A full-reference quality metric for geometrically distorted images

Angela D'Angelo*, Li Zhaoping, and Mauro Barni, *Senior Member, IEEE*

Abstract

In multimedia applications there has been an increasing interest in the use of quality measures based on human perception, however, research has not dealt with distortions due to geometric transformations. In this paper we propose a method to objectively assess the perceptual quality of geometrically distorted images, based on image features processed by human vision. The proposed approach is a full-reference image quality metric focusing on the problem of local geometric distortions and is based on the use of Gabor filters that have received considerable attention because the characteristics of certain cells in the visual cortex of some mammals can be approximated by these filters. The novelty of the proposed technique is that it considers both the displacement field describing the distortion and the structure of the image. The experimental results show the good performances of the proposed metric.

Index Terms

Geometric distortions, Human Visual System (HVS), Perceptual quality, Image quality assessment.

I. INTRODUCTION

Digital images are subject to various kinds of distortions that may result in a degradation of visual quality during acquisition, processing, compression, storage, transmission and reproduction. It is therefore necessary for many applications to be able to quantify the image quality degradation that occurs in a system, so that it is possible to control and enhance the quality of the images it produces.

The best way of assessing the quality of an image is through subjective evaluation because in most cases human eyes are the ultimate receivers. The Mean Opinion Score (MOS) [1], which provides a numerical

A. D'Angelo is with the Department of Information Engineering, University of Siena, Italy (e-mail: angela.dangelo@unisi.it)

L. Zhaoping is with the Department of Computer Science, University College of London, United Kingdom (e-mail: z.li@ucl.ac.uk)

M. Barni is with the Department of Information Engineering, University of Siena, Italy (e-mail: barni@dii.unisi.it)

indication of the perceived quality of a media and is obtained from a number of human observers, has been used for many years; however, the MOS is tedious, quite expensive in terms of time and human resources. Furthermore the subjective results depend on several external factors such as the observer's background, motivation, etc.

The goal of objective image quality assessment is to design a quality measure based on the human vision model that can predict the perceived image quality automatically. The objective measure should give a numerical value quantifying the dissatisfaction of a typical human viewer when he/she observes the reproduced image in place of the original.

Despite the existence of several studies on human perception of image quality, only few works can be found in the scientific literature dealing with the quality assessment of geometrically distorted images and the limitation of the few metrics proposed so far is that they only rely on the displacement field defining the distortion without taking into account the characteristics of the images. This is a relevant problem since the same distortion applied to two different images can result in a different perceived image quality.

In the last years the problem of assessing the quality of geometrically distorted images has received an increasing attention from the watermarking community, due to the central role that such distortions play in watermarking theory. As a matter of fact, the application of a geometric distortion to a watermarked image causes a de-synchronization between the watermark embedder a detector that in most cases prevents the correct extraction of the watermark [2]. All the more, that in most cases the geometric distortion is less annoying than other kind of distortions like noise addition, blurring or lossy compression. This is especially true for local or spatially varying distortions, for which an exhaustive search of the watermark is unfeasible due to the huge size of the search space. A first step to solve the problems with geometric attacks is the characterization of the class of perceptually admissible distortions, defined as the class of geometric distortions whose effect can not be perceived, or is judged acceptable, by a human observer. It is clear that the availability of an objective quality metric capable of dealing with geometric distortions would be of invaluable help in this sense.

Watermarking is not the only field where an image quality metric for geometrically distorted images would be needed. Interested applications include registration of biomedical images that usually requires the application of local and nonlinear transformations [3], whose strength should be constrained to satisfy certain quality constraints, collusion-secure fingerprinting techniques by random pre-warping [4], the problem of recovering 3D models from uncalibrated images of architectural scenes [5].

The goal of this paper is to introduce an objective quality metric for geometrically distorted images, which has a good correspondence to the human perception of the distortions. Specifically, we focus our

attention on the problem of local geometric transformations. Global transformations, in fact, usually do not affect image quality at all[1]or introduce a degradation proportional to the parameters defining the distortion. This is the case of the application of two different scaling factors to the vertical and horizontal dimensions of an image, for which the ratio between the two scaling factors is directly linked to the perceived distortion. Moreover in some applications, like digital watermarking, it is more difficult to deal with local transformations with respect to global distortions for which many watermarking schemes have been proposed during the last years [6].

The basic idea of our approach is that human vision is sensitive to structures in images, thus a measurement of structural distortions should be a good approximation of the perceived image distortion. In order to provide such a measurement, we need to analyze how the displacement field describing the distortion affects the structure of the image from a perceptual point of view. In our approach we identify image structures with edges and bars and we use Gabor filters to detect them since these filters provide better results with respect to classical features detectors such as Robert Cross or Sobel operators. The use of Gabor filters is also justified by the consideration that their shapes are quite similar to receptive fields found in V4 [7], a visual cortical area of the primary visual cortex. Finally we link bars and edges in the image to the displacement field describing the geometric distortions.

In doing so, we adopt a *full-reference* approach since we assume that the displacement field describing the distortion is known. If the original image is available, the displacement field can be estimated by many available techniques like image registration, optical flow or motion estimation. Note that is some cases, the estimation of the displacement field may not be necessary, since the the purpose of the system is just to build a displacement field whose application does not degrade significantly the perceived image quality. This is the case, for instance, in image registration applications, where the displacement field used to register the image at hand on a target image is the final outcome of the registration process, or in digital watermarking, where the goal of the attacker is to build a displacement field that does not impair image quality, but it is strong enough to de-synchronize the detector. The main novelty of the proposed quality metric is that it relies on both the displacement field describing the distortion and the structure of the image.

The paper is divided into two parts. In the first part we describe the problem of the assessment of geometric distortions in images, and, in section III, we present the mathematical background behind the idea of the proposed method. The second part of the paper regards the design and implementation of two

[1]This is the case, for example of rotations, zooming, or translations.

user-tests to measure human perception of geometric distortions in images. The first set of experiments, explained in section IV, has been performed to set the parameters of the model. The second set is used to cross-validate, with a different dataset, the performance of the new metric. Finally, in section VI, we derive our conclusions and propose some ideas for future works.

In order to ensure the reproducibility of the experimental results, the software we used for the experiments, such as all the image databases, is available on the web site *http://www.dii.unisi.it/∼vipp/*.

II. IMAGE QUALITY ASSESSMENT

The simplest and most widely used full-reference quality metric is the mean squared error (MSE), computed by averaging the squared intensity differences of the distorted and reference image pixels, along with the related quantity of Peak Signal-to-Noise Ratio (PSNR)[2]. The assumption is that a distorted image can be seen as the sum of the reference image and an error signal and the loss of perceptual quality is directly related to the visibility of the error signal. This method, as well as all the statistics based measures, is simple to calculate, has clear physical meanings, and is mathematically convenient in the context of optimization. Unfortunately, statistics based measures often fail to reflect accurately the subjective observation since they ignore human vision characteristics. One of the most popular objective quality metrics is the SSIM (Structural SIMilarity) index [8], a full-reference method for measuring the similarity between two images comparing local patterns of pixel intensities that have been normalized for luminance and contrast.

In the last three decades, a great deal of effort has been paid to develop quality assessment methods that take into account of the characteristics of the human visual system (HVS). It is a matter of fact that the visual assessment task involves high complex psychophysical mechanisms, however, the HVS is too complex to be completely understood with current psychophysical knowledge, thus current HVS models rely on some simplified assumptions and they may not provide fully reliable results [9]. Many image quality assessment algorithms have been proposed in the last decades exploiting the knowledge

[2]The PSNR of a monochrome image is defined via the mean squared error (MSE) as follows:

$$MSE = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \|I(x,y) - Z(x,y)\|^2$$
$$PSNR = 20 \log \left(\frac{\max(I)}{\sqrt{MSE}} \right)$$

where I is the original image, Z is the image to be evaluated, m and n are the width and the height of the two images and $\max(I)$ is the maximum possible pixel value of the images.

of the HVS and some of them have been shown to behave consistently when applied to certain kinds of distortions (e.g. JPEG compression, gaussian noise, median filtering, etc.).

Among the metrics inspired to human vision characteristics, promising results have been recently reported in [10], where a reduced-reference metric for color images, called C4, is introduced. This criterion implements an operating and organizational model of the HVS, including some of the most important stages of vision (perceptual color space, CSF, psychophysical sub-band decomposition, masking effect modeling). The novelty of this approach is that it extracts structural information from the representation of images in a perceptual space. Extracted features are stored in a reduced description which is generic as it is not designed for specific types of distortions. The results reported in [10] indicate that C4 provides much better performances than classical RMSE or PSNR methods and that it competes favorably with state-of-the-art criteria like UQI [11] and SSIM which are full reference quality criteria.

Unfortunately, as we will see in section V, the effectiveness of this metric, such as that of statistically-flavored measures, degrades when dealing with images affected by geometric distortions.

Let us introduce now a definition of geometric distortions. In general, a geometric distortion can be seen as a transformation of the position of the pixels in the image. Let I be an original image, neglecting the border effects the distorted image Z is obtained by assigning to each pixel $I(x, y)$ a displacement vector $D\left(x, y\right) = \left(D_h\left(x, y\right), D_v\left(x, y\right)\right)$, where $D_h(x, y)$ and $D_v(x, y)$ are, respectively, the horizontal and vertical displacements. In the following D is called displacement field, depending on how it is produced it is possible to distinguish between global and local geometric distortions.

A global transformation is defined by a mapping function that relates the points in the input image to the corresponding points in the output image. The mapping function is defined by a set of operational parameters and applied to all the image pixels, that is, the same operation, under the same parameters, affects all the image pixels. An image rotation, for example, is uniquely defined by the rotation angle and all the pixels in the image rotate by the same angle (other examples of global geometric distortions are zooming, translation, affine transformation, etc.).

On the other hand, local distortions refer to transformations affecting in different ways the position of the pixels of the same image or affecting only part of the image. Fig. 1 shows two examples of local geometric distortions applied to the Barbara image (the first image on the left) with a different perceived image quality: the image on the right has a good quality thanks to the smoothness of the displacement field, while the image in the center presents a very annoying distortion.

Geometric distortions and geometric correction (or image registration, that is the procedure that corrects spatial distortions in an image), play a central role in different fields of image processing. In the context of

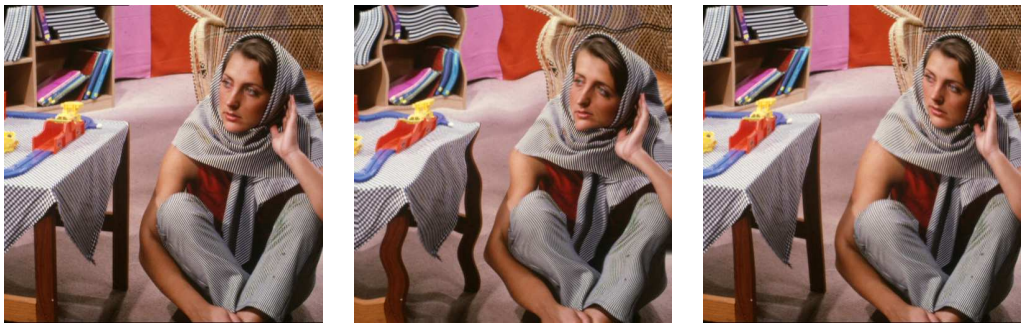


Fig. 1. Local geometric distortions applied to Barbara standard image (the original image is on the left).

medical imaging, for example, great attention has been paid to study and correct the distortions that affect magnetic resonance imaging in the radiotherapy [12]. In the same way, geometric distortions introduced by the Landsat imaging system are well-known [13]: the satellite does not provide information with sufficient accuracy to determine the geographical position of each pixel thus, to map the data onto the desired space, it is necessary to model a geometric transformation from a set of ground control points.

Lately, local geometric distortions have been extensively studied in the field of digital watermarking, as already explained in the introduction. The use of digital watermarking in real applications, in fact, is impeded by the weakness of current available algorithms against signal processing manipulations leading to the de-synchronization of the watermark embedder and detector. For this reason the problem of watermarking under geometric attacks has received considerable attention. Specifically it could be very useful to know the amount of geometric distortions that can be applied before the distorted image loses its commercial value or its meaning (this happens if the distortion affects too much the image quality) in order to develop an ad-hoc decoding algorithm and eventually obtain watermark synchronization through exhaustive search [14].

In all the above applications, an objective metric capable of assessing the quality of geometrically distorted images is missing. In fact, only few works can be found in literature regarding the problem of the assessment of geometrically distorted images. A simple measure, proposed by Licks et al. [15] and Baüml et al. [16], is based on the evaluation of the variance of the sampling grid jitter. However this method does not take into account the spectral features of the jitter on which the perception of the transformation depends. This means that two jitter noises with completely different spectral characteristic but with the same variance will be evaluated in the same way.

The method proposed by Desurmont et al. [17] is based on the average gradient of the sampling

grid transformation. This method does not work in presence of geometric distortions based on local permutations of the position of the pixels in the image, as the LPCD model proposed in [18].

Setyawan et al. [19] proposed an objective quality measurement scheme based on the use of simple transformation models, for example RST (rotation, scaling and translation) or affine transform, to approximate the underlying complex geometric transform. The approach is based on the assumption that a complex geometric transformation applied on a global scale can be approximated by simpler transformations applied on a local scale and the visual quality of a distortion is determined by the degree of homogeneity of the geometric transformations applied locally. The less homogenous the distortions, the worse the visual quality will be. This method is very expensive from a computational point view due to the optimization involved in its computation. In their paper the authors set to 32 pixels the minimum block size: the choice of this parameter is a trade-off between the precision and the reliability of the approximation, as block sizes that are too small will make the approximation less reliable.

In [20] D'Angelo et al. proposed an objective metric based on the theory of Markov Random Fields. The geometric distortion can be described as a Markov Random Field defined on the set of the image pixels and the displacement field generating the distortion is a possible configuration of the field. The proposed metric relies on the assumption that the potential function of the Markov Random Field describing the distortion gives an indication of the perceptual degradation of the distorted image.

The limitation of all the above works, is that they rely only on the displacement field defining the distortion without taking into account the characteristics of the images, that is, the same distortion applied to different images returns the same value of the objective metric. This limitation can be easily understood by looking at the images in Fig. 2. These are two distorted images generated by using the same displacement field. In this example all the metrics discusses above would return the same score, however the visual quality of the two distorted images is drastically different since the distortion in the image on the right is more annoying. Consequently, an efficient image-quality measure would need to be able to consider the structural information in the image.

III. Overall architecture of the proposed metric

Images are highly structured since their pixels exhibit strong dependencies, especially when they are spatially close, and these dependencies carry important information about the structure of the objects in the visual scene. It is well known that one of the main functions 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. Based on this idea, the overall

Fig. 2. Two distorted images generated by using the same displacement field: the image on the right is more annoying.

architecture of the proposed metric is described in Fig. 3.

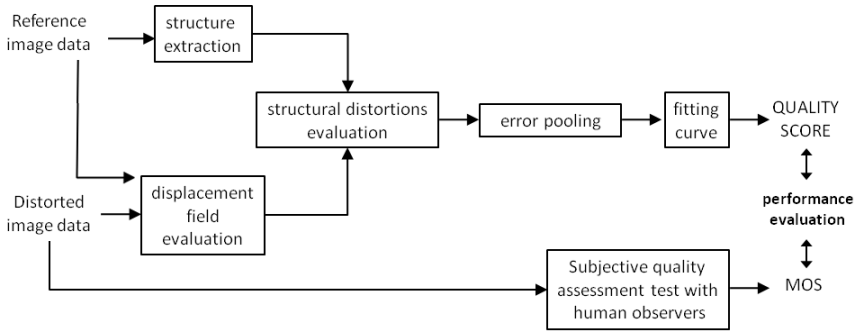


Fig. 3. Full process of the objective quality criterion

Given an original image and the corresponding distorted version, in order to evaluate the perceived image quality, the displacement field is first estimated (*displacement field evaluation*)[3]. At the same time a set of features describing the structures contained in the image are extracted (*structure extraction*). The impact of the displacement field on the extracted structural features is then evaluated (*structural distortions evaluation*) providing a local quality score. The local scores are then pooled to quantify the overall perceived image quality (*error pooling*). Finally a fitting curve is applied with the twofold aim of accounting for saturation effects typical of human quality judgement and to obtain an objective quality score with the same range as the subjective scores (*fitting curve*).

[3]As we already noted, in some applications the displacement may be known in advance, hence this step can be skipped.

The objective quality score may then be compared to the MOS, that is the scores given by human observers (*subjective quality assessment test with human observers*).

By inspecting the scheme reported in Fig. 3, the full-reference nature of the proposed metric clearly comes out. In particular, the availability of the original and the distorted images is exploited to recover the displacement field describing the distortion. Note however, that the choice of the algorithm used to estimate the displacement field is not fixed, hence leaving the possibility of tailoring such a step to the application at hand. For this reason, and to develop a metric that is independent of the effectiveness of the displacement estimation step, in the next two sections we assume that the displacement field generating the distortion is known. We will evaluate the impact of this step experimentally, in Sec.V-C.

A. The main idea

Before describing the details of all the steps described in Fig. 3, we explain the main rationale behind the proposed metric, i.e. the approach we chose to measure how the displacement field modifies the structural information contained in the image.

Psychophysical studies show that human vision is sensitive to edges and bars in images, and structures of objects in images are typically outlined by edges and bars [21]. Hence, we expect that a measure that links the geometric distortions to the presence of edges and bars in the image is likely to provide an adequate measure of image quality. Based on this considerations, the idea behind our metric is that a geometric distortion causes a degradation of the structure of the objects in the visual scene when the displacement field describing the distortion is orthogonal to the direction of the image bars and edges.

In the following this idea is exemplified by means of examples applied to the synthetic image shown in Fig. 4. This image is well suited to show our approach due to the presence of several bars and edges.

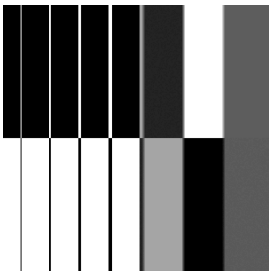


Fig. 4. Sample synthetic image.

For the sake of simplicity let us consider a geometric distortion described by a horizontal displacement field (the vertical displacement field is set to zero), that is each pixel in the image is assigned a

displacement vector with an horizontal direction and a specific magnitude. Let us now evaluate the perceived degradation when applying this displacement field to images with differently orientated features. Applying the distortion to an image with several vertical features (Fig. 5.(a)) results in a very annoying distortion (Fig. 5.(b)), while the distortion is almost imperceptible, except for the border effects (Fig. 5.(d)) when the displacement field is applied to an image containing only with horizontal structures (Fig. 5.(c)).

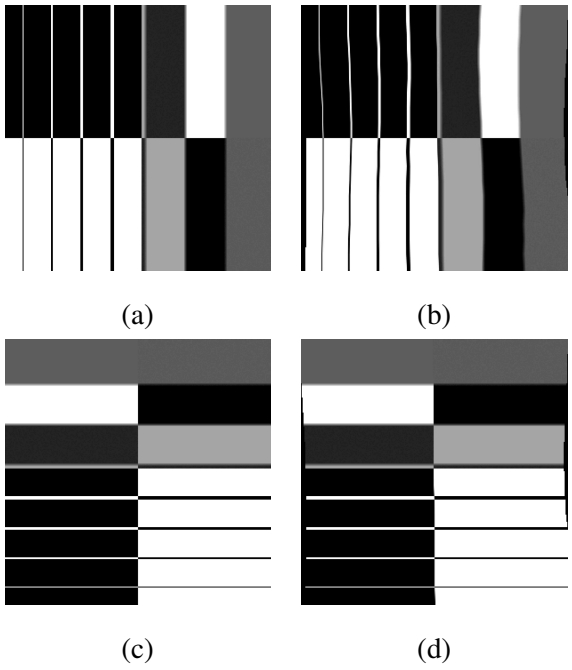


Fig. 5. Example of perceptibility of a geometric distortion: (a) original image; (b) application of the horizontal displacement field on the original image; (c) rotated image ; (d) application of the horizontal displacement field on the rotated image.

The above considerations guided us in the definition of the *structure extraction* and the *structural distortions evaluation* modules in Fig. 3.

B. Structural feature extraction

We decided to use Gabor filters to extract bar and edges information from the images and to use these features to evaluate the perceptibility of the distortions.

Two-dimensional Gabor functions were firstly proposed by Daugman [22] to model the spatial summation properties of the receptive fields of simple cells in the visual cortex. These filters have been shown to posses optimal localization properties in both spatial and frequency domain and thus are well

suited for image processing applications. Gabor filters have been used in many applications, like texture segmentation, target detection, fractal dimension management, document analysis, edge and bar detection, retina identification, image coding and image representation. In the context of edge and bar detection, typical detectors such as the Roberts Cross or Sobel operators, which measure the gray-scale gradient component in a given orientation, cannot discriminate between edges or bars while this is possible by using symmetric and antisymmetric Gabor filters [23] (as we will show in the following).

A Gabor filter can be viewed as a sinusoidal plane wave of particular frequency and orientation, known as *carrier*, modulated by a 2D Gaussian shaped function, known as *envelope*. It is generally described by the following equation:

$$g_{\lambda,\theta,\varphi,\sigma,\gamma}(x,y) = e^{\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)} \cos\left(2\pi\frac{x'}{\lambda}+\varphi\right) \tag{1}$$

where:

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

A Gabor filter is uniquely defined once the $\lambda, \theta, \varphi, \sigma$ parameters are known. λ is the wavelength of the cosine factor of the Gabor filter kernel, θ specifies the orientation of the normal to the parallel stripes of a Gabor function, φ is the phase offset and the aspect ratio γ specifies the ellipticity of the support of the Gabor function. The half-response spatial frequency bandwidth b of a Gabor filter is related to the ratio $\frac{\sigma}{\lambda}$,where σ and λ are the standard deviation of the Gaussian factor of the Gabor function and the preferred wavelength, respectively, as follows:

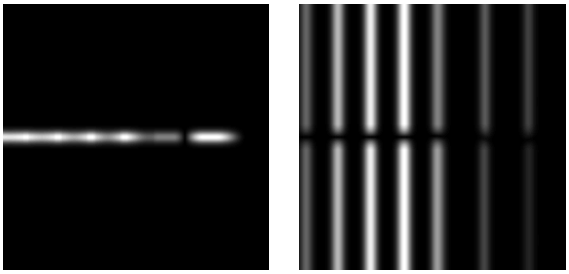
$$b = \log_2 \frac{\frac{\sigma}{\lambda}\pi + \sqrt{\frac{\ln 2}{2}}}{\frac{\sigma}{\lambda}\pi - \sqrt{\frac{\ln 2}{2}}}, \quad \frac{\sigma}{\lambda} = \frac{1}{\pi}\sqrt{\frac{\ln 2}{2}}\frac{2^b+1}{2^b-1}. \tag{2}$$

A 2D Gabor kernel for the edge detection in images, also called antisymmetric Gabor function, can be mathematically defined as:

$$\text{GaborE}_{\lambda,\theta,\sigma,\gamma}(x,y) = e^{\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)} \sin\left(2\pi\frac{x'}{\lambda}\right), \tag{3}$$

while the corresponding equation for the bar detection, or symmetric Gabor function, is:

$$\text{GaborB}_{\lambda,\theta,\sigma,\gamma}(x,y) = e^{\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)} \cos\left(2\pi\frac{x'}{\lambda}\right). \tag{4}$$



(a) (b)

Fig. 6. Example of filtered image: (a) filtered image If with $\theta = \frac{\pi}{2}$; (b) filtered image If with $\theta = 0$.

The parameters we used for the design of the filters are the following: $\gamma = 0.5$ and $b = 1$ (these are the default values in typical applications of Gabor filters). In the next section we will see how to set the value of θ and λ.

Once defined the parameters of the filters, and fixed a particular θ, we use the functions described in equations (3) and (4) to filter the original image and to find edges and bars in the direction orthogonal to θ. The filtering process is described by the following equation:

$$I_f = \sqrt{I^2_{f,\mathrm{bar}}(x,y) + I^2_{f,\mathrm{edge}}(x,y)} \tag{5}$$

where I_f is the filtered image; $I_{f,\mathrm{edge}}$ is obtained by convolving the original image with the Gabor filter described by equations (3); $I_{f,\mathrm{bar}}$ is obtained by convolving the original image with the Gabor filter described by equations (4) and then by subtracting the mean value of the Gabor-filtered image (in order to compensate the DC component).

In the scientific literature, the outputs of a symmetric and an antisymmetric kernel filter at each image point are usually combined into a single quantity as described by equation 5. This quantity is called the Gabor energy and is related to the model of a specific type of orientation selective neuron in the primary visual cortex called the complex cell [24].

Fig. 6 gives an example of the original synthetic image (see Fig. 4) filtered by using equation (5) in the two directions $\theta = \frac{\pi}{2}$ and $\theta = 0$.

C. Local distortions computation

To find the quality score associated to each pixel quantifying the perceivable degradation of the image at that pixel, we need to link the edges and bars of the image with the displacement field at the corresponding

location. This is a consequence of the basic idea of our approach, as explained in Sec. III-A, and clear by looking at Fig. 5. Specifically, we are interested in the displacement field D_θ (the projection of D along θ), that is orthogonal to the bars and edges of the image, and we have to consider the gradient of D_θ with respect to the direction orthogonal to θ, since we assume that this gradient gives us an indication of the degradation of the structures in the image and thus of the perceptual degradation of the image. Following the above argument, the local quality score associated to each pixel can be defined by the following equation:

$$\mathrm{Obj}(x,y) = \sum_{\theta} (I_{f,\theta}(x,y))^{\alpha} \left(\frac{\partial D_\theta}{\partial d_\theta^{\perp}}(x,y) \right)^{\beta} \tag{6}$$

where $I_{f,\theta}$ is the filtered image described by equation (5) in the θ direction (providing the structural information of the image in the direction orthogonal to θ), α and β are two constants to be fixed, and the notation $\frac{\partial D_\theta}{\partial d_\theta^{\perp}}$ indicates the gradient of the displacement field in the θ direction with respect to the direction orthogonal to θ (applying the gradient of the displacement field we can associate a score equal to zero to translations that do not affect image quality and we can evaluate the smoothness of the displacement field, since very smooth displacements return small values of the local gradients). The exponents α and β have been inserted to allow a different weighting of the two components of the metric, namely the displacement gradient and the contrast of image structures.

For example, for $\theta = 0$, $I_{f,\theta}$ contains vertical edges and bars, like in Fig. 6(b); D_θ is the displacement field in the horizontal direction, $d_\theta^{\perp}$ is in the vertical direction and $\partial D_\theta / \partial d_\theta^{\perp}$ is the gradient of the horizontal displacement D_θ in the vertical direction.

The summation over θ in equation (6) is needed to extract the salient features along different orientations. Typical applications of Gabor filters require the exploitation of at least 4 different orientations: 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$.

Be aware that the amplitude of the response of the Gabor filter depends on the contrast, as we can see by looking at the images in Fig. 6, thus, using this approach, the perceived distortion will depend on the contrast too. This is an important point since it is well known that human vision is more sensitive to high contrast areas of the image. However, there is no evidence that the relationship between the contrast and the perceived distortion is linear, rather, with higher probability, it is a sigmoid function to take into account the saturation effects typical of human senses. In our metric, such a saturation effect is taken into account by means of a fitting function whose goal is to fit the human perception of geometric artifacts with the values provided by the objective metric.

D. Error pooling

In order to pass from the local scores to an overall metric, that is an overall score that quantifies the perceived distortion globally, we included and Error Pooling step. Pooling refers to the task of arriving at a single measurement of quality by starting from local artifacts. The main problem of pooling is that it is not quite understood how the HVS performs pooling, though is quite obvious that pooling involves cognition, where a perceptible distortion may be more annoying in some areas of the scene (such as human faces) than in others. Most quality assessment metrics use Minkowski relation to pool the error signals from the different frequency and orientation selective streams, as well as across spatial coordinates, to arrive at a single fidelity measurement.

In our approach, the global quality score is computed by using the Minkowsky relation, as follows:

$$\text{Score} = \left(\sum_{x,y} |\text{Obj}\,(x,y)|^p \right)^{\frac{1}{p}} \tag{7}$$

where p is typically a constant between 1 and 4 and whose value was set experimentally as shown in the next section.

All the steps described so far are exemplified in Fig. 7 with reference to the synthetic image given in Fig. 4. The figure has been obtained by using two orientations ($\theta = 0$ and $\theta = 90$) and the following parameters: $\gamma = 0.5$, $\lambda = 10$, $\alpha = 1$, $\beta = 1$. Fig. 7 shows an example of distortion, the filtered images obtained using equation (5) and the pixels scores evaluated using equation (6) for two separate values of θ ($\theta = 0$ and $\theta = 90$). Fig. 7.(f) shows the total score associated to each pixel (found by using equation (6)) for both the orientations $\theta = 0$ and $\theta = 90$. By looking at Fig. 7.(f) and 7.(a) we can see that the local scores are higher at locations with a higher perceived degradation.

E. Fitting curve

To produce an objective quality score with the same range as the subjective scores provided by the users, and to account for the saturation effect typical of the HVS, a fitting curve is applied to the global scores obtained trough the error pooling stage. The purpose of the fitting function is to associate the values given by the objective metric to the subjective scores provided by the subjects and this step is always necessary to take into account the saturation effect typical of human senses. Moreover we need to take into account, as explained in the previous section, the non linear relationship between the contrast and the perceived distortion. Through the fitting procedure, a match between the human perception of geometric artifacts and the values provided by the objective metric defined by equation (7) has to be established.

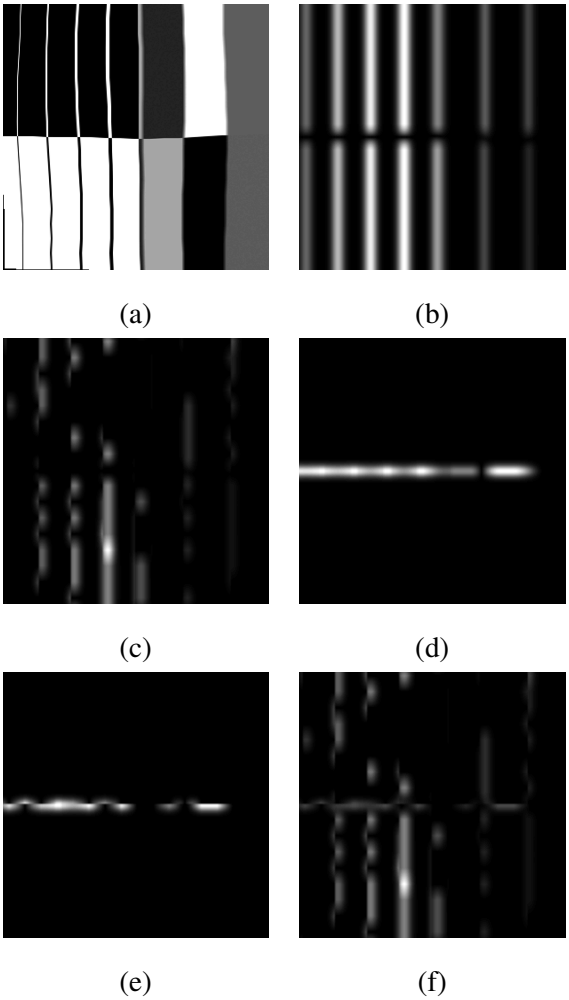


Fig. 7. Example of the proposed approach: (a) distorted image; (b) filtered image $If(x, y)$ with $\theta = 0$; (c) pixel scores $\mathrm{Obj}(x, y)$ along the direction $\theta = 0$; (d) filtered image $If(x, y)$ with $\theta = 90$; (e) pixel scores $\mathrm{Obj}(x, y)$ along the direction $\theta = 90$; (f) total pixel scores.

Many authors transform the value V produced by their criterion into an objective quality score Obj by using the following non-linear function whose parameters are optimized on the tested database:

$$Obj = \frac{a}{1 + \exp\left(b * (V - c)\right)}.$$

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

$$Obj = -4\left(1 - e^{-\left(\frac{V}{c}\right)^k}\right) + 5 \tag{8}$$

where c and k are the parameters to be estimated by fitting the objective metric values to the subjective

data, and V is the global quality score described by equation (7). Thanks to the fitting procedure it is possible to obtain a perceptual quality score, going from 1 to 5, quantifying the dissatisfaction of the viewer observing the distorted image (with 1 corresponding to a bad image quality and 5 to an excellent image quality). We opted for this function because it provides the best fit for our data among the commonly used curves, i.e., Gaussian, logistic and Weibull curves. To estimate the b and k parameters, we used a nonlinear least squares data fitting by the Gauss-Newton method on the tested database, as explained in the next section.

IV. METRIC TUNING

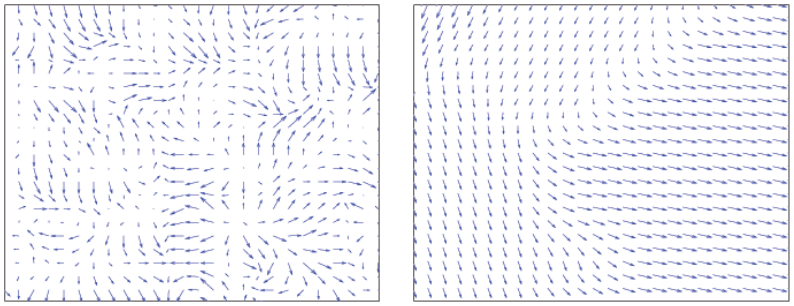
In the previous section we described the overall architecture of our metric. We now must tune all the parameters defining the metric to subjective data and to validate the final metric obtained in this way. To this aim, two sets of subjective experiments were carried out with different purposes. The first set of experiments, that is presented in this section, was performed to set the parameters of the model $\theta, \lambda, \alpha, \beta, p$ (the other parameters are the default values of the Gabor filters: $\gamma = 0.5$ and $b = 1$). The second set of experiments was conducted, as we will see in the next section, to validate the proposed metric.

A. The image database

The image database used for the first test included fifteen gray scale high quality images, 512×512 pixel in size, and was derived from a set of source images that reflects adequate diversity in image contents. The images of the database include pictures of faces, houses and landscape scenes. Some images have high activity, while some do not have much structures and are mostly smooth.

To automatically generate the local geometric distortions to be applied to the images and to have a broad range of image impairments we used the Constrained LPCD model [18] and the Markov Random Field (MRF) model [25] with different parameters in order to obtain different kinds of distortions going from invisible distortions to very annoying distortions, for a total of nine different distortions for each image and a total of 135 different images to be evaluated. We decided to use these models because since they allow to generate a wide range of distortions, by changing the parameters of the models, and because they represent two recently introduced and remarkable models in the digital watermarking research, a field that is one of the main motivation behind our work.

Two examples of distortions that is possible to generate by using these models are given in Fig. 8. With the MRF model it is possible to obtain larger displacement vectors than with the LPCD model,



(a) (b)

Fig. 8. Examples of displacement fields: (a) C-LPCD, (b) MRF.

while keeping the distortion less annoying thanks to its ability to generate a very smooth field (readers may refer to [25] for details).

B. Test methodology

We need to measure the perceived quality of geometrically distorted images trough subjective scaling methods. The subjective scaling method we used is the Absolute Category Rating (ACR) method [1], a category judgement where the test images are presented one at a time and are rated independently on a category scale. Even if a strictly distinction does not exist, a Single Stimulus Method, such as the ACR test, is well-suited for qualification tests, when quantifying the magnitude of the degradation, since the presentation of the stimuli is similar to that of the common use of the system. Double Stimulus Methods like the Degradation Category Rating (DCR), instead, are preferred when testing the fidelity of transmission with respect to the source signal. This is frequently an important factor in the evaluation of high quality systems, when it is important to check the fidelity with respect to the source signal and to find the perceptibility threshold.

Procedures for ACR experiment have been designed by following the ITU-T Recommendation P.910,13 which suggests standard viewing conditions, criteria for the selection of observers and test material, assessment procedures, and data analysis methods.

The ACR method specifies that after each presentation the subjects are asked to evaluate the quality of the stimulus shown using the following five-level scale: 1) Bad; 2) Poor; 3) Fair; 4) Good; 5) Excellent.

The time pattern for the stimulus presentation is set in the following way: the voting time equals to 5 seconds and the images presentation time to 8 seconds.

The experiments were conducted in a dark room by using the VP800 video card of the Cambridge

Research Systems together with a high resolution 21-inch digital monitor Mitsubishi DiamondPro 2070 with the external adaptor ViSaGe 71.02.00D2.14[4]. To have a correct color representation a luminosity calibration was previously carried out through a videocamera ColorCAL.14. Subjects watched the images at a distance of 4 times the height of the image displayed on the monitor. All the parameters regarding the test environment have been set consistently with the international standard in terms of illumination, background, display terminal, etc.

The tests involved a panel of sixteen subjects with a good vision, all naives with respect to image quality assessment methods and image impairments. Subjects were shown images in a random order, the randomization was different for each subject.

The experiment followed a five-stage procedure. The stages were: (1) oral instructions, (2) training, (3) practice trials, (4) experimental trials, (5) interview. In the first stage, the subjects were verbally given instructions and made familiar with the task and the graphic interface. The scenario of the intended application of the system under test was explained to the subjects. In addition, a description of the type of assessment, the opinion scale and the presentation of the stimuli was given. In the training the original models and the distorted models were shown to establish the range for the impairment scale. The practice trials stage was used to familiarize subjects with the experimentation. The range and type of impairments were presented in preliminary trials, which contained images and distortions other than those used in the real tests. In the experimental stage, the subject had to give a score to indicate how much the distortions were evident. Finally, in the interview stage, we asked the test subjects for a qualitative description of the perceived distortions.

C. Processing of data

The subjective scores must be analyzed with statistical techniques [1] to yield results which summarize the performance of the metric. The averaged score values (MOS), that is the arithmetic mean of all the individual scores, are considered as the amount of distortions that anyone can perceive on a particular image. However, impairment is measured according to a certain scale, as explained before, and such a scale may vary from person to person. For this reason, we used standard methods based on Kurtosis coefficient to screen the judgments provided by the subjects, that is to eliminate viewers with extreme scores.

[4]http://www.crsltd.com/catalog/visage/index.html

As explained before, the goal of the first experiment was to match human perception of geometric artifacts with the values provided by the objective metric and to set the parameters of the metric so to maximize this matching.

We tested the objective metric with the following values of the parameters: $\alpha \in \{\frac{1}{2}, 1, 2, 3\}$, $\beta \in \{\frac{1}{2}, 1, 2, 3\}$ and $p \in \{1, 2, 3, 4\}$. Regarding the number of orientations to be used, we tested the proposed approach with different number of orientations going from 2 ($\vartheta \in \{0, \frac{\pi}{2}\}$) to 8 ($\vartheta \in \left\{0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}\right\}$).

In order to find the optimum value of the parameters maximizing the match between the quality score and the human responses, we followed the metric performance evaluation procedures employed in the Video Quality Experts Group (VQEG) Phase I FR-TV test [26]. Performance of the objective metric is usually evaluated with respect to three aspects of its ability to estimate subjective assessment of image quality. **Prediction accuracy** is the ability to predict the subjective quality ratings with low error; **prediction monotonicity** is the degree to which the models prediction agrees with the relative magnitudes of subjective quality ratings, the objective image quality measures scores should be monotonic in their relationship to the performance scores; **prediction consistency** is the degree to which the model maintains prediction accuracy over the range of test sequences, i.e., the objective quality measures capability to provide consistently accurate predictions for all types of images and not to fail badly for a subset of images.

To remove any nonlinearities due to the subjective rating process and to facilitate comparison of the models in a common analysis space, the relationship between objective data and the subjective ratings was estimated by using a nonlinear regression, the Weibull functions described by the following equation:

$$y = 4\left(1 - e^{-\left(\frac{x-1}{c1}\right)^{k1}}\right) + 1 \tag{9}$$

Once the nonlinear transformation was applied, the metric attributes listed above were evaluated by using different performance metrics, applied on the fitted values, as specified in the report of the VQEG. The first metrics are 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. The second metric is the Spearman rank-order correlation coefficient between the objective/subjective scores that is considered a good evaluation of prediction monotonicity. Finally, the third metric is 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.

By using the standard performance evaluation method described above we found that the values that maximize the correlation between the objective metric and the subjective scores are the following: $\alpha = 1$, $\beta = 3$ and $p = 1$. The value of p shows that in the case of geometric distortions the Minkowsky equation is not useful to understand the visual perception mechanism of the human brain, thus the overall score associated to each distortion, will be simply given by the sum of all pixels scores.

Regarding the number of orientations to be used in equation (6), we observed that adding more than 2 orientations does not increase the correlation between the objective metric and the MOS data, for this reason we decided to use only two orientations: 0, $\frac{\pi}{2}$. This result can be explained by considering that in real word-images, including both natural landscapes and man-made environments, vertical and horizontal orientations are more frequent than obliques ones [27]. Furthermore it is well known in neurophysiological studies of visual pathways, that the performance for a large variety of perceptual tasks is superior for stimuli aligned in horizontal or vertical orientations, as compared to stimuli in oblique orientations. This phenomena is called the *Oblique Effect* [28].

In table I the performance of the proposed quality metric, for different number of orientations θ and different values of p in the Minkowski error pooling, is shown. These are four of all the possible configurations we tested to find the optimum values of the parameters used in the metric. As already said, by looking at the table we can observe that increasing the number of orientations does not improve the performance of the metric while changing the value of p in the Minkowski error pooling results in lower correlation coefficients.

	Pearson correlation	RMSE MOS - MOSp	Spearman correlation	Outlier ratio
$\alpha = 1$,$\beta = 3$,$\vartheta = 2$,$p = 1$	0.8419	0.5928	0.8182	0.1333
$\alpha = 1$,$\beta = 3$,$\vartheta = 2$,$p = 4$	0.7181	0.7945	0.6706	0.1852
$\alpha = 1$,$\beta = 3$,$\vartheta = 8$,$p = 1$	0.8435	0.6108	0.8100	0.1300
$\alpha = 1$,$\beta = 3$,$\vartheta = 8$,$p = 4$	0.6636	0.8547	0.6372	0.2519

TABLE I

PERFORMANCE OF THE PROPOSED PERCEPTUAL METRIC AS A FUNCTION OF θ AND p

One more parameter need to be evaluated in this experiment, that is the wavelenght λ.

Because real-world images contain distinct features at various resolutions, efficient features extraction may require the filtering process across several scales. In our approach this consideration corresponds to find the correct value of σ (the standard deviation of the Gaussian factor of the Gabor filter) for each

image. Using an adaptive value of σ, linking the structure of the image to the scene scale, seems to be a cumbersome task, thus we decided to find the correct scale to be used for each class of images assuming that images belonging to the same class can be described at the same level of resolution. We followed two different approaches to evaluate the performances of the proposed metric either considering all the images together (fixing an unique value of the image scale) or applying the metric to different classes of image. For this purpose we tested the objective metric with different values of λ (as we said σ and λ are connected as follows: $\sigma = 0.56\lambda$) and we found, for each category of images, the value of the wavelenght of the cosine factor of the Gabor filter that minimizes the error in the fitting procedure ($\lambda = 9$ for all images together, $\lambda = 10$ for the class of house images, $\lambda = 6$ for landscape images, $\lambda = 8$ for face images). This approach can be justified by considering that natural images can be classified in basic-level categories [29] and images belonging to the same categories share some common features and statistics [27] including image scale. In fact, statistics of natural images have been found to follow particular regularities, specifically different categories exhibit different orientations and spatial frequency distributions, captured in the averaged power spectra. These results are usually applied to the problem of scenes categorization in the area of computer vision.

The results of the subjective test, using the values found for $\theta, \lambda, \alpha, \beta, p$, are described in Fig. 9 which shows the scatter plot of the Mean Opinion Score versus the objective metric evaluated by using equation (7). Specifically, Fig. 9.(a) shows the scatter plot for all the 135 images while the following three graphs present the results obtained for each class of images.

The equation of the overall score associated to each distortion, using the values for $\theta, \lambda, \alpha, \beta, p$ found with the subjective test, is given by the following formula:

$$\mathrm{Score} = \sum_{x,y} \sum_{\theta \in \mathrm{S}} \mathrm{I_{f,\theta}(x,y)} \left(\frac{\partial \mathrm{D}_\theta}{\partial \mathrm{d}_\theta^\perp}(\mathrm{x},\mathrm{y}) \right)^3 \tag{10}$$

where $S = \{0, \frac{\pi}{2}\}$ and the final perceptual metric is described by the following equation:

$$\mathrm{GaborMetric} = -4 \left(1 - e^{-\left(\frac{\mathrm{Score}}{c}\right)^k} \right) + 5 \tag{11}$$

where c and k are parameters whose value is reported in table II. The Weibull function in equation (11) describes our metric: for each distortion it returns a numerical score, going from 1 to 5, quantifying the dissatisfaction of the viewer observing the distorted image (with 1 corresponding to a bad image quality and 5 to an excellent image quality).

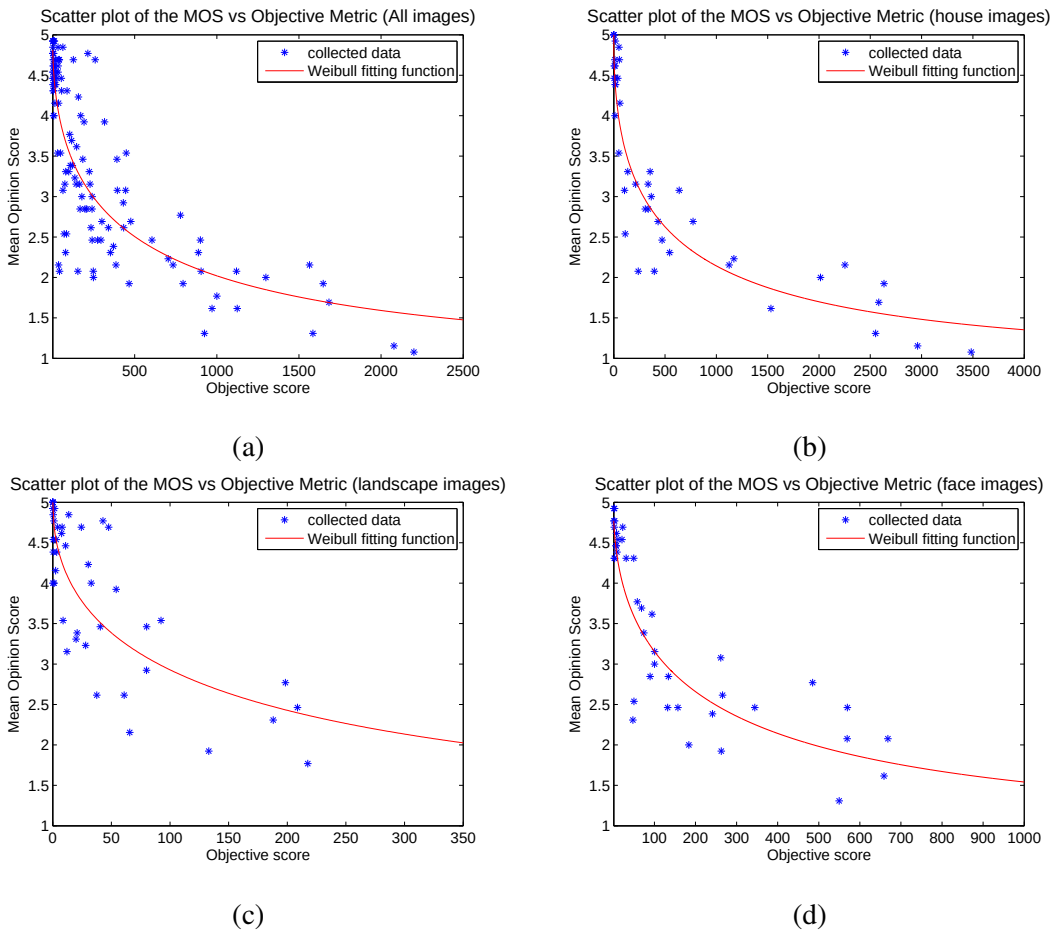


Fig. 9. Scatter plot of the Mean Opinion Score versus Objective Metric: (a) all images; (b) houses images; (c) landscape images; (d) faces images.

	All images	House images	Landscape images	Face images
c	524.58	623.5	188.4	257
k	0.4838	0.4771	0.4979	0.5101

TABLE II

VALUE OF THE PARAMETERS USED FOR THE WEIBULL FITTING FUNCTION

V. METRIC VALIDATION

After the definition of the metric, we need to cross-validate it. Cross-validation is an important step towards successful development of practical image quality measurement systems and the most standard form of validation is to compare objective quality measures with ratings by human subjects on an extensive

database of images.

A. Subjective test

To cross-validate the proposed metric a new subjective test was designed and performed. Once again we used the ACR test following the procedures explained in the previous section. A new dataset of fifteen images was built according to the class of images explained previously and new nine distortions for each image were generated by using the same models C-LPCD and MRF. The tests involved a panel of other sixteen subjects.

The results of the test are shown in Fig. 10 that describes the scatter plot of the MOS versus the perceptual metric described by equation (11) for all the images and for classes of images. The interval of confidence (IC) at 95% has been added to the plots to have an idea of the accuracy of the MOS value (for the purpose of visibility the IC has not been added to the plot of all images).

The performances of the proposed metric, through the standard performance evaluation methods described in the previous section, are shown in table III.

	All images	House images	Landscape images	Face images
Pearson correlation coefficient	0.8322	0.8808	0.7465	0.9054
RMSE MOS - MOSp	0.5858	0.5163	0.5370	0.4421
Spearman correlation coefficient	0.8482	0.8969	0.7778	0.8772
Outlier ratio	0.1407	0.0667	0.1111	0.0222

TABLE III

PERFORMANCE OF THE PROPOSED PERCEPTUAL METRIC

By referring to this table and by looking at the scatter plot in Fig. 10, the following considerations are in order:

- Both the Pearson and the Spearman coefficient are quite high for all the classes of images revealing a good prediction accuracy and monotonicity of the model. The outlier ratio is quite low revealing that the metric maintains prediction accuracy over the range of image sequences.
- Applying the model per class of images instead than to all the images together brings a little improvement of the objective metric for the class of house and face images but it is not so relevant considering the disadvantage of having different objective metrics for different classes of images.

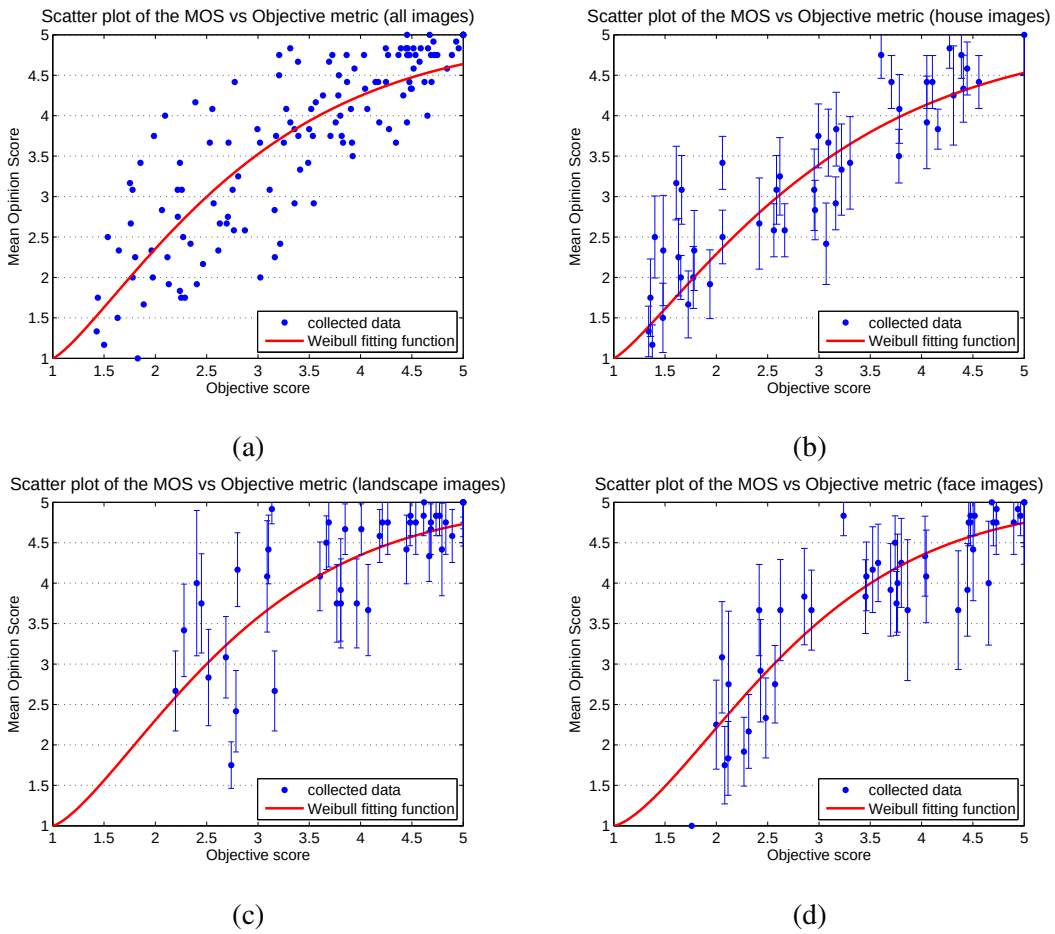


Fig. 10. Scatter plot of the Mean Opinion Score versus Gabor metric: (a) all images; (b) house images; (c) landscape images; (d) face images.

- The performance of the proposed metric slightly decreases for the class of landscape images. This result can be explained by observing that it is more difficult to perceive the distortions in this kind of images due to the absence of edges and bars.
- After the nonlinear regression the relationship between the objective and subjective data is almost linear thanks to the fitting procedure that we used for the metric design and described in the previous section.

B. Comparison with other metrics

We now provide a comparison of the proposed technique with other metrics in the literature. Specifically we considered the PSNR measurement and the SSIM-index [8], that are two widely used full-reference

quality metric thanks to their simple formulation and computation, the C4 metric [10] that is a reduced-reference metric, the Gibbs metric [20] and the metric based on the variance of the jitter noise [15], that are specifically designed to deal with geometric distortions.

1) PSNR and SSIM index: We expect that statistical measures based on pixel-wise comparisons between the original and the distorted image, like PSNR and SSIM index, cannot work in presence of geometric distortions. However, just for completeness, Fig. 11(a) and Fig. 11(b) show the scatter plot of the MOS versus PSNR and SSIM index, respectively, for all the images (similar results are for classes of images). As we expected there is no evident correlation between the two objective metrics and the users response.

Depending on the application, it is also possible to compute a weighted average of the different samples in the SSIM index map.

A weighted version of the SSIM index, as described in [30], was also tested and the scatter plot of the Mean Opinion Score versus WSSIM (weighted SSIM) is shown in Fig. 11(c). As we expected there is no evident correlation between the improved version of the SSIM index and the users response.

2) C4 metric: C4 [10] is a metric based on the comparison between the structural information extracted from the distorted and the original images. Perceptual features are extracted from some selected characteristic points and their neighborhoods. The extracted features characterize the neighborhood of characteristic points in terms of oriented segments with contrast computed at different resolutions. Such features correspond to the type of information which is extracted in area V1 of the HVS since they indicate the orientation, length, width and magnitude of the contrast at the characteristic point.

Fig. 11(d) shows the scatter plot of the Mean Opinion Score versus C4 metric for all the images. It is possible to observe that there is no evident correlation between the objective metric and the users response.

3) Jitter noise variance: The measure proposed by Licks et al. [15], focuses the attention on the case where the jitter function $g(t)$ describing the geometric distortion is such that its sample values are equal to $g(n2T) = J = (j1, j2)$ where $j1$, and $j2$, are i.i.d. samples drawn from the marginal distribution $f_J(j) \sim N(0, \sigma_J^2)$. The metric is given by the variance of the jitter noise $\sigma_{n_J}^2$ as follows:

$$E\left[n_J^2\right] \cong E\left[\left(\nabla_Z^T J\right)^2\right] = \sigma_J^2 \sigma_{\nabla_z}^2$$

where ∇_Z is the gradient of the distorted image (readers may refer to [15] for details).

Fig. 12 shows the scatter plot of the Mean Opinion Score versus the jiiter noise variance for all the images. Although this measure is specifically designed to deal with geometric distortions we can observe

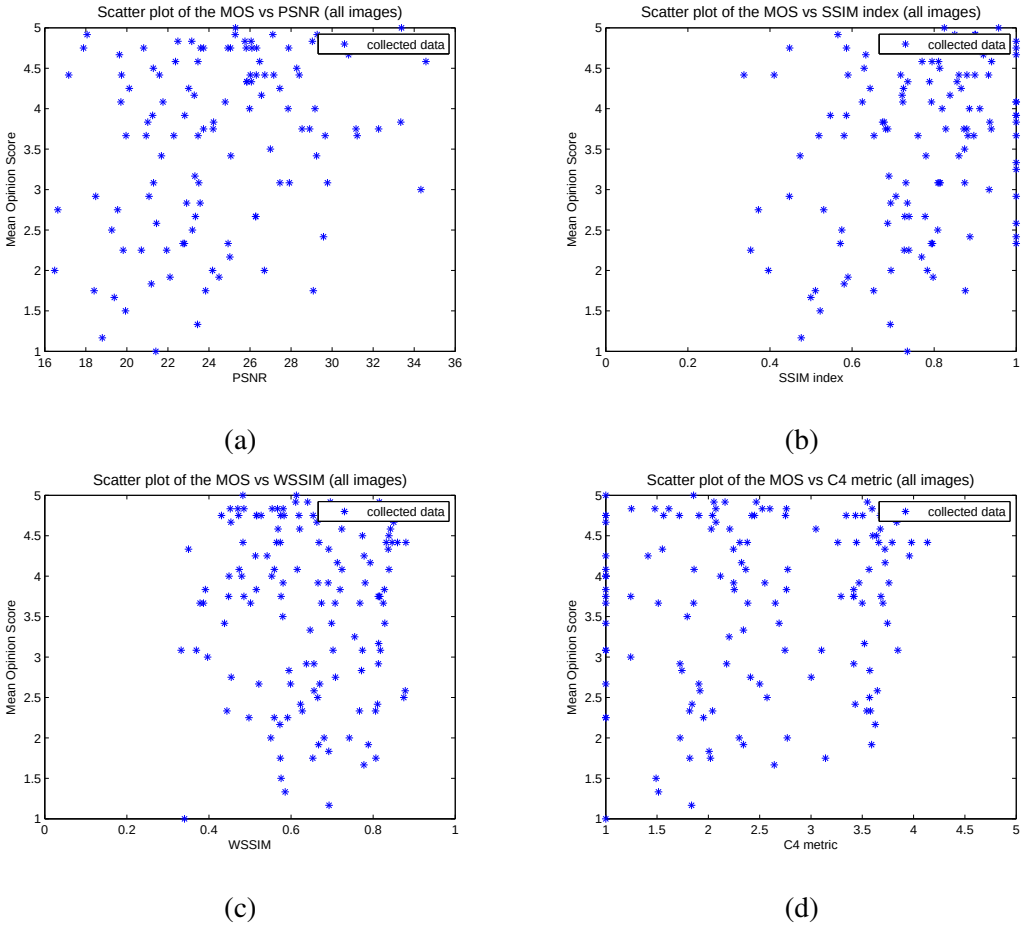


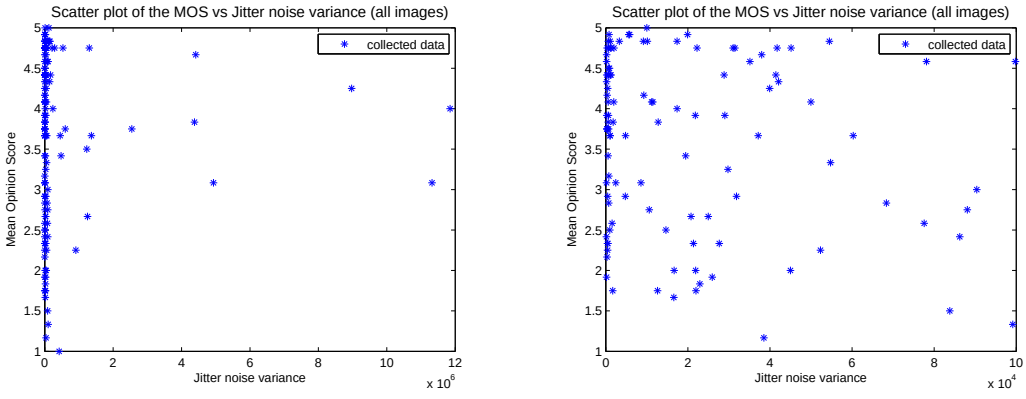
Fig. 11. Scatter plot of the Mean Opinion Score versus PSNR (a), SSIM index (b), Weighted SSIM (c) and C4 metric (d) for all images

that there is no correlation between the objective metric and the MOS.

4) Gibbs potential: We tested the Gibbs metric [20] based on the theory of the Markov Random Fields and relying on the assumption that the potential function of the configuration defining the geometric distortion gives an indication of the degradation of the distorted image. The results of the performance evaluation of the proposed algorithm through the three metrics explained above are shown in table IV.

By looking at the table we can see that Gibbs metric maintains prediction accuracy over the range of image sequences, and both the Pearson and the Spearman coefficient are quite high revealing a quite good prediction accuracy and monotonicity of the model, except for the class of landscape images, for which the objective metric is not able to predict the human responses.

By comparing the new, Gabor-based, approach with the Gibbs metric we can observe the significant



(a) (b)

Fig. 12. Scatter plot of the Mean Opinion Score versus Jitter noise variance (on the left there is the total plot, on the right a zoomed version of it).

	All images	House images	Landscape images	Face images
Pearson correlation coefficient	0.7910	0.8766	0.5018	0.8547
RMSE MOS - MOSp	0.6455	0.4792	0.5412	0.5098
Spearman correlation coefficient	0.8054	0.8993	0.4842	0.8447
Outlier ratio	0.2148	0.1433	0.1111	0.0667

TABLE IV

PERFORMANCE OF THE GIBBS METRIC

improvement brought by the Gabor-based metric. The value of both the correlation coefficients improves approximately from 0.80 to 0.84 for all the images together. There is a significantly improvement for the class of landscape and face images, while there is no improvement (the correlation coefficient is almost the same) for the class of house images.

C. Dependency of the method on the displacement field evaluation

So far we have assumed that the displacement field defining the geometric distortion is known. The reason for such an assumption was that we wanted to test the validity of the main ideas behind our metric, regardless of the accuracy of the displacement field estimation. We now evaluate the performance of the proposed metric when a particular displacement estimation algorithm is adopted. The algorithm we used for the displacement field evaluation, given the original reference images and the distorted ones,

is the well known image registration technique described in [31]. This is a general purpose registration algorithm designed for medical images/volumes. Given a source image (the distorted images) and a target image (the original images), it automatically estimates a smooth warp field that brings the source image into register with the target image. This approach is built upon a differential multi-scale framework, allowing to capture both large and small scale transformations.

The warp fields found by applying this registration technique to the image dataset were then used to find the corresponding quality scores with the proposed approach, as described in the previous sections. The efficiency of the quality metric in the presence of the particular displacement field estimation technique we used is evaluated in table V. As we can see, the performance of the metrics decrease only slightly with respect to the results shown in table III.

	All images	House images	Landscape images	Face images
Pearson correlation coefficient	0.7814	0.8815	0.7865	0.8358
RMSE MOS - MOSp	0.6888	0.5158	0.5053	0.6545
Spearman correlation coefficient	0.7937	0.8875	0.7594	0.8260
Outlier ratio	0.1926	0.1111	0.1111	0.1778

TABLE V

PERFORMANCE OF THE PROPOSED METRIC IN THE PRESENCE DISPLACEMENT FIELD ESTIMATION

VI. CONCLUSIONS AND FUTURE WORKS

In this paper we proposed a full-reference method of objectively assessing the perceptual quality of images distorted with local geometric distortions. To best of our knowledge only few works can be found in literature addressing this problem and the novelty of the proposed technique is that it considers both the displacement field describing the distortion and the structure of the image. The proposed metric is based on the use of Gabor filters to extract the structures of the image and on the evaluation of how such structures are distorted by the displacement field describing the geometric transformation between the original and the distorted image.

The experimental results show good performances of the metric, either when the scheme is applied to a generic image or to a specific class of images. Though the correlation with human observation is not perfect, as shown by the plots in the previous section, the performance of the mew metric greatly

outperforms all the metric proposed so far, be them general purpose metrics or metrics explicitly designed to work with geometric distortions.

Some ideas for future works could be the use of a multi-scale approach to incorporate image details at different resolutions and to apply high level perceptual factors that are not captured by the low level processing of edges and bars.

An interesting feature of the propose technique is that the value of σ could be used to take the viewing distance into account, a possibility that is not allowed by most available metrics (e.g. the SSIM metric). We plan to investigate this issue in a future research.

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] *Subjective Video Quality Assesment Methods for Multimedia Applications Recommendation P.910*, International Telecommunication Union, Geneva, Switzerland, 1996.

[2] M. Barni and F. Bartolini, *Watermarking systems engineering: enabling digital assets security and other applications*. CRC Press, 2004.

[3] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992.

[4] M. Celik, G. Sharma, and A. Tekalp, "Collusion-resilient fingerprinting by random pre-warping," *Signal Processing Letters, IEEE*, vol. 11, no. 10, pp. 831–835, 2004.

[5] R. Cipolla, T. Drummond, and D. Robertson, "Camera calibration from vanishing points in images of architectural scenes," *Proceedings of British Machine Vision Conference*, vol. 2, no. 382-391, p. 2, 1999.

[6] V. Licks and R. Jordan, "Geometric Attacks on Image Watermarking Systems," *IEEE MULTIMEDIA*, pp. 68–78, 2005.

[7] J. Gallant, J. Braun, and D. Van Essen, "Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex," *Science*, vol. 259, no. 5091, pp. 100–103, 1993.

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

[9] Z. Wang, A. Bovik, and L. Lu, "Why is image quality assessment so difficult?" *Acoustics, Speech, and Signal Processing, 2002. Proceedings.(ICASSP'02). IEEE International Conference on*, vol. 4, 2002.

[10] M. Carnec, P. Le Callet, and D. Barba, "Objective quality assessment of color images based on a generic perceptual reduced reference," *Signal Processing: Image Communication*, 2008.

[11] Z. Wang and A. Bovik, "A universal image quality index," *Signal Processing Letters, IEEE*, vol. 9, no. 3, pp. 81–84, 2002.

[12] B. Petersch, J. Bogner, A. Fransson, T. Lorang, and R. Pötter, "Effects of geometric distortion in 0.2 T MRI on radiotherapy treatment planning of prostate cancer," *Radiotherapy and Oncology*, vol. 71, no. 1, pp. 55–64, 2004.

[13] S. Shlien, "Geometric correction, registration, and resampling of Landsat imagery," *Canadian Journal of Remote Sensing*, vol. 5, pp. 74–89, 1979.

[14] M. Barni, "Effectiveness of exhaustive search and template matching against watermark desynchronization," *IEEE Signal Processing Letters*, vol. 12, no. 2, pp. 158–161, 2005.

[15] V. Licks, F. Ourique, R. Jordan, and F. Perez-Gonzalez, "The effect of the random jitter attack on the bit error rate performance of spatial domain image watermarking," *Image Processing, 2003. Proceedings. 2003 International Conference on*, vol. 2, 2003.

[16] R. Bauml, J. Eggers, and J. Huber, "A Channel Model for Watermarks Subject to Desynchronization Attacks," *4th International ITG Conference on Source and Channel Coding, Berlin, Germany, January 28*, vol. 30, pp. 28–30, 2002.

[17] X. Desurmont, J. Delaigle, and B. Macq, "Characterization of geometric distortions attacks in robust watermarking," *Proceedings of SPIE*, vol. 5306, pp. 870–878, 2004.

[18] A. D'Angelo, M. Barni, and G. Menegaz, "Perceptual Quality Evaluation of Geometric Distortions in Images," in *Proc. of SPIE Human Vision and Electronic Imaging XII*, vol. 6492, 2007.

[19] I. Setyawan, D. Delannay, B. Macq, and R. Lagendijk, "Perceptual quality evaluation of geometrically distorted images using relevant geometric transformation modeling," *Proceedings of SPIE, Security and Watermarking of Multimedia Contents V*, vol. 5020, pp. 85–94, 2003.

[20] A. D'Angelo, M. Pacitto, and M. Barni, "A Psychovisual Experiment on the Use of Gibbs Potential for the Quality Assessment of Geometrically Distorted Image," in *Proc. of SPIE Human Vision and Electronic Imaging XIII*, vol. 6492, 2008.

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

[22] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Optical Society of America, Journal, A: Optics and Image Science*, vol. 2, pp. 1160–1169, 1985.

[23] C. Grigorescu, N. Petkov, and M. Westenberg, "Contour detection based on nonclassical receptive field inhibition," *Image Processing, IEEE Transactions on*, vol. 12, no. 7, pp. 729–739, 2003.

[24] S. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on Gabor filters," *Image Processing, IEEE Transactions on*, vol. 11, no. 10, pp. 1160–1167, 2002.

[25] A. D'Angelo, M. Barni, and N. Merhav, "Stochastic Image Warping for Improved Watermark Desynchronization," *EURASIP Journal on Information Security*, vol. 2008, 2008.

[26] P. Corriveau and A. Webster, "VQEG evaluation of objective methods of video quality assessment," *SMPTE journal(1976)*, vol. 108, no. 9, pp. 645–648, 1999.

[27] A. Torralba and A. Oliva, "Statistics of natural image categories," *Network: Computation in Neural Systems*, vol. 14, no. 3, pp. 391–412, 2003.

[28] S. Appelle, "Perception and discrimination as a function of stimulus orientation: the oblique effect in man and animals," *Psychology Bull*, vol. 78, no. 4, pp. 266–278, 1972.

[29] B. Tversky and K. Hemenway, "Categories of enviromental scenes," *Cognitive Psychology*, vol. 15, no. 1, pp. 121–149, 1983.

[30] Z. Wang, L. Lu, and A. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121–132, 2004.

[31] S. Periaswamy and H. Farid, "Medical image registration with partial data," *Medical Image Analysis*, vol. 10, no. 3, pp. 452–464, 2006.