A Machine Learning-based Color Image Quality Metric

Christophe Charrier Gilles Lebrun Olivier Lezoray

Université de Caen-Basse Normandie, LUSAC EA-2607, Groupe Vision et Analyse d'Images
120, route de l'exode, 50000 Saint-Lô, France – Phone : +33 (0)2 33 77 55 11 - Fax : +33 (0)2 33 77 11 67
E-mail: {c.charrier,g.lebrun,o.lezoray}@chbg.unicaen.fr

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. In that way, a machine learning expert, providing a final class number is designed.

Introduction

When designing a quality metric (including or not Human Visual System features), the main weakness is the computation of the final score. Actually, 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 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 obtained 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.

Actually, the final goal of each metric is to score by a single note the quality of an image. Then, to assess how the metric performs well, a correlation measure between the designed quality metric value and the Mean Opinion Score (MOS) is computed. The MOS is obtained from a set of human observers scores with respect to a normalized scale defined by the UIT [2]. The higher degree, the more the metric proceeds as the human observer does.

From the above remarks, one main constatation can be formulated: the need to obtain a final quality score is not necessary the best way to quantify the quality. Actually, in the recommendations given by the UIT [2], the human observers have to choose a quality class from a scale containing five notes (cf. Table 1). Those notes characterize the quality of the reconstructed images. In that way, the human observers make then neither more nor less one classification.

In this paper, the quality measure is based on a learned classification process in order to respect the one of human observers. 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 ordered from 1 (the worst quality) to 5 (the best quality). The selected class of the proposed method represents the opinion score OS. In that way, a machine learning expert, providing a final class number is designed.

Table 1: Quality scale of the UIT-R.

	Quality
5	Excellent
4	Good
3	Quite good
2	Bad
1	Very bad

The proposed approach

The used 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 [3], and to their high generalization abilities.

The SVMs were developed by VAPNIK ET AL. [4] 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 $\mathscr{S} = \{(x_i, y_i)\}_{i=\{1,\ldots,m\}}, x_i \in \mathscr{R}^n$, $y_i \in \{-1,+1\}$, SVM maps the input vector x into a high-dimensional feature space $\mathbf{H}$ through some non linear mapping functions $\phi : \mathscr{R}^n \rightarrow \mathbf{H}$, 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:

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

where the constraint $\forall_{i=1}^{m} : y_i[w \cdot \phi(x_i) + b] \geq 1 - \xi_i$, $\xi_i \geq 0$ 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 [4] is to maximize the following optimization problem:

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

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:

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

where the threshold b^* is computed via the unbounded support vectors [4] (*i.e.*, $0 < \alpha_i^* < C$). An efficient algorithm SMO (Sequential Minimal Optimization) [5] and many refinements [6, 7] 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 [8].

The used parameters

In order to use SVMs, a vector of features has to be designed. To create this vector, the HVS models have been considered.

Color space transformation

The first step of the proposed scheme concerns the colorimetric transformation of the initial coordinates system, *i.e.*, the *RGB* space. 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) [9]. 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 [10]. 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 [11] one is selected. This coordinates system is computed from the LMS primaries that correspond to the HVS cone responses :

$$\begin{aligned} A &= \frac{1}{2}(L+M) \\ C_1 &= \frac{1}{2}(L-M) \\ C_2 &= \frac{1}{2}(S - \frac{L+M}{2}) \end{aligned} \quad (4)$$

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 [12]. 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 [13].

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 [14] 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:

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

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:

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

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:

$$\text{M}_i(u,v) = \text{M}_0(\tau_i u, \tau_i v) \quad (7)$$

where f_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

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

where $\text{H}(v)$ 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:

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

For an orientation θ, the echelon filter is:

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

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

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

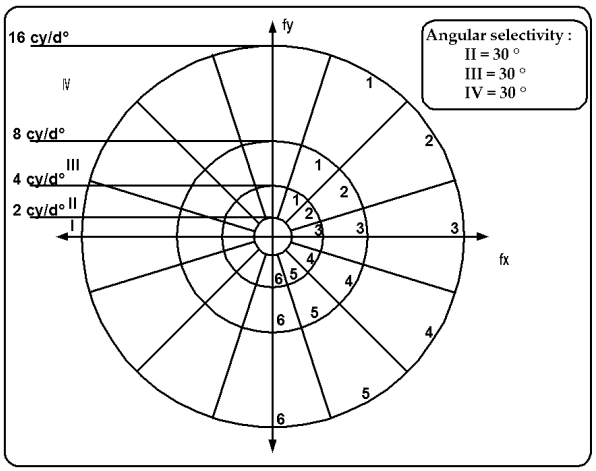
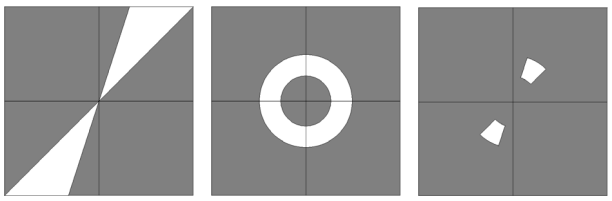


Figure 1. *Cortex filter layout in frequency domain.*



(a) Orientation (*Fan*) filter. (b) Radial (*Dom*) filter. (c) Resulting cortex filter.

Figure 2. *Example of cortex filter obtained from the product of radial and orientation filters.*

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

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

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

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

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 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 2 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.

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 [15] 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:

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

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.

Structural criteria

In addition, the three criteria integrated in the metric proposed by WANG and BOVIK [16] 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.

color criteria

Two local descriptors based on visual attention are used [17]. 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).

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 as follows:

$$D_{\mathrm{C}}(x,y) = 1 - \frac{\sqrt{(\mu^1_{a^*}(x,y) - \mu^2_{a^*}(x,y))^2 + (\mu^1_{b^*}(x,y) - \mu^2_{b^*}(x,y))^2}}{\sqrt{(I^M_{a^*}(x,y) - I^m_{a^*}(x,y))^2 + (I^M_{b^*}(x,y) - I^m_{b^*}(x,y))^2}} \quad (18)$$

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 is defined as:

$$D_{\mathrm{Co}}(x,y) = \frac{\|\sum_{i=1}^{3}\mathrm{cov}^{1,2}_{\mathrm{c}_i}(x,y)\|}{\sqrt{\sum_{i=1}^{3}\sigma^1_{\mathrm{c}_i}(x,y)^2}\cdot\sqrt{\sum_{i=1}^{3}\sigma^2_{\mathrm{c}_i}(x,y)^2}} \quad (19)$$

c_i represents the considered color component. $\sigma^1_{c_i}(x,y)$ represents the neighborhood variance $V(x,y)$ of the specified pixel from image 1, while $\mathrm{cov}^{1,2}_{c_i}(x,y)$ is the neighborhood covariance $V(x,y)$ of the specified pixel from image 1 with respect to image 2. Thus

$$\sigma^1_{c_i}(x,y)^2 = \frac{1}{\mathrm{card}\, V(x,y)}\sum_{(x',y')\in V(x,y)}\left(I^1_{c_i}(x',y') - \mu^1_{c_i}(x,y)\right)^2$$

$$\mathrm{cov}^{1,2}_{c_i}(x,y) = \frac{1}{\mathrm{card}\, V(x,y)}\sum_{(x',y')\in V(x,y)}\left(I^1_{c_i}(x',y') - \mu^1_{c_i}(x,y)\right).\left(I^2_{c_i}(x',y') - \mu^2_{c_i}(x,y)\right)$$

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.

Finally two commonly used quality measures have been selected as features: 1) the color MAE (Mean Average Error) and 2) the color PSNR (Peak Signal to Noise Ratio) [18].

Therefore the final feature vector contains 38 attributes $(s_i)_{i\in[1,\ldots,38]}$.

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

$$K_\beta(s_i,s_j) = exp(-\sum_{k=1}^{n}\beta_k(s^k_i - s^k_j)^2/r^2) \quad (20)$$

where s^k_i 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 [19]. 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 [8]. 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 2 and 3 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 2: 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

Table 3: 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

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 [8]. The majority vote criterium is the usual way to do this. Let $V_j(x) = \sum_{i=1}^{n_b} LO_1\left(h_i(x), t_{i,j}\right)$ 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: $D(x) = \arg\max_{1 \le j \le n_c} \left(V_j(x)\right)$ (when conflicts exist, the SVM output is used to break it).

Measure of performance

Two datasets are realized from 227 different JPEG2000 compressed image versions of 25 initial images in the LIVE image database [20]. 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 4 illustrates the percentage of images in each quality class category in function of observer s Q_{MOS} and dataset.

Table 4: 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:

$$M_a = 1 - \frac{1}{m}\sum_{i=1}^{m} LO_a\left(D(i), Q_{\mathrm{MOS}}(i)\right) \tag{21}$$

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:

$$LO_1(y_1, y_2) = \begin{cases} 0 & \text{if } y_1 = y_2 \\ 1 & \text{else} \end{cases} \tag{22}$$

$$LO_2(y_1, y_2) = \begin{cases} 0 & \text{if } |y_1 - y_2| \le 1 \\ 1 & \text{else} \end{cases} \tag{23}$$

$$LO_3(y_1, y_2) = \begin{cases} 0 & \text{if } y_1 = y_2 \\ \frac{m}{m_{y_2}} & \text{else} \end{cases} \tag{24}$$

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 5 and 6 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 5: Obervers 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 6: Obervers 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, three experiments have been realized for each combination scheme: 1) all the 38 features are used, 2) only the 31 features provided by the cortex decomposition are used, and 3) the more relevant features are selected by using the best-first-search algorithm [21] in a wrapper feature selection approach [19].

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 [8]. For each of the six designed MLE, coherence measures are computed from the training and the test one. The obtained results are summarized in Tables 7 and 8. When comparing the results from Table 6 and Table 8, one observes that the efficiency of ours MLEs is good. When considering the two first columns of Table 8, our MLEs are more coherent with the Q_{MOS} than any observer (Table 6). The MLE summarizes very well the mean behaviour of the observers, especially when examining the M_2 measure. This measure shows also that our RO combination scheme is more sensitive to the ranking order information. The M_1 and M_3 measures show that our MLE does not neglect the few representative classes.

Table 7: MLE coherence measures using the training set.

	38 features		31 features		feature selection	
	OVA	RO	OVA	RO	OVA	RO
M_1	0.931	0.931	0.922	0.914	0.957	0.965
M_2	0.983	1.000	0.983	1.000	0.991	1.000
M_3	0.873	0.881	0.840	0.863	0.929	0.906

Table 8: MLE coherence measures using the test set.

	38 features		31 features		feature selection	
	OVA	RO	OVA	RO	OVA	RO
M_1	0.801	0.801	0.793	0.784	0.693	0.639
M_2	0.982	1.000	0.991	1.000	0.946	1.000
M_3	0.747	0.781	0.741	0.775	0.672	0.598

From the Table 8, values in the two first columns show how the contrast masking and the structural feature slightly increase the efficiency of the MLE. Nevertheless, the small dataset size does not allow us to conclude more on their relevance. When a feature selection is realized, only 13 and 10 features are respectively used for the OAA and the RO combination schemes.

In addition, the recognition rate increases when the training set is used, whereas it decreases when the test set is used. In that case, an over-fitting effect can be assumed. Yet, the irrelevance (or heavily correlated) status of some attributes could not be determined with a high confidence. The small dataset size is the main reason of the over-fitting effect. On way to overcome this problem could be the use of an efficient bootstrap technique as used in genomic problems [22].

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 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, a study about the importance of each selected features will be investigated. To perfrom this, one have to construct an image database containing several thousand compressed images.

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, "Méthodologie d'évaluation subjective de la qualité des images de télévision," tech. rep., UIT, Geneva, Switzerland, 2000.

[3] 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*, (Rocquencourt, France), pp. 685–692, 2005.

[4] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[5] J. Platt, *Fast Training of Support Vector Machines using Sequential Minimal Optimization, Advances in Kernel Methods-Support Vector Learning*. MIT Press, 1999.

[6] R. Collobert and S. Bengio, "SVMTorch: Support vector machines for large-scale regression problems," *Journal of Machine Learning Research*, vol. 1, pp. 143–160, 2001.

[7] 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.

[8] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 415–425, 2002.

[9] G. Sharma and H. Trussell, "Digital color imaging," *IEEE Transactions on Image Processing*, vol. 6, pp. 901–932, july 1997.

[10] G. D. Finlayson, "Color in perspective," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 1054–1058, 1996.

[11] J. Krauskopf, D. R. Williams, and D. W. heeley, "Cardinal directions of color space," *Vision Research*, vol. 22, pp. 1123–1131, 1982.

[12] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*. John Wiley & sons, second ed., 1982.

[13] F. Vienot, H. Brettel, and J. D. Mollon, "Digital video color maps for checking the legibility of displays by dichromats," *Color Research and Application*, vol. 24, no. 4, pp. 243–252, 1999.

[14] 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.

[15] E. Peli, "Contrast in complex images," *Journal of the Optical Society of America*, vol. 7, pp. 2032–2040, Oct. 1990.

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

[17] 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.

[18] K. N. Platoniotis and A. N. Venetsanopoulos, *Color Image Processing and Applications*. Springer, 2000.

[19] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *JAIR*, vol. 97, no. 1-2, pp. 273–324, 1997.

[20] Laboratory for Image & Video Engineering, University of Texas (Austin), "LIVE Image Quality Assessment Database," *http://live.ece.utexas.edu/research/Quality*, 2002.

[21] J. Pearl, *Heuristics: Intelligent Search Strategies for Computer Problem Solving*. Addison-Wesley, 1984.

[22] 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*, vol. 99, no. 10, pp. 6562–6566, 2002.