Image Quality Assessment with Manifold and Machine Learning

Christophe Charrier, Gilles Lebrun, Olivier Lezoray

Image Quality Assessment with Manifold and Machine Learning

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 crucial step in image compression is the evaluation of its performance, and more precisely the available way to measure the final quality of the compressed image. In this paper, a machine learning expert, providing a final class number is designed. 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). This was done constructing a vector containing many visual attributes. Finally, the final features vector contains more than 40 attibutes.

Unfortunatley, no study about the existing interactions between the used visual attributes has been done. A feature selection algorithm could be interesting but the selection is highly related to the further used classifier. Therefore, we prefer to perform dimensionality reduction instead of feature selection. Manifold Learning methods are used to provide a low-dimensional new representation from the initial high dimensional feature space.

The classification process is performed on this new low-dimensional representation of the images. Obtained results are compared to the one obtained without applying the dimension reduction process to judge the efficiency of the method.

Keywords: Manifold learning, Machine Learning, SVM, Image quality, Human Visual System

1. INTRODUCTION

Nowadays, it is usual for anyone to take photos with digital cameras, to upload the images on computers, and to use some software to apply many image processing algorithms on these images (compression, blurring, ...). This is a simple and representative example of the growing of digital media that is everywhere. By the way, many Tera-bytes transit on the Internet. In order to reduce the amount of transmitted data, one typical applied processing on an image is compression, that allows to reach high compression levels, so that few data is to be further transmitted.

Image compression maps an original image into a bit stream suitable for communication over or storage in a digital medium. It consists in one or more of the following operations :

- Signal representation: the first and the most important element is the transform of the image to the most suitable domain. A frequential or spatial transform is applied to have an efficient image representation. The goal is to concentrate energy in a few coefficient or to provide a useful data structure.
- Quantization: this conversion can operate on individual pixels (scalar quantization) or group of pixels (vector quantization). This operation is nonlinear and noninvertible: it is a lossy process. In our study, this conversion operates on a group of pixels by Vector Quantization (VQ).
- lossless compression: the compression is achieved by invertible code. The idea is to assign codewords with few bits to likely symbols and codewords with more bits to unlikely symbols so that the average number of bits is minimized.

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

A crucial step in image compression is the evaluation of its performance, and more precisely the available way to measure the final quality of the compressed image. There is a very rich literature on image quality criteria, generally dedicated to specific applications (optics, detector, compression, restoration, ...). The quality evaluation is divided into two topics: objective and subjective evaluation. The first topic gives place to two families of criteria: unweighted and weighted criteria. The first family corresponds to the traditional criteria known as mathematical measures, because they result from geometry (concept of distance) or from the signal processing (signal to noise ratio). These criteria do not give an estimate of the visual quality of the image. The second family of criteria takes into account the characteristics of the human visual system in particular by a weighting of the image of error. Lastly, the second topic relates to the psychophysical experiments allowing to add a subjective dimension in the quality evaluation process. Due to the time expensive aspect of this last topic, objective quality measures have been intensively investigated to quantify the quality of a compressed image.

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. A final quality score is obtained by pooling these errors in both spatial and frequential domain. The most common way to perform this pooling is to use the Minkowski error metric. Some studies[1] have shown that this summation does not perform well. The same final value can be computed for two different degraded images even if the visual quality of the two images is drastically different. This can be due to the fact that the implicit assumption of this metric is based on the independancy of all signal samples. It is commonly assumed that this is not true when one uses perceptual channels. This explains the reason why the Minkowski metric might fail to generate a good final score.

In our previous work, a machine learning expert, providing a final class number and its associated confidence probability, has been designed.[2] 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. This was done constructing a vector containing many visual attributes such as those defined by Wang *et al.* in their SSIM measure,[3] and ones obtained computing a cortex filter based on a multi-channel decomposition,[4] and so on. Finally, the final features vector contains more than 40 attibutes.

Unfortunatley, no study about the existing interactions between the used visual attributes has been done so far. A feature selection algorithm could be interesting but the selection is highly related to the further used classifier. Therefore, we prefer to perform dimensionality reduction instead of feature selection. Linear dimensionality reduction methods have shown their limits and a lot a different nonlinear methods have emerged these last years. This is known as Manifold Learning and attempts at providing a low-dimensional representation from an initial high dimensional feature space.

The paper is structured as follows. In Section 2, we present the of features used to describe the quality of images. In Section 3, Manifold Learning methods are described. Section4 details how classification is performed with Support Vector Machines. Section 5 provides obtained results. Last Section concludes.

2. FEATURES VECTOR

This section describes the set of features used to describe the quality of images. Four types of attributes contained in that set are: 1) full-reference image SVH-based features and 2) full-reference image features, both for them a reference image is needed and 3) no-reference images SVH-based features and 4) no-reference images features, both for them no reference image is needed.

2.1 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. From all existing opponent color spaces, the Krauskopf[5] one is selected. This coordinates system is computed from the LMS primaries that correspond to the HVS cone responses. Thus, from this color space one are able to take into account the spatial-frequency sensitivity of the SVH. Actually, 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, concerning the luminance component the most used decompositions are from DALY,[4] WATSON[6] and the LUBIN's one.[7] are often used. 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.

In this paper, the cortex transform introduced by DALY[8] is used. Actually that transformation 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 $dom_i(u,v)$ and a angular selectivity filter $fan_{k,\theta}(u,v)$ are used that are combined to obtain the cortex filter $mboxCortex_{k,\theta,i}(u,v) = dom_i(u,v).fan_{k,\theta}(u,v)$, where u and v are the cartesian spatial frequencies, θ is the orientation and k represents the direction. Then, 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.

2.1.1 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[9] 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:

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

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.

2.2 full-reference image features

2.2.1 Structural criteria

In addition, the three criteria integrated in the metric proposed by WANG and BOVIK[3] 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

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

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

$$l(I,J) = \frac{2\alpha\mu_I^2 + C_1}{(1+\alpha^2)\mu_I^2 + C_1} \tag{3}$$

The contrast distortion measure is defined in a similar form:

$$cd(I,J) = \frac{2\sigma_I\sigma_J + C_2}{\sigma_I^2\sigma_J^2 + C_2} \tag{4}$$

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:

$$s(I,J) = \frac{2\sigma_{I,J} + C_3}{\sigma_I^2\sigma_J^2 + C_3} \tag{5}$$

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.

2.2.2 color criteria

Two local descriptors based on visual attention are used.[10] 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.

2.3 No-reference image SVH-based features

2.3.1 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,,[11] 2) the one developed by VLACHOS[12] and 3) the blockiness measure as defined by WU and YUEN.[13]

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.

2.3.2 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[14] 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.

2.4 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.[15] 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

3. DIMENSIONALITY REDUCTION

Given a set of visual attributes describing an image, we use a Manifold Learning method to project the data on a new low-dimensional space. Thus, nonlinear new discriminant features of the input data are yielded. The obtained low dimensional sub-manifold is used as a new representation that is transmitted to classifiers.

When data objects that are the subject of analysis using machine learning techniques are described by a large number of features (i.e. the data is high dimension) it is often beneficial to reduce the dimension of the data. Dimension reduction can be beneficial not only for reasons of computational efficiency but also because it

can improve the accuracy of the analysis. Indeed, traditional algorithms used in machine learning and pattern recognition applications are often susceptible to the well-known problem of the curse of dimensionality, that refers to the degradation in the performance of a given learning algorithm as the number of features increases. To deal with this issue, dimension reduction techniques are often applied as a data pre-processing step or as part of the data analysis to simplify the data model. This typically involves the identification of a suitable low-dimensional representation for the original high-dimensional data set. By working with this reduced representation, tasks such as classification or clustering can often yield more accurate and readily interpretable results, while computational costs may also be significantly reduced. Dimensionality reduction methods can be divided into two sets wether the transformation is linear or nonlinear. We detail here the principles of two well-known linear and nonlinear dimensionality reduction methods: Principal Components Analysis (PCA)[16] and Laplacian Eigenmaps (LE).[17] Let $X = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\} \in \mathbb{R}^p$ be n sample vectors. Dimensionality reduction consists in finding a new low-dimensional representation in $\mathbb{R}^p$ with $q \ll p$.

3.1 Principal Components Analysis

The main linear technique for dimensionality reduction, principal components analysis (PCA), performs a linear mapping of the data to a lower dimensional space in such a way, that the variance of the data in the low-dimensional representation is maximized. Traditionally, principal component analysis is performed on the symmetric covariance matrix C_{cov} or on the symmetric correlation matrix C_{cor}. We will denote C one of these two matrices in the sequel. From such a symmetric matrix, we can calculate an orthogonal basis by finding its eigenvalues and eigenvectors. Therefore, PCA simply consists in computing the eigenvectors and eigenvalues of the matrix C: $C = U\Lambda U^T$ where $\Lambda = diag(\lambda_1, \cdots, \lambda_n)$ is the diagonal matrix of the ordered eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$, and U is a $p \times p$ orthogonal matrix containing the eigenvectors. Dimensionality reduction is then obtained by the following operator $h_{PCA} : \mathbf{x}_i \rightarrow (y_1(i), \cdots, y_q(i))$ where $y_k(i)$ is the i^{th} coordinate of eigenvector $\mathbf{y}_k$. In the rest of this paper, we will denote h_{PCA}^{Cov} and h_{PCA}^{Cor}, dimensionality reduction performed with PCA of the covariance or the correlation matrix.

3.2 Laplacian Eigenmaps

Given a neighborhood graph G associated to the vectors of X, one considers its adjacency matrix W where weights W_{ij} are given by a Gaussian kernel $W_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = e\left(-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{\sigma^2}\right)$. Let D denote the diagonal matrix with elements $D_{ii} = \sum_j W_{ij}$ and Δ denote the un-normalized Laplacian defined by $\Delta = D - W$. Laplacian Eigenmaps dimensionality reduction consists in searching for a new representation $\{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n\}$ with $\mathbf{y}_i \in \mathbb{R}^n$, obtained by minimizing $\frac{1}{2}\sum_{ij} \left\|\mathbf{y}_i - \mathbf{y}_j\right\|_2 W_{ij} = Tr(\mathbf{Y}^T \Delta \mathbf{Y})$with $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n]$. This cost function encourages nearby sample vectors to be mapped to nearby outputs. This is achieved by finding the eigenvectors $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n$ of matrix Δ. Dimensionality reduction is obtained by considering the q lowest eigenvectors (the first eigenvector being discarded) with $q \ll p$ and is defined by the following operator $h_{LE} : \mathbf{x}_i \rightarrow (y_2(i), \cdots, y_q(i))$ where $y_k(i)$ is the i^{th} coordinate of eigenvector $\mathbf{y}_k$.

4. CLASSIFICATION METHOD

From all existing classification schemes, a Support Vector Machine (SVM)-based technique has been selected due to high classification rates obtained in previous works,[18] and to their high generalization abilities.

The SVMs were developed by Vapnik et al.[19] 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 $\mathcal{S} = \{(x_i, y_i)\}_{i=\{1,\ldots,m\}}, x_i \in \mathbb{R}^n$, $y_i \in \{-1, +1\}$, SVM maps the input vector x into a high-dimensional feature space $\mathbf{H}$ through some non linear mapping functions $\phi : \mathbb{R}^n \rightarrow \mathbf{H}$, 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:

$$\mathcal{V}(w, b, \xi) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{m} \xi_i\right) \tag{6}$$

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[19] is to maximize the following optimization problem:

$$\mathcal{W}(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{7}$$

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:

$$h(x) = \text{sign}(f(x)) \text{ with } f(x) = \sum_{i \in SV} \alpha_i^* y_i K(x_i, x) + b^* \tag{8}$$

where the threshold b^* is computed via the unbounded support vectors[19] (*i.e.*, $0 < \alpha_i^* < C$). An efficient algorithm SMO (Sequential Minimal Optimization)[20] and many refinements[21,22] 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.[23]

4.1 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

$$K_\beta(s_i, s_j) = exp(-\sum_{k=1}^{n} \beta_k (s_i^k - s_j^k)^2 / r^2) \tag{9}$$

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 values 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.[24] 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.[23]

In this paper, the common One-Versus-One (OO) decomposition scheme is used to create 10 binary classifiers. Let $t_{i,j}, \forall i \in [1,5], j \in [2,5]$ be a binary problem with $t_{i,j} \in \{+1, -1\}$. The number 5 represents the final quality classes according to the ones recommended by the UIT. Let $h_i(\cdot)$ be the SVM decision function obtained by training it on the i^{th} binary problem. Table 4.1 gives binary problems transformation used in the OO scheme.

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.[23] The majority vote criterion is the usual way to do this. Let $V_j(x) = \sum_{i=1}^{n_b} LO_1(h_i(x), t_{i,j})$ be the number of votes for

class	$t_{5,4}$	$t_{5,3}$	$t_{5,2}$	$t_{5,1}$	$t_{4,3}$	$t_{4,2}$	$t_{4,1}$	$t_{3,2}$	$t_{3,1}$	$t_{2,1}$
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

Table 1. Binary problems transformation used in a One-Versus-One combination scheme.

the class j, where n_b is the number of binary decision function in a specific combination scheme, and LO_1 is a loss function defined as follows:

$$LO_1(y_1, y_2) = \begin{cases} 0 & \text{if } y_1 = y_2 \\ 1 & \text{else} \end{cases} \tag{10}$$

where m_{y_2} corresponds to the number of images labelled to as class y_2 in the reference dataset. The multiclass decision function D using majority vote is:

$$D(x) = \underset{1 \leq j \leq n_c}{\arg\max} \left(V_j(x) \right) \tag{11}$$

(when conflicts exist, the SVM output is used to break it).

5. EXPERIMENTAL SETUP AND RESULTS

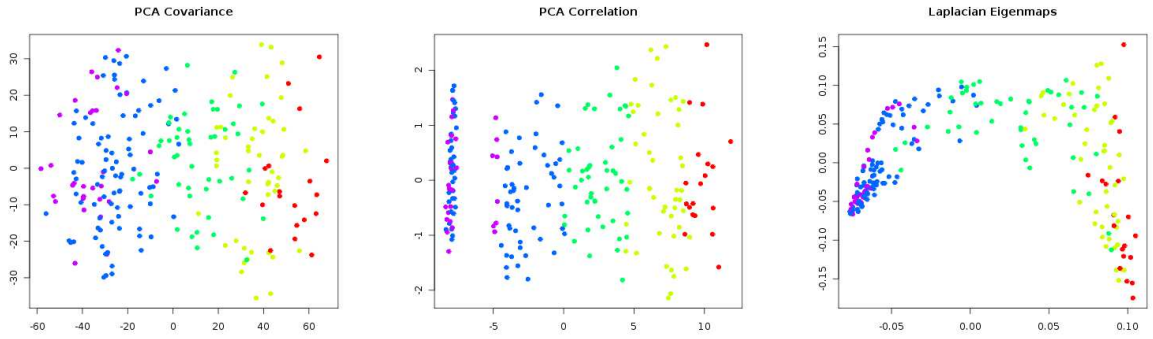
SVM classification results are obtained with JPEG2000 compressed versions of 12 images of the LIVE image database, for which 1) all the visual features have been computed and 2) the class selection of each human observer is available. The JPEG2000 compressed versions of the 13 remaining images constitute the test set, for which the same kind of information are available.

Figure 1 presents the 2D projection of the obtained results using the trial methods to reduce the initial dimension of the data. As we can observe, h_{PCA}^{Cor} seems to perform best than the two others tested dimension reduction techniques. Actually, the separation of the five quality classes is linearly separable in the first case, while using the two other projections classes are not linearly separable.

Considering this point of vue, the results obtained from h_{PCA}^{Cor} is the best candidate to be computed with the SVMs; it seems to be the best representation to be linearly separable. In order to confirm or not this assumption, the SVMs will be applied with the obtained results from the three dimension reduction methods: h_{PCA}^{Cov}, h_{PCA}^{Cor} and h_{LE}.

Figure 2 presents the Mean Opinion Score (MOS) obtained from a set of human observers for each tested image versus its associated compression rate (bpp)– The MOS is not taken into account in the dimension reduction process. On the upper left corner, one notes a vertical set of points. Those points correpond to all the original images (with a compression ratio equal to 0) of the trial test for which the quality is judged as very good or as excellent. When observing the subfigure 1(a), the same remarks can be formulated. Thanks to the reduction of the dimension of the original features vector, one is able to reach a linear separation between classes, except for two classes : "Excellent" and "very good". This is probably due to the original images than can be interprated as noise.

Table 5 presents the recognition rate errors with respect to number of used features issued from the dimension reduction step. From this table one can observe that the recognition error rate is quite always minimized when the h_{PCA}^{Cor} is used to reduce the dimension of the data prior to apply SVMs, whatever the number of new vectors used. On note that the best error rate is given when one only uses one feature that yields to reach an error rate of 14.1% *pm* 7.8%. This means that the initial data are highly correlated. When other features are added to perform the SVM, the recognition rate error increases. This can be interpreted as the following: when one adds new features, in fact one adds noise, and thus the whole performance decreases.



(a) h_{PCA}^{Cov} (b) h_{PCA}^{Cor} (c) h_{LE}

Figure 1. Obtained results from the three dimension reduction methods applied to the initial data. The colors are associated to the quality classes as follows: purple–5 (excellent), blue–4 (very good), green–3 (quite good), yellow–2 (bad) and red–1 (very bad)

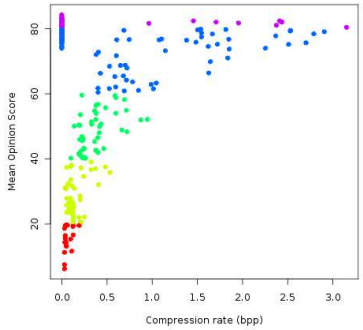


Figure 2. MOS vs. bpp. The colors are associated to the quality classes as follows: purple–5 (excellent), blue–4 (very good), green–3 (quite good), yellow–2 (bad) and red–1 (very bad)

	Recognition rate error with respect to the number of used features				
	1	2	3	4	5
h_{PCA}^{Cor}	0.141 ± 0.078	0.185 ± 0.064	0.181 ± 0.0984	0.207 ± 0.0691	0.191 ± 0.102
h_{PCA}^{Cov}	0.295 ± 0.0895	0.299 ± 0.113	0.308 ± 0.1020	0.291 ± 0.094	0.283 ± 0.166
h_{LE}	0.298 ± 0.111	0.317 ±0.114	–	–	–
	6	7	8	9	10
h_{PCA}^{Cor}	0.229 ± 0.062	0.229 ± 0.106	0.251 ± 0.076	0.219 ± 0.111	0.264 ± 0.133
h_{PCA}^{Cov}	0.317 ± 0.0831	0.321 ± 0.087	0.211 ± 0.09	0.238 ± 0.056	0.277 ± 0.040
h_{LE}	–	–	–	–	–

Table 2. Recognition rate errors with respect to number of used features issued from the dimension reduction step.

Nevertheless, when one compares the obtained results with the one obtained when the training phase is performed on the initial data (*i.e.*, without applying the dimensionality reduction process as a prior step), the recognition rate is equal to 14.2% ± 7.1%. This means that, for the original attributes used to create the features vector characterizing a compressed image, the reduction of the dimension of this vector is not a significant step to improve the quality of the classification process. Nevertheless, using only one attribute, the data, and thus the results obtained at the end of the classification process become more easily understandable by a human being.

More generally, when a dimensionnality reduction process is expected as a prior step to a classification approach, one has to think about an existing correlation between the features of the initial vector.

6. CONCLUSION

In this paper a new approach to design a quality metric is proposed. This approach is based on a classification process such as the human being is supposed to proceed to judge the quality of an object. To apply the classification process, a vector of features has been generated. The selected features are chosen from full-reference image SVH-based features and full-reference image features, both for them a reference image is needed. In addition, no-reference images SVH-based features and no-reference images features (both for them no reference image is needed) are included in the initial features vector.

In order to reduce the number of features within the initial vector, a manifold learning process is applied to reduce the dimension of the vector that will be used during the classification process. Three techniques are tested: principal component analysis performed 1) on the symmetric covariance matrix and 2) on the symmetric correlation matrix, and 3) Laplacian Eigenmaps.

The obtained results shown that the PCA on the correlation matrix gives better results and yields to reach a recognition rate error close to 14%. Analysis tends to prove that the original attributes of the features vectors are highly correlated. In future works, new attributes will be designed, and original images will be removed from the trial images set.

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. C. Charrier, K. Knoblauch, and H. Cherifi, "A color image quality assessment using a reduced-reference image machine learning expert," in *SPIE, Image Quality and System Performance V*, **6808**, (San-Jose, California), Jan. 2008.
3. Z. Wang and A. C. Bovik, "A universal quality index," *IEEE Signal Processing Letters* **9**(3), pp. 81–84, 2002.
4. S. Daly, "A visual model for optimizing the design of image processing algorithm," in *ICIP*, **2**, pp. 16–20, 1994.
5. J. Krauskopf, D. R. Williams, and D. W. heeley, "Cardinal directions of color space," *Vision Research* **22**, pp. 1123–1131, 1982.
6. A. B. Watson, "The cortex transform: Rapid computation of simulated neural images," *Computer Vis. Graphics and image proces.* **39**, pp. 311–327, 1987.
7. 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, 1993.
8. 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.
9. E. Peli, "Contrast in complex images," *Journal of the Optical Society of America* **7**, pp. 2032–2040, Oct. 1990.
10. 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.
11. 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.
12. T. Vlachos, "Detection of blocking artifacts in compressed video," *Electronics Letters* **36**(13), pp. 1106–1108, 2000.
13. 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.
14. A. Cumani, "Edge detection in multispectral images," *Graphical Models and Image Processing* **53**, pp. 40–51, Jan. 1991.
15. 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.

16. L. K. Saul, K. O. Weinberger, J. Ham, F. Sha, and D. D. Lee, "Spectral methods for dimensionnality reduction," in *Semi-supervised Learning*, pp. 279–294, MIT Press, 2006.
17. M. Belkin and P. Niyogi, "Laplacien eigenmaps for dimensionality reduction and data representation," *Neural Computing* **15**(6), pp. 1373–1396, 2003.
18. 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.
19. V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
20. J. Platt, *Fast Training of Support Vector Machines using Sequential Minimal Optimization, Advances in Kernel Methods-Support Vector Learning*, MIT Press, 1999.
21. R. Collobert and S. Bengio, "SVMTorch: Support vector machines for large-scale regression problems," *Journal of Machine Learning Research* **1**, pp. 143–160, 2001.
22. 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.
23. 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.
24. R. Kohavi and G. H. John, "Wrappers for feature subset selection," *JAIR* **97**(1-2), pp. 273–324, 1997.