Quality evaluation of motion estimation algorithms based on structural distortions

Angela D'Angelo #1, Marco Carli *2, and Mauro Barni #3

Department of Information Engineering
University of Siena, Siena, Italy
* *Applied Electronics Department*
University of Roma TRE, Roma, Italy
1 angela.dangelo@unisi.it
2 carli@uniroma3.it
3 barni@dii.unisi.it

***Abstract*—In this paper a methodology for understanding the effectiveness of motion estimation techniques is presented. Unlike other performances evaluation systems, that are based on measuring the errors between the actual and the predicted displacements, the proposed technique is inspired to the Human Visual System. More in detail, the perceptual impact of geometric distortions induced by non accurate motion estimation is considered by means of an objective measure of the perceived distortion impact. Some of the most common block-based motion estimation algorithms have been tested. For each of them the performances have been evaluated by comparing the proposed metric with the state of the art metrics. A subjective experiment has been performed to assess the effectiveness of the estimation algorithms from a perceptual point of view. The obtained results show that the scores obtained with the tested metrics generally do not match with the perceived quality, while the proposed methodology does. Therefore, the presented tool can be used in the design and in the verification of a generic motion estimation algorithm[1].**

I. INTRODUCTION

Properly modeling and estimating the apparent motion of the objects among the frames of a video is a key issue in the design and implementation of efficient video processing systems. The estimation of two dimensional (2D) motion field is used for the reduction of temporal redundancy in video compression methods, in sampling rate conversions, in filtering, or in artificial vision applications such as robotics, remote surveillance, automatic vehicle guidance, etc.

The methodology used for estimating the relative motion is highly dependent on the particular application. For example, when the 2D estimation is employed as preprocessing for 3D structure extraction, few motion vectors, computed at the vertex points of the framed objects can be enough. On the other hand, in video coding applications, the accuracy of the prediction of objects motion directly affects the redundancy reduction achieved by extracting and transmitting only the inter-frame innovation and the relevant motion data. The more accurate is the estimation, the smaller is the amount of extra information to be sent to the decoder.

[1]978-1-4244-4652-0/09/$25.00 2009 IEEE

Several techniques have been designed for estimating the displacement between two frames of a video. They can be classified into five groups, depending on the chosen approach: 1) feature/region matching methods; 2) pixel-recursive methods; 3) deterministic model-based methods; 4) stochastic model-based methods; 5) optical flow based methods. A good review of the above methods is provided in [1].

The performance evaluation of motion estimation techniques depends on the particular application. Commonly used indicators are: the computational cost, strictly depending on the number of operations needed for estimating the displacement field, and the prediction errors, that highly affect the quality of the reconstructed frames. Regarding this second aspect, to evaluate the goodness of a motion estimation algorithm, it is necessary to judge the quality of the reconstructed frames. In multimedia applications, in fact, the effectiveness of video processing/transmission is evaluated by measuring the satisfaction of the final user, that is a human being.

The classical scheme for evaluating the quality of the reconstructed frames is based on pixel-difference error metrics. In this paper we propose a new methodology, based on the characteristics of the Human Visual System (HVS), for predicting the quality of the motion field produced by a given estimation algorithm. We demonstrate that the criteria generally adopted for the quality evaluation of motion estimation algorithms are not suitable when a human is the final user of the video transmission/processing system. The motivation is that the prediction errors often result in blocking artifacts: these effects cause a degradation of the structures contained in the video that is a deformation of the shape of the objects in the scene.To this purpose, we adopt a recently introduced quality metric [2], [3] based on the use of Gabor filters, that is able to consider the impact of geometric distortions in images. This quality metric is here used in a tool for measuring the impact of non perfect reconstruction due to motion estimation errors.

The proposed evaluation scheme has been applied to several block-based motion estimation techniques, since they are commonly used, mainly due to their easy implementation, in most common video coding standards. Some state of the art metrics have been used to assess the performances of these estimation methods and their results have been compared with the proposed technique. Finally, a subjective experiment has been performed to collect human judgments. As explained in Sec. V the proposed metric is the one that better predicts the human scores.

The rest of the paper is divided as follows. In Sec. II a brief description of the motion estimation methods under investigation is provided. In Sec. III we introduce the quality metrics used for comparison, while in Sec. IV the proposed methodology is presented. In Sec. V the comparison among the scores produced by the various metrics and the scores obtained by the subjective test are reported and finally, in Sec. VI, we draw our conclusions.

II. Block based motion estimation algorithms.

Several motion compensation algorithms for interframe predictive coding have been presented in literature. In this work, we focus on block-based (or block matching) motion estimation algorithms. The idea behind block matching is to divide the current frame into macroblocks and to find, for each block, the one in the previous frame that best *matches* the current block, based on certain distance criteria. This search generates a vector representing the displacement of a macroblock between two frames. In the matching process, it is assumed that pixels belonging to the same macroblock are subject to the same displacement. Consequently, the whole moving area will be described by one single vector. This process, repeated for all the macroblocks of a frame, returns the motion field estimated for the current frame. The search area used for the matching is constrained up to p (*search parameter*) pixels around the position of the macroblock in the target frame. Faster motions require a larger p but the larger the search parameter the more computationally expensive the process of motion estimation becomes. Although this model considers only translational motion, more complex movements, like rotation and zooming, can be accounted for - assuming that they are small enough - by approximating them with a piecewise translation of the block.

Matching one macroblock with another in the target frame is based on the minimization of a cost function, called *distance measure*. The most popular and simpler cost functions are the Mean Absolute Difference (MAD), the Mean Squared Difference (MSD), and the Pixel Difference Classification (PDC) [4].

In the following we refer to the estimation of the motion vector $\mathbf{D}$ between the frame $I(x, y, t_1)$ at time t_1 (*current frame*) and the frame $I(x, y, t_2)$ at time t_2 (*target frame*). $D(x, y)$ represents the displacement of a generic point at location (x, y) from t_1 to t_2. The temporal change in luminance (or color) magnitude is the rational behind the motion estimation approach. The algorithms under test belong to the class of the block based methods [2]: Exhaustive Search (ES), Three Step Search (TSS), New Three Step Search (NTSS), Simple and Efficient Search (SES), Four Step Search (4SS), Diamond Search (DS), Diamond Search (DS), and Adaptive Root Pattern Search (ARPS).

[2]Please refer to [5] for details.

III. Motion estimation algorithms performance evaluation.

The methodology generally used for evaluating the performances of a motion estimation technique is based on two factors: the computational cost and the prediction errors. The first one is related to the number of operations needed for estimating the displacement. The second one concerns the artifacts created in the reconstruction of the target frame $I(x, y, t_2)$ based on the estimations provided by the $\mathbf{D}$ vector. To be complete, the evaluation should also consider the impact of these artifacts on the HVS. Subjective tests can adequately asses the human perceived quality; however, these tests are costly and time consuming. For this reason, several objective metrics, some of them *inspired* to the HVS, have been developed to measure the quality of the reconstructed frame. They can be classified according to the amount of original information required: Full reference (FR), Reduced Reference (RR), and No Reference (NR) where full, partial, and no information of reference data are available. The classical metrics, used to evaluate the quality of the reconstructed video sequence, are image quality metrics applied to video content frame by frame. Even if they produce reliable scores when applied to still images, they can provide totally wrong indications for video applications since they do not consider the main feature of a video: the temporal behavior. In this work we have considered three state of the art metrics: the average Peak Signal to Noise Ratio (PSNR), the SSIM index [6], and the Video Quality Metric (VQM) [7] developed by the ITS/NTIA.

The ability of these metrics to evaluate the quality of the predicted video sequence is compared to the proposed methodology, in the following called Gabor metric, which will be detailed in the next section. In the following a brief summary of the adopted metrics is reported.

A. PSNR

The most used metric in image processing applications is the PSNR. This FR metric is based on the Mean Squared Error (MSE) and it is defined as follows:

$$MSE = \frac{1}{N \times M} \sum_{x=1}^{N} \sum_{y=1}^{M} (I(x, y) - \hat{I}(x, y))^2,$$

$$PSNR = 10 \log_{10} \frac{L^2}{MSE},$$

where $N \times M$ is the number of pixels in the image, and $I(x, y)$ and $\hat{I}(x, y)$ are the $(x, y)^{th}$ pixels in the original and in the distorted signals. When dealing with video, these signals are the corresponding original and predicted frames.

L is the dynamic range of the pixel values. As it can be easily demonstrated, PSNR is not always well correlated with the quality perceived by a human observer; nevertheless, it is commonly used for both image and video quality assessment, due to its low computational cost.

B. SSIM

The SSIM index is a widely used system for measuring the similarity between two images by comparing the local patterns of luminance and contrast normalized pixel intensities. This approach considers image degradations as the perceived changes in structural information. The SSIM index between two images I and $\hat{I}$ is defined as follows:

$$\mathrm{SSIM}\left(I,\hat{I}\right)=\frac{\left(2\mu_I\mu_{\hat{I}}+c_1\right)\left(2\mathrm{cov}_{I\hat{I}}+c_2\right)}{\left(\mu_I^2+\mu_{\hat{I}}^2+c_1\right)\left(\sigma_I^2+\sigma_{\hat{I}}^2+c_2\right)} \quad (1)$$

where $c_1=(k_1L)^2$, $c_2=(k_2L)^2$, L is the dynamic range of the pixel-values, k_1 and k_2 are two constants equal to 0.01 and 0.03 respectively and μ, cov and σ are the mean, the covariance and the standard deviation functions, respectively. The SSIM index is able to predict the HVS quality assessment for certain kind of distortions. Thanks to its effectiveness the SSIM index has been often applied to the frame based video quality evaluation.

C. VQM

PSNR and SSIM have been designed for still image quality assessment. Therefore they do not consider the main peculiarity of a video: the motion factor. Even if a quality ranking can be obtained, it will be difficult to exploit the annoyance of temporal artifacts introduced by subsampling, format conversion, wrong motion displacement estimation, and so on.

VQM is a standardized method to objectively assess the quality of a video closely predicting the subjective quality ratings that would be obtained from a panel of human viewers. The VQM system extracts some parameters from the processed and the original video sequences and then compares the features. The extraction is performed by applying a perceptual filter, dividing the video sequence into spatial-temporal regions and extracting the needed parameters. The considered features can be classified as: features based on spatial gradients (the amount of perceivable edge distortion or the annoyance of horizontal or vertical edges introduced by the processing), features based on chrominance information (the frame is represented in Y, Cb, Cr color space; more perceptual weight is given to the Cr component), features based on contrast information (blurring and added noise are considered), and features based on absolute temporal information (distortions in the flow of motion due to dropped or repeated frames and added noise).

IV. Proposed Framework

In this contribution we propose the use of a new metric to assess a particularly important feature affecting the human perception: the geometric distortion. In a recent work [2] a full-reference method of objectively assessing the perceptual quality of geometric distortions in images has been described. This quality metric has been introduced in the field of digital watermarking to investigate the perceptual quality impact of geometric attacks on the watermarked images. However the same methodology can be applied to several applications in different image processing fields.

The basic ideas of the metric is that the main function of the HVS when looking at an image is to extract structural information from the viewing field. Therefore a measurement of structural distortions should be a good approximation of perceived image distortion, since the more distortion affects the structure of the objects in the visual scene, the more the corresponding degradation is visible and annoying.

In this work we apply these considerations for the quality evaluation of the motion estimation algorithms cited in Sec. II. In fact, wrong predictions result in blocking artifacts that degrade the structures contained in the frame, as it can be noticed in Fig.1. This effect is particularly evident for classical block matching algorithms, however it is still present in deformable block matching algorithms [8] or in mesh-based algorithms [9].

Fig. 1. Original frame (on the left) and compensated frame (on the right): the prediction errors result in a loss of the structural information on the image (noticeable in the hat of the foreman).

It is well known that human vision is sensitive to bars and edges [10], for this reason structures of objects in images are typically outlined by edges and bars. Hence, we expect that a measure that links the motion vector of each macroblock with the presence of edges and bars in the current frame is likely to provide an adequate measure of the perceived quality.

In [2] Gabor filters were applied to the image for extracting bar and edges information and to use these features to evaluate the perceptibility of the distortions. Two-dimensional Gabor functions were firstly proposed to model the spatial summation properties of the receptive fields of simple cells in the visual cortex. These filters are characterized by optimal localization properties in both spatial and frequency domain [11].

A 2D Gabor kernel can be mathematically defined as:

$$\mathrm{Gaborf}_{\lambda,\theta,\sigma,\gamma,\varphi}\left(x,y\right)=e^{\left(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2}\right)}\cos\left(2\pi\frac{x'}{\lambda}+\varphi\right) \quad (2)$$

where:

$$x'=x\cos\theta+y\sin\theta$$
$$y'=-x\sin\theta+y\cos\theta$$

λ (whose value is specified in pixels) is the wavelength of the cosine factor of the Gabor filter kernel, φ is the phase offset of the cosine factor, θ specifies the orientation of the normal to the parallel stripes of the filter, γ describes the ellipticity of the support of the Gabor function and σ is the standard deviation of the Gaussian factor of the function. The value of σ cannot be specified directly, it can only be changed through the half-response spatial frequency bandwidth b (default is $b = 1$, in which case σ and λ are connected as follows: $\sigma = 0.56\lambda$).

For $\varphi = 90$ degrees (or -90) the filter in Eq.(2) deploys an antisymmetric Gabor function and gives a maximum response at an edge. A symmetric Gabor function ($\varphi = 0$ or 180 degrees) can be used for the bar detection.

For filter design Daugman in [11] suggests the widely adopted parameters $\gamma = 0.5$ and $b = 1$. Using these values, the resulting filters for the edge and bar extraction, respectively, are those shown in Fig. 2 (with $\theta = 0$ degrees and $\lambda = 10$ pixels).

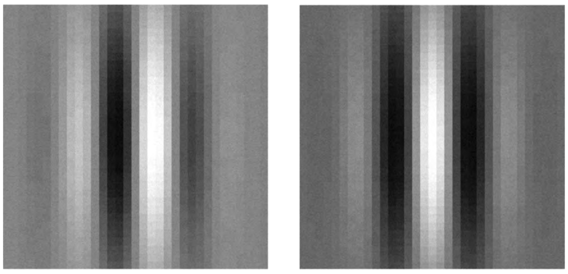


Fig. 2. Filters for the edges (a) and bars (b) detection with $\gamma = 0.5, b = 1, \theta = 0$ degree, $\lambda = 10$ pixel

Once defined the filter parameters, and set a particular θ, the function described in Eq. (2) is used to filter each frame of the original video sequence and to find edges and bars in the direction orthogonal to θ. The filtering function is described by the following equation:

$$If_\theta(x,y) = \sqrt{If^2_{\theta,\mathrm{bar}}(x,y) + If^2_{\theta,\mathrm{edge}}(x,y)} \qquad (3)$$

where If_θ is the filtered frame and $If_{\theta,\mathrm{edge}}$ and $If_{\theta,\mathrm{bar}}$ are the original frame convolved with the Gabor filters described by Eq. (2) with $\varphi = -90$ and $\varphi = 0$ respectively.

The quality of the compensated video is measured at three levels: the local region level, the frame level, and the video sequence level.

To obtain the score associated to the perceivable degradation for each pixel in the frame, we link edges and bars of the frame in the original video with the motion vector $\mathbf{D}$ of the corresponding macroblock in the processed video. Specifically, we first consider the vector $\mathbf{D}_\theta$ (the projection of $\mathbf{D}$ along θ), orthogonal to bars and edges of the original frame. Then, to estimate the loss of structure in the compensated frame, we evaluate the gradient of $\mathbf{D}_\theta$ with respect to the direction orthogonal to θ. The local score associated to each pixel, quantifying the perceived degradation in that pixel, is defined by the following equation:

$$\mathrm{Score}(x,y) = \sum_{\theta \in S} If_\theta(x,y) \left(\frac{\partial D_\theta}{\partial d_\theta^\perp}(x,y) \right) \qquad (4)$$

where $If_{\theta(x,y)}$ is the filtered frame described by Eq. (3) in the θ direction and the notation $\frac{\partial D_\theta}{\partial d_\theta^\perp}$ indicates the gradient of the motion vector in the θ direction with respect to the direction orthogonal to θ. The summation over θ in Eq. (4) is needed to take into account the salient features along different orientations, specifically $S = \{0, \frac{\pi}{2}\}$.

At the second level of quality evaluation, the local quality values are combined into a frame-level score by summarizing the scores obtained for each pixel, as follows:

$$\mathrm{Obj}(i) = \sum_{x,y} \mathrm{Score}\,(x,y) \qquad (5)$$

To produce an objective quality score with the same range as the objective scores, and to account for the saturation effect typical of the HVS, a fitting curve is applied to the global quality score in Eq. (5).

We decided to use the Weibull cumulative distribution function defined as follows:

$$\mathrm{Gabor}(i) = -4 \left(1 - e^{-\left(\frac{\mathrm{Obj}(i)}{\zeta_1}\right)^{\zeta_2}} \right) + 5 \qquad (6)$$

where $\mathrm{Gabor}(i)$ is the quality measure of the i-th frame in the video sequence, ζ_1 and ζ_2 are the Weibull parameters defined in [2]. In the performed experiment $\zeta_1 = 0.01445$ and $\zeta_2 = 0.4141$. By using the fitting function it is possible to obtain a perceptual quality score, in the range 1 to 5, quantifying the perceived quality by the viewer observing the compensated frame (with 1 corresponding to a bad image quality and 5 to an excellent image quality). This fitting function provided the best fit for our data among the commonly used curves, i.e., Gaussian, logistic, and Weibull (please refer to [2] for details).

Finally, in the third level, the overall quality of the entire processed video sequence is given by:

$$\mathrm{Gb} = \frac{1}{\mathrm{F}} \sum_{i=1}^{F} \mathrm{Gabor}\,(i) \qquad (7)$$

where F is the number of frames.

V. Experimental results

Several tests have been performed to validate the proposed methodology. In the following the results we obtained are presented and discussed. In our test we used four QCIF videos (Carphone, Foreman, Coastguard, and Container) generally employed in video processing testing. Those video are characterized by different motion rates, different content, and varying regions of interest. The seven block based motion estimation algorithms cited in Sec. II have been used. For each test, the size of the macroblock was $MB = 16$ pixels and the search window size $p = 7$.

We are interested in evaluating how the artifacts created by each motion estimation method are detected by the previously described metrics: PSNR, SSIM, VQM, and the proposed one, the Gabor metric defined by Eq. (7). We would like to stress that the purpose of the performed tests and more generally of this work is not to provide a comparison among the motion estimation algorithms under investigation, but to show the inability of state-of-the-art metrics to evaluate the effectiveness of the methods from a quality point of view. The choice to use more than one estimation method is to prove that the obtained results are independent from the particular adopted algorithm.

A. Visual inspection

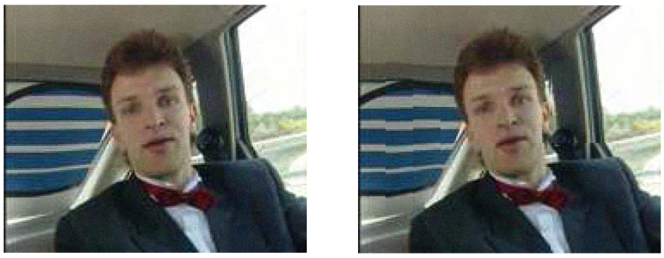
The first comparison of the performances of the estimation methods can be made by means of visual inspection of the predicted frames. To this aim, in Fig. 3 two frames of the Carphone video sequence obtained by using the ES method are reported. As it can be noticed, even by using the most accurate estimation method, some artifacts are created. This is evident in Fig. 3.(a) in the region containing the horizontal blue bars in the background. These artifacts are less annoying in Fig. 3.(b), even if they are still present in the right part of the image, among the leafs of the trees. The different perception of the image quality is due to the fact that in the frame on the top the artifacts cause a degradation of the structures in the image.

The scores provided by the quality metrics are reported in the figure caption. A lower value of PSNR indicates a lower quality for the predicted image (b). The same results are obtained with VQM and SSIM metrics: the frame in Fig. 3.(b) presents worst scores than the one in (a). The only metric that is able to catch the presence of the misalignment in the horizontal bars and that is able to consider the importance of incongruence in horizontal lines is the Gabor metric.

This simple example shows a mismatch between the subjective judgment and the objective scores of the three classical metrics. Even if this fact already demonstrates the effectiveness of the proposed methodology, it can not be generalized due to two motivations: i) this comparison should be performed and validated for each couple of frames for all the video under investigations and for all the motion estimation methods, and, most of all, ii) the overall quality of a video is not strictly related to the quality of the single frames. Several aspects influence the overall quality as motion homogeneity, saliency of objects in some shots of the video, the particular application, and so on. To obtain a real assessment of the artifacts annoyance on a human being, a subjective experiment has been performed.

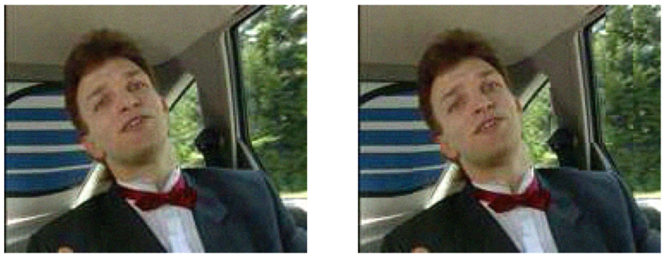
B. Subjective tests

The subjective scaling method we used is the Double Stimulus Impairment Scale (DSIS)[12], which is the most suitable system for collecting the artifact visibility threshold level. In fact, even if a strict definition does not exist, the methods that use explicit references should be used when testing the fidelity of transmission with respect to the source signal (this is frequently an important factor, for example, in the evaluation of high quality systems).



PSNR = 30,15 db, SSIM = 0.8802, VQM= 2.7828, Gabor = 1.8505

(a)



PSNR = 23,14 db, SSIM = 0.8416, VQM = 3.400, Gabor = 2.0143

(b)

Fig. 3. Quality metrics comparison: original frame on the left, compensated one on the right; (a) frame n. 5, (b) frame n. 298

Videos are shown in pairs: the reference and the impaired video. After their playback, the expert is asked to give his/her opinion by using a five level impairment scale: 1) very annoying, 2) annoying, 3) slightly annoying, 4) perceptible but not annoying, 5) imperceptible.

The experience in the assessment of high quality video images during the activities of ITU-R SG6 Task Group 6/9 suggested the usage of expert other than naive viewers. This experience allowed to conclude that the assessment by a group of experts is highly reliable and able to provide results as reliable and stable as those obtainable by performing a standard subjective test. This conclusion led to the approval of the recommendation dedicated to the expert viewing described in [13]. Following this recommendation, an expert viewing test with five experts was performed, allowing to considerably reduce the total length of the test.

Fig. 4 shows the collected Mean Opinion Score (MOS) values versus the objective scores obtained with PSNR (a), SSIM (b), VQM (c), and Gabor (d) metrics, averaged for the 28 videos under test. From the analysis of the plots in (a),(b), and (c) it is clear that no evident relation exists between the objective and subjective scores. It is almost impossible to find any relation between the MOS and the relative metric score. This means that the objective metrics can not be used to predict the humans response. The correlation between the objective data given by the Gabor metric and the users response is, instead, evident in the plot in Fig. 4.(d).

To provide quantitative measures on the performance of the proposed metric, we followed the performance evaluation

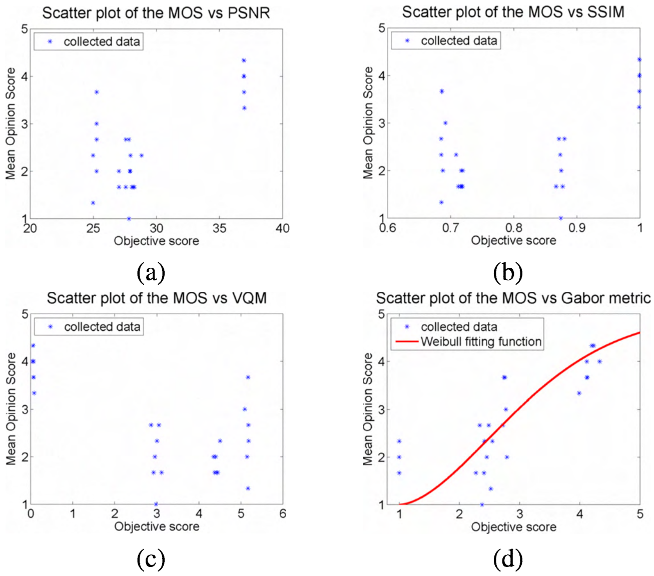


Fig. 4. Mean Opinion Score collected by subjective experiments versus the objective score obtained with PSNR (a), SSIM (b), VQM (c), and Gabor (d) metrics, averaged for the 28 videos under test.

procedures employed in the VQEG Phase I FR-TV test[3]. The relationship between objective data and the subjective ratings was estimated by using a nonlinear regression, the Weibull function described by the solid plot in Fig. 4(d). Then, the objective metric was evaluated through different performance attributes, applied on the fitted values, as specified in the report: 1) the Pearson linear correlation coefficient between the objective/subjective scores and the RMSE (root MSE) between MOS and MOSp (MOS predicted), that are measures of the prediction accuracy of a model; 2) the Spearman rank-order correlation coefficient between the objective/subjective scores that is considered a good evaluation of prediction monotonicity; 3) the outlier ratio (percentage of the number of predictions outside the range of twice the interval of confidence at 95%) of the predictions after the nonlinear mapping, which is a measure of prediction consistency. The results we obtained are summarized in table I. The high value of both Pearson and Spearman correlation coefficients, demonstrate the effectiveness of the proposed scheme.

	Gabor metric
Pearson correlation coefficient	0.8052
RMSE MOS - MOSp	0.6767
Spearman correlation coefficient	0.8005
Outlier ratio	0

TABLE I
PERFORMANCE OF THE PROPOSED GABOR METRIC

VI. Conclusions and Future Works

In this paper a novel method for evaluating the effectiveness of motion estimation methods is presented. The proposed method is tuned to the HVS by considering the perceptual impact of geometric distortions induced by non-accurate motion estimation. As demonstrated by a subjective experiment, this methodology better matches the quality perceived by human beings with respect to state of the art metrics. The proposed tool can be used both in the design and in the verification of a motion estimation algorithm. The promising results obtained till now push for further studies. Different motion estimation algorithms (i.e. mesh based, deformable regions based, etc.), different estimation settings (i.e. MS size, p-search window, frame rate, etc.) can be studied to verify the applicability and generalization of the proposed method to evaluate different types of artifacts other than blocking artifacts. Furthermore, from a theoretical point of view, an extension of the Gabor based metric to video quality needs to be investigated to incorporate the temporal behavior.

[3]http://ftp.crc.ca/test/pub/crc/vqeg/phase1_obj_test_plan.rtf

Acknowledgment

The research described in this paper has been partially funded by the Italian Ministry of Research and Education under FIRB project no. RBIN04AC9W.

References

[1] Y. Q. Shi and H. Sun, *Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards (Image Processing Series)*. CRC Press, 1999.

[2] A. D'Angelo and M. Barni, "A structural method for quality evaluation of desynchronization attacks in image watermarking," in *2008 IEEE 10th Workshop on Multimedia Signal Processing*, 2008, pp. 754–759.

[3] ——, "Multiresolution quality evaluation of geometrically distorted images," in *IV International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2009.

[4] C. Bowling and R. Jones, "Motion compensated image coding with a combined maximum a posteriori and regression algorithm," *IEEE Trans. on Communications*, vol. 33, no. 8, pp. 844–857, Aug 1985.

[5] B. Furht and B. Furht, *Motion estimation algorithms for video compression*. Kluwer Academic Publishers Norwell, MA, USA, 1996.

[6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, 2004.

[7] S. Wolf and M. Pinson, "Video quality measurement techniques," NTIA Report 02-392, Tech. Rep., 2002.

[8] O. Lee and Y. Wang, "Motion-compensated prediction using nodal-based deformable block matching," *Journal of Visual Communication and Image Representation*, vol. 6, no. 1, pp. 26–34, March 1995.

[9] P. Van Beek, A. Tekalp, and A. Puri, "2-d mesh geometry and motion compression for efficient object-based video representation," in *Int. Conf. on Image Processing*, 1997.

[10] B. Wandell and E. Simoncelli, "Foundations of Vision," *Journal of Electronic Imaging*, vol. 5, p. 107, 1996.

[11] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, vol. 2, no. 7, pp. 1160–1169, 1985.

[12] [ITU (2002)], "Methodology for subjective assessment of the quality of television pictures," Rec. BT.500-11. International Telecommunication Union, Geneva, Switzerland., Tech. Rep., 2002.

[13] [ITU R], "Expert viewing methods to assess the quality of systems for the digital display of lsdi in theatres," Rec. BT.1663. International Telecommunication Union, Geneva, Switzerland., Tech. Rep., 2003.