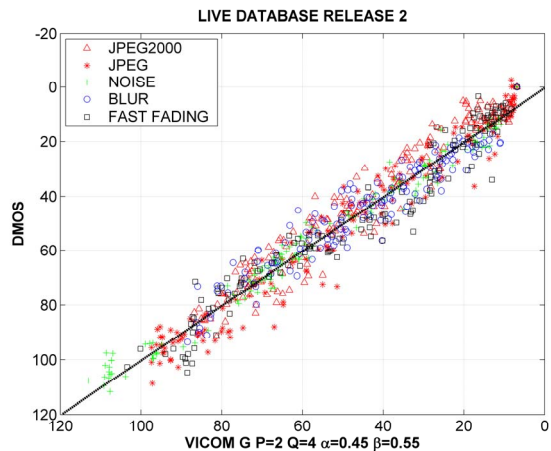


Fig. 4 The 2nd order VICOM G scatterplot for the LIVE database. RMSE=6.20, SROCC=0.974.

From the calibration view-point, parameter setting of the VICOM GL version is greatly simpler than the logistic correction required by other method, because of the pronounced linearity of the predicted DMOS with respect to the empirical DMOS

Model	SROCC	RMSE
PSNR	0.876	13.43
MSSIM	0.953	9.37
VIF	0.958	8.25
FSIM	0.963	7.67
GMSD	0.960	7.62
VICOM 6 par.	0.971	6.65
VICOM GL	0.966	7.18
VICOM G P=2 Q=4	0.974	6.20

Table I Performance of various estimators on the realigned LIVE DMOS set (779 distorted images) [2], [6], [7], [12].

Model	SROCC	RMSE
PSNR	0.842	9.50
MSSIM	0.886	8.15
VIF	0.908	7.20
FSIM	0.938	6.68
GMSD	0.935	7.47
VICOM 6 par.	0.930	6.28
VICOM GL	0.906	7.40
VICOM G P=2 Q=4	0.921	6.52

Table II SROCC performance of various estimators on the TID2008 subset [3], [6], [10].

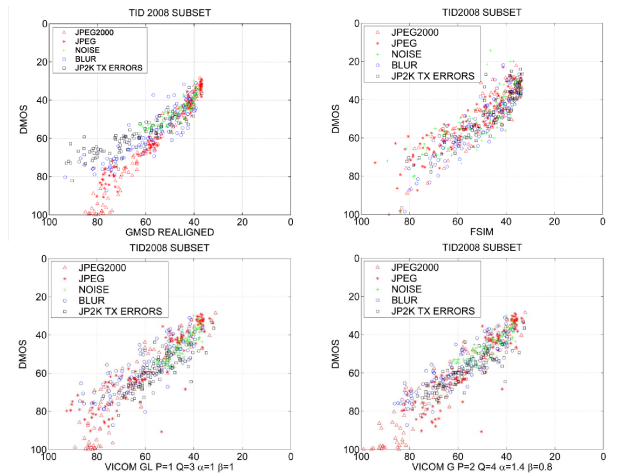


Fig. 5. The scatterplots of the linearized GSMD (upper left), the linearized FSIM (upper right), the VICOM GL (lower left, $P=1$, $Q=3$) and the 2nd order VICOM G (lower right, $P=2$, $Q=4$) for the five distortions TID2008 subset.

REFERENCES

- [1] H. R. Sheikh, A. C. Bovik, "Image information and visual quality," *IEEE Trans. on Image Processing*, vol. 15, no. 2, pp. 430-444, Feb. 2006.
- [2] L. Zhang, D. Zhang, M. Xuanqin, D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. on Image Processing*, vol. 20, no. 8, pp. 2378-2386, Aug. 2011.
- [3] L. Capodiferro, G. Jacovitti, E. D. Di Claudio, "Two-dimensional approach to full-reference image quality assessment based on positional structural information," *IEEE Trans. Image Processing*, vol. 21, n. 2, pp. 505-516, Feb. 2012.
- [4] R. Xiaonong, N. Farvardin, "A perceptually motivated three-component image model-Part I: description of the model," *IEEE Trans. on Image Processing*, vol. 4, no. 4, pp. 401-415, Apr. 1995.
- [5] S. Li, F. Zhang, Lin Ma, K. Ngi Ngan, "Image quality assessment by separately evaluating detail losses and additive impairments," *IEEE Trans. on Multimedia*, vol. 13, no. 5, pp. 935-949, Oct. 2011.
- [6] Ke Gu, G. Zhai, X. Yang, W. Zhang, "An improved full-reference image quality metric based on structure compensation," *Signal & Information Processing Association Annual Summit and Conference (APSIPAASC), 2012 Asia-Pacific*, pp.1-6, Hollywood, CA, USA, 3-6 Dec. 2012.
- [7] L. Capodiferro, E. D. Di Claudio, G. Jacovitti, F. Mangiardi, "Structure oriented image quality assessment based on multiple statistics," *Proc. of VPQM10 - Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale Plaza Resort, AZ, USA, Jan. 2010.
- [8] J. Bigun, G. H. Granlund, J. Wiklund, "Multidimensional orientation estimation with application to texture analysis and optical flow," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 775-790, Aug. 1991.
- [9] H. R. Sheikh, Z. Wang, L. Cormack, A. C. Bovik, *LIVE Image Quality Assessment Database Release 2* [Online]. Available: <http://live.ece.utexas.edu/research/quality>.
- [10] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, "TID2008: A database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelectron.*, vol. 10, pp. 30-45, 2009.
- [11] E. D. Di Claudio, G. Jacovitti, A. Laurenti, "Maximum likelihood orientation estimation of 1-D patterns in Laguerre-Gauss subspaces," *IEEE Trans. on Image Processing*, vol. 19, no. 5, pp. 1113-1125, May 2010.
- [12] H. R. Sheikh, M. F. Sabir, A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. on Image Processing*, vol. 15, no. 11, pp. 3441-3452, Nov. 2006.
- [13] W. Xue, L. Zhang, X. Mou, A. C. Bovik, "Gradient Magnitude Similarity Deviation: A highly efficient perceptual image quality index," *IEEE Trans. on Image Processing*, Vol. 23, no. 2, pp. 684-695, Feb. 2014.