A color image quality assessment using a reduced-reference image machine learning expert

Christophe Charrier, Gilles Lebrun, Olivier Lezoray

A color image quality assessment using a reduced-reference image machine learning expert.

Christophe Charrier Gilles Lebrun Olivier Lezoray
Université de Caen-Basse Normandie, GREYC UMR CNRS 6072, Equipe Image
120, route de l'exode, 50000 Saint-Lô, France

ABSTRACT

A quality metric based on a classification process is introduced. The main idea of the proposed method is to avoid the error pooling step of many factors (in frequential and spatial domain) commonly applied to obtain a final quality score. A classification process based on the Support Vector Machine method is designed to obtain the final quality class with respect to the standard quality scale provided by the UIT. Thus, for each degraded color image, a feature vector is computed including several Human Visual System characteristics, such as, contrast masking effect, color correlation, and so on. Selected features are of two kinds: 1) full-reference features and 2) no-reference characteristics. That way, a machine learning expert, providing a final class number is designed.

Keywords: Machine Learning, HVS, Classification, Image quality

1. INTRODUCTION

In language abuse, there often is an amalgam between the terms quality and fidelity. For the Artist, the concept of quality would be what the concept of fidelity would be to the forger. The Artist generally works starting from the concepts, impressions related to their social environment and/or professional and places himself in an existing artistic current(relation Master-student) or in the new current that he creates. The works carried out are thus regarded as original, and the experts speak about the quality of the works. Behind this approach, one realizes that to the word quality, the concept of originality is associated. Who has never found themselves on front of a work which left him perplex while his neighbor was filled with wonder? It is enough to saunter in the museums to see this phenomenon. Thus one qualifies the quality of a work according to his conscience and his personal sensitivity presets from his economic and social environment. Thus, many internal subconscious scales are implied when looking at an object for the first time, the human observer is able to say if what he sees is pleasant or not. Then he just makes one classification of the perception of this object according to the feeling he has in two categories: "I like" or "I don't like".

Such an ability to classify the visual feelings is indeniably to put in relation with the inherent conscience of each human being. The conscience is related to what Freud calls "the perception-conscience system". It concerns a peripheral function of the psychic apparatus which receives information from the external world and those coming from the memories and the internal feelings of pleasure or displeasure. The immediate character of this perceptive function involves an unability for the conscience of keeping a lasting trace of this information. It communicates them to the preconscience, place of a first setting in memory. The conscience perceives and transmits significant qualities. Freud employs formulae like "index of perception, of quality, of reality " to describe the content of the operations of the perception-conscience system.

Thus, perception is to be regarded as one of the internal scales of a process driving to an overall quality assessment of an object or an image.

Any designed quality metrics (including or not Human Visual System features) tends to make as well as any human observer, *i.e.*, one tries to clone the Human Visual System. Until now, to develop a quality metric, the usually applied scheme consists in performing 1) a color space transformation to obtain decorrelated color coordinates and 2) a decomposition of these new coordinates towards perceptual channels. An error is then estimated for each one

Further author information: (Send correspondence to C.C.)
C.C.: E-mail: christophe.charrier@unicaen.fr, Telephone: 33 233 77 11 66
G.L.: E-mail: gilles.lebrun@unicaen.fr, Telephone: 33 233 77 11 66
O.L.: E-mail: olivier.lezoray@unicaen.fr, Telephone: 33 233 77 11 66

of these channels. The final quality score is obtained by pooling these errors in both spatial and frequential domain. Nevertheless, the pooling stage is based on the use of the Minkowski error metric. Recent studies[1] have shown that this summation does not perform well even if it is the most widely used. One can obtain the same Minkowski value for two different distorted images while the visual quality drastically differs from one distorted image to the other. This can be explained by the fact that the implicit assumption of this metric is that all signal samples are independant. Yet, this is not the case when one uses perceptual channels. When applying an "error pooling" of the obtained estimated errors within each perceptual channels, the Minkowski metric fails to generate a good final score.

Furthermore, it can be very hard to interpret the difference of quality for two images whose scores are very close, since they are obtained from measures that usually do not integrate all the degradation measures and that those degradation are not necessary located in the same area of the image.

One main observation can be formulated: the need to obtain a final quality score is not the best way to quantify the quality. Actually, in the recommendations given by the UIT,[2] the human observers have, for example, to choose a quality class from a scale containing five notes from Very bad (1) to Excellent (5). Those notes characterize the quality of the reconstructed images. That way, the human observers only make one classification or categorization. This is the way investigated in this paper. No final score is computed, only the final quality class is given with an associated a posteriori probability. The latter yields us to make final decision concerning degraded color images for with the quality class is close to the frontier between two contiguous classes.

In this paper, the quality measure is based on a learned classification process in order to respect the one of human observers.[3] Instead of computing a final note, our method classifies the quality using the quality scale recommended by the UIT. This quality scale contains 5 ranks ranginng from 1 (the worst quality) to 5 (the best quality). The selected class of the proposed method represents the opinion score OS. That way, a machine learning expert, providing a final class number is designed.

2. THE PROPOSED APPROACH

2.1. Selection of the classifier

From all existing classification schemes, a Support Vector Machine (SVM)-based technique has been selected due to high classification rates obtained in previous works,[4] and to their high generalization abilities.

The SVMs were developed by Vapnik et al.[5] and are based on the structural risk minimization principle from statistical learning theory. SVMs express predictions in terms of a linear combination of kernel functions centered on a subset of the training data, known as support vectors (SV).

Given the training data $\mathcal{S} = \{(x_i, y_i)\}_{i=\{1,\ldots,m\}}, x_i \in \mathcal{R}^n$, $y_i \in \{-1, +1\}$, SVM maps the input vector x into a high-dimensional feature space $\mathbf{H}$ through some non linear mapping functions $\phi : \mathcal{R}^n \rightarrow \mathbf{H}$, and builds an optimal separating hyperplane in that space. The mapping operation $\phi(\cdot)$ is performed by a kernel function $K(\cdot, \cdot)$ which defines an inner product in $\mathbf{H}$. The separating hyperplane given by a SVM is: $w \cdot \phi(x) + b = 0$. The optimal hyperplane is characterized by the maximal distance to the closest training data. The margin is inversely proportional to the norm of w. Thus computing this hyperplane is equivalent to minimize the following optimization problem:

$$\mathcal{V}(w, b, \xi) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{m} \xi_i\right) \tag{1}$$

where the constraint $\forall_{i=1}^{m} : y_i \left[w \cdot \phi(x_i) + b\right] \geq 1 - \xi_i$, $\xi_i \geq 0$ requires that all training examples are correctly classified up to some slack ξ and C is a parameter allowing trading-off between training errors and model complexity.

This optimization is a convex quadratic programming problem. Its whole dual[5] is to maximize the following optimization problem:

$$\mathcal{W}(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{2}$$

subject to $\forall_{i=1}^{m} : 0 \leq \alpha_i \leq C$, $\sum_{i=1}^{m} y_i \alpha_i = 0$.

The optimal solution α^* specifies the coefficients for the optimal hyperplane $w^* = \sum_{i=1}^{m} \alpha_i^* y_i \phi(x_i)$ and defines the subset SV of all support vector (SV). An example x_i of the training set is a SV if $\alpha_i^* \geq 0$ in the optimal solution. The support vectors subset gives the binary decision function h:

$$h(x) = \text{sign}(f(x)) \text{ with } f(x) = \sum_{i \in SV} \alpha_i^* y_i K(x_i, x) + b^* \tag{3}$$

where the threshold b^* is computed via the unbounded support vectors[5] (*i.e.*, $0 < \alpha_i^* < C$). An efficient algorithm SMO (Sequential Minimal Optimization)[6] and many refinements[7,8] were proposed to solve dual problem. SVM being binary classifiers, several binary SVM classifiers are induced for a multi-class problem. A final decision is taken from the outputs of all binary SVM.[9]

2.2. SVM model selection

Kernel function choice is critical for the design of a machine learning expert. Radial Basic Function (RBF) kernel function is commonly used with SVM. The most important reason is that RBF functions work like a similarity measure between two examples. As no a priori knowledge exists on the relative importance of each feature s^k, the classical RBF function has been extended in order to reflect this fact and the kernel function has been defined as follow

$$K_\beta(s_i, s_j) = exp(-\sum_{k=1}^{n} \beta_k (s_i^k - s_j^k)^2 / r^2) \tag{4}$$

where s_i^k is the k^{th} feature of the i^{th} image. To have efficient SVM inducers, a parameter tuning process has to be realized. This procedure is the so-called model selection. The selection of the SVM hyper-parameter (C), the radius of RBF function (r) has been realized by using cross-validation. In this paper, β_k could only take binary value and modelize if the s^k feature is used or not. When β_k values are not fixed by human priors, they are determined by using a feature selection paradigm. The quality of a subset of features for the design of a binary SVM is measured by its recognition performance. This corresponds to a wrapper feature selection approach.[10] SVMs being binary classifiers, multi-class decision using SVMs are usually implemented by combining several two-classes SVM decision. Several combination schemes of binary classifiers exist.[9] Two schemes are used: 1) the common One-Versus-All (OVA) scheme and 2) a second one designed to take into account the existing Rank Ordering (RO) between the classes. Let $\mathbf{t}_i = \{t_{i,1}, \ldots, t_{i,n_c}\}$ be a class map vector to transform a n_c classes problem to a binary problem with $t_{i,j} \in \{+1, -1\}$. $t_{i,j}$ means that in the i^{th} binary problem, images initially located in class j now belong to the class $t_{i,j}$. Let $f_i(\cdot)$ and $h_i(\cdot)$ respectively be the SVM output and the SVM decision function obtained by training it on the i^{th} binary problem. Tables 1 and 2 respectively give binary problems transformation used in OVA and RO combination schemes. t_1 and t_5 transformation in OVA scheme are identical to t_1 and t_4 (class label switch is not significant) in RO scheme. The difference is concentrated to the others binary class maps. In the RO scheme, the information about the class label rank is preserved, but this is not true when using the OVA scheme (i.e. $\forall c_1, c_2 : t_{i,c_1} > t_{i,c_2} \rightarrow c_1 > c_2$). Moreover, discriminative function corresponding to t_2, t_3 or t_4 map in OVA is more difficult to realize when excellent and very bad images are merged in the same class and are used to identify *quite good* images.

Table 1. Binary problems transformation use in One-Versus-All combination scheme

class	t_1	t_2	t_3	t_4	t_5
5	+1	-1	-1	-1	-1
4	-1	+1	-1	-1	-1
3	-1	-1	+1	-1	-1
2	-1	-1	-1	+1	-1
1	-1	-1	-1	-1	+1

The binary problem transformation is the first part of a combination scheme. A final decision must be taken from all binary decision functions. Many combination strategies can be used to obtain the final decision.[9] The majority

Table 2. Binary problems transformation use the Rank Ordering combination scheme

class	t_1	t_2	t_3	t_4
5	+1	+1	+1	+1
4	-1	+1	+1	+1
3	-1	-1	+1	+1
2	-1	-1	-1	+1
1	-1	-1	-1	-1

vote criterium is the usual way to do this. Let $V_j(x) = \sum_{i=1}^{n_b} LO_1(h_i(x), t_{i,j})$ the number of votes for the class j (n_b is the number of binary decision function in a specific combination scheme, and LO_1 is defined in the next section). The multiclass decision function D using majority vote is: $D(x) = \arg\max_{1 \le j \le n_c} (V_j(x))$ (when conflicts exist, the SVM output is used to break it).

3. SELECTION OF THE FEATURES VECTOR

As a classification process is used, one has to design a set of features that will be used by the machine learning expert. That way, the class of the quality is obtained from a reduced reference image.

Yet, there are four types of attributes contained in that set: 1) full-reference image SVH-based features and 2) full-reference image features where a reference image is needed and 3) no-reference images SVH-based features and 4) no-reference images features for which no reference image is needed. Actually, when a human observer judges the quality of an image, it is not obvious that he takes its decision only making comparison between the reference image and the reconstruced one. Probably for some degradations, only the perception of those ones are implied. In addition, internal scales imply within the final decision is not entirely well-known.

3.1. full-reference image SVH-based features

When using full-reference image features, both original and degraded images are first subject to a transformation towards an antogonist luminance chrominance color space. The aim of such transformation is to take into account the spatial-frequency sensitivity of the SVH. The latter analyzes input signals by averaging a set of radial and orientation selective channels. Two questions may have to be asked to implement this sensitivity. The first one concerns the choice of the colorimetric decomposition. For the luminance component the most used decomposition are from Daly,[11] Watson[12] and the Lubin's one.[13] The two first are characterized by a diadic radial selectivity (five one octave bandwidth channels) and a constant angular selectivity. The main difference is situated around the value of the bandwith: 30 degrees for Daly's model and 45 degrees for the Watson's one. Lubin's decomposition needs seven radial bands and four orientations. For the chrominance component, the same decomposition is applied with fixed maximal frequency that is different for the C_1 axis from the C_2 axis.

The second question is linked to the choice of linear transformation to apply. Actually this one has to meet many properties such as the radial and angular selectivity, the linear phasis, a minimal recovery between adjacent channels, and so on. Only one transformation meets all the required properties: the Cortex Transform.[12]

Then, for each one of the resulting images (one luminance image and txo chrominance images) a local contrast is computed for each point of the three images. Then pooling those measures a global contrast score is generated.

3.1.1. Color space transformation

The perception of color differences in RGB is highly nonuniform. The study of perceptual uniformity concerns numerical differences that correspond to color differences at a perceptibility threshold (just noticeable differences, or JNDs).[14] In its second sense, color difference refers to color components where brightness has been removed. Actually, the Human Visual System has poor response to spatial detail in colored areas of the same luminance, compared to its response to luminance spatial detail.[15] The easiest way to remove brightness information to form

two color channels is to subtract it. The luma (luminance) component already contains a large fraction of the green information from the image, so it is standard to form the other two components by subtracting luma from nonlinear blue (to form B-Y) and by subtracting luma from nonlinear red (to form R-Y). These are called chroma. Various scale factors are applied to (B-Y) and (R-Y) for different applications.

From all existing opponent color spaces, the Krauskopf[16] one is selected. This coordinates system is computed from the LMS primaries that correspond to the HVS cone responses. The coordinates L , M and S represent the non-linear values due to the non-linear processing of the HVS. This transformation is obtained using 1) a logarithm function or 2) a power raise of 1/3. To compute those three non linear components, one need to apply a non linear transfer function (known to be the gamma function) to each of the component of the *RGB* color space. Then from those new components R', G' and B', one compute the XYZ transformation.[17] Then the LMS color space is obtained by apply a 3×3 matrix transformation on the three *XYZ* components corresponding to the Simth-Pokorny matrix.[18]

3.1.2. Cortex Filter decomposition

It is well known that the HVS analyzes the visual input by a set of channels, each of them being selectively sensitive to a restricted range of spatial frequencies and orientations. Several psychophysical experiments have been conducted by different researchers to characterize these channels. Currently, two transforms are often used. The cortex transform introduced by DALY[19] uses a radial frequency selectivity that is symmetric on a log frequency axis with bandwidths nearly constant at one octave. Their decompositions consist in one isotropic low-pass and three bandpass channels. The angular selectivity is constant and is equal to 45 degrees. Many different filters have been proposed as approximations to the multi-channel representation of visual information in the HVS. In this paper, a radial selectivity filter and a angular selectivity filter are used that are combined to obtain the cortex filter. Then, the reconstructed image is filtered with each cortex filter in order to obtain 31 filtered images.

Dom filter The *dom* filters–*d*ifference *o*f *m*esa–are generated by computing a difference between two consecutive *mesa* filters:

$$\mathrm{Dom}_i(u,v) = M_{i-1}(u,v) - M_i(u,v), \tag{5}$$

where u and v are the cartesian spatial frequencies. The *mesa* filter $M_i(u,v)$ is a low-pass filter of radial frequencies generated from the initial *mesa* filter $M_0(u,v)$ given by:

$$\mathrm{M}_0(u,v) = \left(\frac{\gamma}{f_0}\right)^2 \exp\left[-\pi\left(\frac{\omega\gamma}{f_0}\right)^2\right] \otimes \prod\left(\frac{\omega}{2f_0}\right) \tag{6}$$

where $\omega^2 = u^2 + v^2$.

The function $\prod\left(\frac{\omega}{2f_0}\right)$ represents a 2D gate function with circular symmetry, centered to 0 with a radius equal to f_0. γ is an attenuation parameter, linked to the standard deviation σ_0 og the Gaussian by $\sigma_0 = \frac{1}{\sqrt{2\pi}}\frac{f_0}{\gamma}$.

The *mesa* filter of index i can be expressed by:

$$\mathrm{M}_i(u,v) = \mathrm{M}_0(\tau_i u, \tau_i v) \tag{7}$$

where $\int_i$ is a scale factor given by $\tau_i = \prod_{j-1}^{i-1} \tau_j$. From eq. 7, a set of K filters can be generated from the initial filter M_0 by reducing the cut frequency of the obtained filter by the factor τ.

Fan filter The *fan* filters model the orientation attibutes of spatial frequency selectivity. This is obtained by applying a Gaussian diffusion on an ideal angular filter. From the vertical direction, this evolution is given by

$$\mathrm{M}'_0(u,v) = \mathrm{H}(v) \otimes \gamma_b \exp(-\pi\gamma_b^2 v^2) \tag{8}$$

where $\mathrm{H}(v)$ is the echelon filter, γ_b is an attenuation parameter. As the echelon filter has no variation on the axis u, the convolution can be expressed:

$$\mathrm{M}'_0(u,v) = \mathrm{F}(\gamma_b, v) = \int_{-\infty}^{v} \exp(-\pi\gamma_b^2\omega^2) d\omega \tag{9}$$

(a) Cortex filter layout in frequency domain. (b) Orientation (*Fan*) filter. (c) Radial (*Dom*) filter. (d) Resulting cortex filter.

Figure 1. Example of cortex filter obtained from the product of radial and orientation filters.

For an orientation θ, the echelon filter is:

$$\mathrm{M}'_\theta(u,v) = \mathrm{F}(\gamma_b, (v\cos\theta - u\sin\theta)). \tag{10}$$

The *fan* filter corresponding to the k^{th} direction is given by:

$$fan_{k,\theta}(u,v) = \mathrm{M}'_{\theta k}(u,v) - \mathrm{M}'_{\theta k+1}(u,v) \tag{11}$$

Cortex filter The cortex filters are simply the product of a *dom* filter and a *fan* filter in the frequency domain :

$$\mathrm{Cortex}_{k,\theta,i}(u,v) = dom_i(u,v).fan_{k,\theta}(u,v) \tag{12}$$

The image is then filtered by each one of the cortex filter to obtain a set of subimages $a_{k,\theta,i}(u,v)$ defined by

$$a_{k,\theta,i}(u,v) = \mathrm{Cortex}_{k,\theta,i}(u,v).S(u,v)$$

where $S(u,v)$ represents the image spectrum.

Each one of those images corresponds to the structural content of the image with respect to the frequency and the orientation.

Figure 1(a) shows the layout of all *dom* and *fan* filters in the spatial frequency plane. The series of arabic numbers specifies the *fan* filters. The center band orientation of the filter desinged by the number 6 is 90 degrees or -90 degrees. Each *fan* filter covers a range of 30 degrees. After six sequential *fan* filters, the same order repeats.

The series of roman numbers represents the *dom* filters at different frequency levels. The lower the serial roman number of the *dom* filter is, the lower the frequency range it resides in. Each frenquency band covers a range of one octave in the frequency domain. The first ring I is a non directional low-pass channel.

Figure 1 presents an example of a cortex filter obtained by combination of a *dom* filter and a *fan* one. Thus, when a *dom*and a *fan* filter are applied together, information od a certain frequency range and a certain orientation can be filtered out from the source image.

3.1.3. Contrast masking

Then, from each one of those filtered images, a contrast masking score is computed.

To obtain a good definition of the masking contrast, one have to take into account together the spatial and frequential resolution. Peli[20] has proposed such a model known as the limited band local contrast. This contrast is local since it quantifies the human observer's sentivitity to the luminance variation with respect to the local mean luminance. In addition, it is a limited band contrast since the degradation perception depends on its spectral location. When using the above mentionned cortex decomposition, one has to take into account both angular and radial to define the limited band local contrast such as:

$$c_{i,j}(u,v) = \frac{L_{i,j}(u,v)}{\sum_{k=0}^{i-1}\sum_{l=0}^{\mathrm{card}(l)} L^i_{k,l}(u,v)} \tag{13}$$

where $L_{ij}(u,v)$ and $c_{i,j}(u,v)$ respectively specifies the luminance and the contrast located to the coordinates (u,v) of the i^{th} radial channel and the j^{th} angular sector. card(l) represents the number of angular sectors of the k^{th} radial band.

Then, the perceived errors are modeled by the contrast masking for one spatial frequency and orientation channel and one spatial location, into a single objective score for each one of the 31 filtered image.

From this step, 31 scores, labeled to as feature s_i, are available and integrated within the feature vector.

3.2. full-reference image features

3.2.1. Structural criteria

In addition, the three criteria integrated in the metric proposed by WANG and BOVIK[21] are added to the vector. These criteria are 1) a luminance distorsion, 2) a constrast distortion and 3) a structure comparison. The authors proposed to represent an image as a vector in an image space. In that case, any image distortion can be interpreted as adding a distortion vector to the reference image vector. In this space, the two vectors that represent luminance and contrast changes span a plane that is adapted to the reference image vector. The image distortion corresponding to a rotation a such a plane by an angle can be interpreted as the structural change.

The luminance comparison is defined as

$$l(I,J) = \frac{2\mu_I\mu_J + C_1}{\mu_I^2\mu_J^2 + C_1} \tag{14}$$

where μ_I and μ_J respectively represent the mean intensity of the image I and J, and C_1 is a constant avoiding instability when $\mu_I^2 + \mu_J^2 \approx 0$. According to the Weber's law, the magnitude of a just-noticeable luminance change δL is proportional to the background luminance L. In that case, $\mu_I = \alpha\mu_J$, where α represents the ratio of the luminance of the distorted signal relative to the reference one. The luminance comparison can be now defined as

$$l(I,J) = \frac{2\alpha\mu_I^2 + C_1}{(1+\alpha^2)\mu_I^2 + C_1} \tag{15}$$

The contrast distortion measure is defined in a similar form:

$$cd(I,J) = \frac{2\sigma_I\sigma_J + C_2}{\sigma_I^2\sigma_J^2 + C_2} \tag{16}$$

where C_2 is a non negative constant, and σ_I (resp. σ_J) represents the standard deviation .

The structure comparison is performed after luminance substraction and contrast normalization. The structure comparison function is defined as:

$$s(I,J) = \frac{2\sigma_{I,J} + C_3}{\sigma_I^2\sigma_J^2 + C_3} \tag{17}$$

where $\sigma_{IJ} = \frac{1}{N-1}\sum_{i=1}^{N}(I_i - \mu_i)(J_i - \mu_J)$, and C_3 is a small constant. $s(I,J)$ can take negative values which is interpreted as local image structures inversion.

3.2.2. color criteria

Two local descriptors based on visual attention are used.[22] Those descriptors are not ponctually defined in $I(x,y)$ but with respect to the mean value $\mu(x,y)$ of neigborhood V of the pixel (x,y) . $I^M_{(c_i)}(x,y)$ and $I^m_{(c_i)}(x,y)$ respectively represent the maximal and minimal value of the c_i axis within V for the image I at the pixel located to (x,y).

The two used features are:

1. local chrominance that measures the sensitivity of an observer to color degradation within a uniform area. The calculation of this descriptor is performed in the $L^*a^*b^*$ color space.
2. local colorimetric dispersion that measures the spatio-colorimetric dispersion in each one of the two color images. This comparison which is performed over a neighborhood.

These descriptors have been defined according to the same scale ranging from 0 to 1 ; 0 corresponding to the most noticeable differences and 1 corresponding to the least noticeable difference.

3.3. No-reference image SVH-based features

3.3.1. Blockiness measures

The measure of blocking artefact has an important weight in the final judgment of the image quality. Blocking artefact results from a visible block structure appearing in reconstructed images. This is mainly due to the block-based compression algorithms used to compressed images.

Many techniques to quantify blockiness effects have been proposed. Yet, these techniques require to have access to the original image. In our case, three no-reference blockiness metrics have been used: 1) the WANG *et al.* one,,[23] 2) the one developed by VLACHOS[24] and 3) the blockiness measure as defined by WU and YUEN.[25]

WANG *et al.* model the blocky image as a non-blocky image interfered with a pure blocky signal. To estimate the blockiness of an image M_B, they assume that the vertical M_{B_v} and th horizontal M_{B_h} effects are of the same importance, and the relationship between M_B and both M_{B_v} and M_{B_h} effects is $M_B = (M_{B_v} + M_{B_h})/2$.

In order to define the two effects, they apply a 1-D FFT to the horizontal and vertical difference signals. From these signals, the average horizontal and vertical power spectra is computed. Peaks in these spectra are then identified by their locations in the spectra. The power spectra of non-blocky images is approximated by applying a median filter on these curves. The final blockiness measure is then computed as the difference between the resulting power spectra and the peaks location.

In the VLACHOS's model, the used algorithm is based on the cross-correlation of subsmapled images. The original image is first decomposed in 8×8 size blocks. Every generated sub-image contains one specefic pixels from each original blocks. Eight sub-images are generated as follows: 1) four sub-images are generated from the four corner pixels of each blocks and 2) four other sub-images are constructed from four neighboring pixels in the top left corner of each block. Then, the cross-correlations from the former four sub-images are normalized by the computed cross-correlations of the latter four sub-images to score the blockiness measure.

For WU and YUEN the blockiness measure is based on the vertical and horizontal differences between the columns and the rows all 8×8 boundaries. The mean and the standard deviation obtained from adjacent blocks to each boundary are respectively used to define weight respectively dedicated to perceptual luminance effects and texture masking effects. Then, the final blockiness measure is obtained from the latter measure normalized by the mean of the same measures computed at non-boundary columns and rows.

3.3.2. Blurriness measure

Blur in an image is due to the attenuation of high spatial frequencies in the image. It is characterized by a smearing of sharp edges and a general loss of details. This artefact commonly occurs during a compression process.

To measure the blurry effect on an image, one first have to detect edges. In order to detect edges in a color image, the Cumani edge detector[26] is used on the color image expressed in the Krauskopf space. Then an opening operator from mathematical morphology is applied on the resulting thresholded image in order to remove noise and non-important edges. Then, taking into account the gradient orientation, a measure of blurriness is performed along the actual local edges. To measure the width of a located edge, one used the Achromatic component and seeds obtained from the former edge image. Then both local maximum and local minimum are extracted from the achromatic image and correspond to luminance extrema closest to the edge seed. The difference between the location of those two extrema is computed to define the width of the edge. Finally the global blurriness measure is obtained by averaging the local blurriness measure over all detected edges.

3.4. No-reference image features

As no-reference image features, one has used three of the objectives features used by GASTALDO *et al.* to provide an objective assessment of JPEG compressed image using neural networks.[27] The three features are expressed from a color correlogram that allows us to have information about the spatial correlation of color changes with distance. The used features are:

1. Energy, that corresponds to a summation of all squared elements of the color correlogram,
2. Information Entropy correponding to amount of information bringing by the color correlogram

3. Coefficient of homogeneity indicating the degree data approximates the Guttman implicatory scales. It measures the consistency of data matrices

Therefore the final feature vector contains 42 attributes $(s_i)_{i\in[1,...,42]}$.

4. MEASURE OF PERFORMANCE

Two datasets are realized from 227 different JPEG2000 compressed image versions of 25 initial images in the LIVE image database.[28] The 38 factors given in the previous section are computed for each compressed image. 25 initial images are used as reference for the computation of those factors. First dataset of 116 compressed images defines the training set used to learning phase of the machine expert. Second dataset of 111 remaining compressed images defines the test set used to evaluate the efficiency of the machine learning expert. Respectively 29 and 25 observers give an opinion score for images in training and test set. Opinion scores and mean opinion score of observers are converted to quality scale of the UIT and are respectively noted Q_{OS} and Q_{MOS}. The table 3 illustrates the percentage of images in each quality class category in function of observer s Q_{MOS} and dataset.

Table 3. Percentage of images in each quality class.

Q_{MOS} class	1	2	3	4	5
Training set	12.9%	39.7%	25.9%	16.4%	5.1%
Testing set	13.5%	36.1%	14.4%	24.3%	11.7%

To measure the efficient of machine learning expert, three coherence measures M are defined from three loss functions LO:

$$M_a = 1 - \frac{1}{m}\sum_{i=1}^{m} LO_a\left(D(i), Q_{\mathrm{MOS}}(i)\right) \tag{18}$$

where $D(i)$ is the quality decision from machine learning expert for the image i and $Q_{\mathrm{MOS}}(i)$ represents the MOS for the image i. $a \in \{1,2,3\}$ corresponds to the loss function used. The three loss functions are the following:

$$LO_1\left(y_1, y_2\right) = \begin{cases} 0 & \text{if } y_1 = y_2 \\ 1 & \text{else} \end{cases} \tag{19}$$

$$LO_2\left(y_1, y_2\right) = \begin{cases} 0 & \text{if } |y_1 - y_2| \leq 1 \\ 1 & \text{else} \end{cases} \tag{20}$$

$$LO_3\left(y_1, y_2\right) = \begin{cases} 0 & \text{if } y_1 = y_2 \\ \frac{m}{m_{y_2}} & \text{else} \end{cases} \tag{21}$$

where m_{y_2} corresponds to the number of images labelled to as class y_2 in the reference dataset. M_1 is a classical measure of recognition rate. M_2 measures the rank coherence of a quality decision prediction. For example, if *excellent* or *quite good* is image quality prediction and its associated Q_{MOS} is *good*, then this small difference in appreciation could be tolerated, but not for a *bad* or *very bad* quality prediction. Particularly, when observing the Q_{OS} values, for many images 90% of the observers select 2 classes (3 sometimes) which are very close in terms of their ranking. From this remark, the importance of the M_2 measure is highlighted. M_3 is a measure that takes into account the relative proportion of each quality class in a dataset. This permits to verify that low representative class are not discarded by classifier. This effect could artificially increase both M_1 and M_2 measures, but it is a kind of over-fitting effect which must be avoided. A great difference between the M_1 and the M_3 measure could detect this effect.

Before measuring the efficiency of machine learning expert, we have measured how each observer is confident with Q_{MOS}. To doing this, the Q_{OS} of each observer is used as a decision function D with respect to the three coherence

measures M. Tables 4 and 5 respectively show statistical informations on the Q_{MOS} observer s confidences obtained from the training set and the test one. Those results show that the observer's opinion have a great variability. This variability is greatly independent of the used dataset. The M_2 confidence measure shows that divergence with Q_{MOS} rarely exceeds one class for the raking order (even for an observer which is the most faraway from the Q_{MOS} reference).

Table 4. Observers statistics of coherence measures for the training set.

	Mean	min	Max
M_1	0.558 ± 0.078	0.405	0.698
M_2	0.989 ± 0.019	0.914	1.000
M_3	0.594 ± 0.057	0.462	0.669

Table 5. Observers statistics of coherence measures for the test set.

	Mean	min	Max
M_1	0.529 ± 0.103	0.324	0.712
M_2	0.975 ± 0.026	0.909	1.000
M_3	0.550 ± 0.090	0.396	0.706

The Machine Learning Expert (MLE) is built using the two binary SVM combination schemes defined in the previous section. To measure the influence of the used features, four experiments have been realized for each combination scheme with: 1) all full-reference features, 2) a feature selection from all full-reference features, 3) all no-reference features and 4) a feature selection from all no-reference features. For experiments 2) and 4), the most relevant features are selected by using the best-first-search algorithm[29] in a wrapper feature selection approach.[10]

Due to the small size of the training set, a model selection for each binary SVM involved in the MLE is performed by a leave-one-out cross-validation (LOO-CV) measure. For speed-up LOO-CV evaluation with each binary SVM, especially when a feature selection is realized, a specific alpha seeding SVM method is used.[30] For each of the eight designed MLE, coherence measures are computed from the training and the test one. The obtained results are summarized in Tables 6 and 7.

When comparing the results from Table 5 and Table 7, one observes that the efficiency of ours MLEs is good. When considering the full-reference columns of Table 7, our MLEs are more coherent with the Q_{MOS} than the average selection of the observers (Table 5). The MLE summarizes very well the mean behaviour of the observers, especially when the M_2 measure is observed. Then, all the errors of final class prediction measured by M_1 or M_3 criterion on the test set correspond to a choose of the quality class which differs from the Q_{MOS} with no more than one class. Statistics shown in table 4 and 5 present the same similarity for human beings. That way, our MLEs are in concordance with the human being's prediction.

When comparing no-reference columns to full-reference columns, it could be observed a fewer decrease of the MLEs efficiency. No-reference MLEs remains good quality predictor. Especially, when observing M_2 measure and also taking into account that no-reference problem is more difficult than full-reference problem.

For the two reference problems, the M_3 measures show that our MLE does not neglect the few representative classes. Moreover, our RO combination scheme is more sensitive to the ranking order information and could improve MLEs when M_1 and M_3 measures are considered. That combination scheme could also improve MLEs with M_2 mesure under the no-reference problem. When a feature selection is realized with full-reference features (no-reference features), in average only 4.2 and 3.5 (2.6 and 2.75) features are respectively used by binary classifiers in the OAA and the RO combination schemes. From the Table 7, the M_2 measure show that efficient MLEs could be produce by using few features, expecially on full-reference problem. Even if, using both M_1 and M_2 one observes taht MLEs are

less efficients. However, the small dataset size does not allow us to conclude more on relevance or irrelevance status of some features with a high confidence for the two reference problems. On way to overcome this problem could be the use of an efficient bootstrap technique as used in genomic problems.[31]

Table 6. Table 7: MLE coherence measures using the training set.

	full-reference				no-reference			
	all-features		feature selection		all-features		feature selection	
	OVA	RO	OVA	RO	OVA	RO	OVA	RO
M_1	0.922	0.922	0.870	0.888	0.991	0.802	0.853	0.870
M_2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
M_3	0.880	0.880	0.753	0.844	0.986	0.659	0.811	0.840

Table 7. Table 8: MLE coherence measures using the test set.

	full-reference				no-reference			
	all-features		feature selection		all-features		feature selection	
	OVA	RO	OVA	RO	OVA	RO	OVA	RO
M_1	0.730	0.747	0.703	0.712	0.504	0.703	0.558	0.504
M_2	1.000	1.000	1.000	1.000	0.973	1.000	0.892	0.982
M_3	0.689	0.727	0.686	0.697	0.419	0.651	0.444	0.399

5. CONCLUSION

A color image quality metric based on Machine Learning Expert is introduced. The MLE only learns on 1) the MOS of the observers and 2) several human visual system features to characterize the quality of color images. The used features are no-reference characteristics and full-reference features. One observes that the obtained results only using the no-reference features are close to those obtained only using full-reference features to predict the final image quality. The MLE can modelize with a great efficiency the mean behaviour of the observers. The efficiency of the MLE is deeply linked to the design of a good similarity measure. In this paper, the construction of such a measure is based machine learning approach. The obtained results shows that this kind of strategy is a new promising way to investigate the image quality measure. Atually, it is more natural for human beings to classifiy the quality of a color image than to score it.

In future works, to each decision taken by the classifier an a posteriori probability will be associated in order to define a meta-classifier by combining multi-model classifiers.

REFERENCES

1. Z. Wang, A. C. Bovik, and E. P. Simoncelli, "Structural approaches to image quality assessment," in *Handbook of Image and Video Processing*, Academic Press, 2nd ed., 2005.
2. UIT-R Recommendation BT.500-10, "Mthodologie d'valuation subjective de la qualit des images de tlvision," tech. rep., UIT, Geneva, Switzerland, 2000.
3. C. Charrier, G. Lebrun, and O. Lezoray, "Selection of features by a machine learning expert to design a color image quality metric," in *Third Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, pp. 113–119, (Scottsdale, Arizona), 2007.
4. G. Lebrun, C. Charrier, O. Lezoray, C. Meurie, and H. Cardot, "Fast pixel classification by SVM using vector quantization, tabu search and hybrid color space," in *the 11th International Conference on CAIP*, pp. 685–692, (Rocquencourt, France), 2005.

5. V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
6. J. Platt, *Fast Training of Support Vector Machines using Sequential Minimal Optimization, Advances in Kernel Methods-Support Vector Learning*, MIT Press, 1999.
7. R. Collobert and S. Bengio, "SVMTorch: Support vector machines for large-scale regression problems," *Journal of Machine Learning Research* **1**, pp. 143–160, 2001.
8. C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines." Sofware Available at http://www.csie.ntu.edu.tw/~cjlin/libsvm, 2001.
9. C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks* **13**(3), pp. 415–425, 2002.
10. R. Kohavi and G. H. John, "Wrappers for feature subset selection," *JAIR* **97**(1-2), pp. 273–324, 1997.
11. S. Daly, "A visual model for optimizing the design of image processing algorithm," in *ICIP*, **2**, pp. 16–20, 1994.
12. A. B. Watson, "The cortex transform: Rapid computation of simulated neural images," *Computer Vis. Graphics and image proces.* **39**, pp. 311–327.
13. J. Lubin, *Digital Images and Human Vision*, ch. The use of psychophysical data and models in the analysis of display system performance, pp. 163–178. MIT Press, a. watson, ed. ed., 1993.
14. G. Sharma and H. Trussell, "Digital color imaging," *IEEE Transactions on Image Processing* **6**, pp. 901–932, july 1997.
15. G. D. Finlayson, "Color in perspective," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**, pp. 1054–1058, 1996.
16. J. Krauskopf, D. R. Williams, and D. W. heeley, "Cardinal directions of color space," *Vision Research* **22**, pp. 1123–1131, 1982.
17. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, John Wiley & sons, second ed., 1982.
18. F. Vienot, H. Brettel, and J. D. Mollon, "Digital video color maps for checking the legibility of displays by dichromats," *Color Research and Application* **24**(4), pp. 243–252, 1999.
19. S. Daly, "The visible differences predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, pp. 179–206, The MIT Press Cambridge, 1993.
20. E. Peli, "Contrast in complex images," *Journal of the Optical Society of America* **7**, pp. 2032–2040, Oct. 1990.
21. Z. Wang and A. C. Bovik, "A universal quality index," *IEEE Signal Processing Letters* **9**(3), pp. 81–84, 2002.
22. A. Trémeau, C. Charrier, and E. Favier, "Quantitative description of image distorsions linked to compression schemes," in *Proceedings of The Int. Conf. on the Quantitative Description of Materials Microstructure*, (Warsaw), Apr. 1997. QMAT'97.
23. Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in *International Conference on Image Processing*, pp. 981–984, (Vancouver, BC), Sept. 2000.
24. T. Vlachos, "Detection of blocking artifacts in compressed video," *Electronics Letters* **36**(13), pp. 1106–1108, 2000.
25. H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Processing Letters* **4**(11), pp. 317–320, 1997.
26. A. Cumani, "Edge detection in multispectral images," *Graphical Models and Image Processing* **53**, pp. 40–51, Jan. 1991.
27. P. Gastaldo, G. Parodi, J. Redi, and R. Zunino, "No-reference quality assessment of JPEG images by using CBP neural networks," in *ICANN 2007*, **LNCS 4669**, pp. 564–572, 2007.
28. Laboratory for Image & Video Engineering, University of Texas (Austin), "LIVE Image Quality Assessment Database," *http://live.ece.utexas.edu/research/Quality* , 2002.
29. J. Pearl, *Heuristics: Intelligent Search Strategies for Computer Problem Solving*, Addison-Wesley, 1984.
30. G. Lebrun, O. Lezoray, C. Charrier, and H. Cardot, "Speed-up LOO-CV with SVM classifier," in *IDEAL*, pp. 108–115, 2006.
31. C. Ambroise and G. J. McLachlan, "Selection bias in gene extraction on the basis of microarray gene-expression data.," *Proc Natl Acad Sci U S A* **99**(10), pp. 6562–6566, 2002.