17

# Image Similarity based on a Distributional "Metric" for Multivariate Data

Christos Theoharatos, Nikolaos A. Laskaris, George Economou
& Spiros Fotopoulos
*Electronics laboratory, Dept. of Physics, University of Patras*
*Greece*

## 1. Introduction

The problem of image similarity has become a challenging task in the field of computer vision through the last two decades. The assessment of (dis)similarity between color (or multichannel, in general) images or parts of images has been studied on several image processing application domains such as image indexing and retrieval, classification and unsupervised segmentation (Rubner et al., 2001). The basic operations that need to be carried out in order to estimate the similarity between two color images are three-fold (Stricker & Orengo, 1995): first, choose an appropriate color space for image representation; then, extract a signature for each image (using, commonly, low-level features) to construct a theoretically valid distribution; finally, establish pairwise comparisons based on these signatures. Each signature constitutes the content description of a corresponding image. It is summarized based on pixel attributes and provides a representation of the image in a multidimensional feature space. There, a proper (dis)similarity measure is defined in order to act as a general rule for comparing any given pair of images.

In these directions, several (dis)similarity measures have been developed and used as empirical estimates of the distribution of image features, confirming that distribution-based measures exhibit excellent performance in all areas (Rubner et al., 2001). In the context of visual image similarity, we make use of a nonparametric test from the field of multivariate statistics that deals with the "Multivariate Two-Sample Problem", originally presented by Friedman and Rafsky (1979). The specific test is a multivariate extension of the classical Wald-Wolfowitz test (WW-test) and compares two different samples of vectorial observations (i.e. two sets of points in $\mathbf{R}^p$) by checking whether they form different branches in the overall minimal spanning tree (MST) (Zahn, 1971). It provides an aggregate gauge of the match between color images, taking into consideration all the selected characteristics, while alleviating correspondence issues. The output of this test can be expressed as the probability that the two point-samples are coming from the same distribution. We have proven that this is a powerful measure for image similarity, relying on the statistical comparison of content representations in a properly defined feature space (Theoharatos et al., 2005).

Here, the above distributional-'metric' is introduced in conjunction with a prototyping method that dramatically speeds up the execution of the involved computations and results



Source: Vision Systems: Segmentation and Pattern Recognition, ISBN 987-3-902613-05-9,
Edited by: Goro Obinata and Ashish Dutta, pp.546, I-Tech, Vienna, Austria, June 2007

in an efficient overall methodology (e.g. so as to be used in highly demanding applications such as image retrieval tasks). The current proposal incorporates the use of a computational intelligent module for content representation based on self-organizing neural networks (SONNs), the Neural-Gas algorithm (Martinez et al., 1993), which is responsible for generating a parsimonious description of the color distribution of each image. The multivariate distributions representing the individual images are then compared via the standard WW-test, providing enhanced performance when evaluated via a query-by-example image retrieval scheme (Theoharatos et al., 2006a).
Finally, we are discussing the applicability of the same distributional distance in order to compare images following a standard JPEG-format (Wallace, 1991) and with the scope to emphasize texture characteristics during the visual search. Color and texture features are directly extracted from the DCT-compressed domain, in the form of an ensemble of feature vectors that are the inputs to a standard WW-test. The emerging indexing scheme is found to be robust, providing invariant similarity results when image rotation is considered (Theoharatos et al., 2006b).

## 2. Background and related work

Research on image similarity has expanded lately, mainly due to the increased interest of content-based image retrieval (CBIR), which constitutes a highly challenging research area with the emerging techniques sharing many advantages (Smeulders et al., 2000). Even though the focus of interest for image similarity and retrieval has recently shifted towards the identification of high-level semantics from the content of the images (Eakins, 2002), not much success has been achieved so far. This is mainly due to the great difficulties in the derivation of semantically meaningful information at a general level (Sheikholeslami et al., 2002). As a consequence, nowadays methods are still constrained to use low-level visual features such as color, shape and texture to represent the image content.
Considerable investigation has been carried out on the basis of color content (Schettini et al., 2001). Color information has been recognized as the most important indicator of the general 'mood' of an image and is considered to capture, to a certain extent, image semantics. In the existing literature, researchers have experimented with different color spaces such as RGB, CIE-Lab, etc. (Castelli & Bergman, 2002), various color descriptors such as color histogram (Swain, & Ballard, 1991), color moments (Stricker & Orengo, 1995) and chromaticity moments (Paschos et al., 2003), and also miscellaneous similarity measures such as histogram intersection (Swain & Ballard, 1991), quadratic form distance functions and statistical indices (Rubner et al., 2001). The most popular representation of color information is the global histogram, which statistically denotes the joint probability of intensities of the three-color channels, thus describing the global color distribution in an image. In general, the color histogram provides useful clues for the subsequent expression of similarity between images, due to its robustness to background complications and object distortion. Moreover, it possesses translation, scale and rotationally invariant characteristics. A profound number of (dis)similarity measures have been proposed for computing the distance between histograms from two different images. In their work, Rubner et al. (2001) distinguished these measures generally into four categories: heuristic histogram distances, nonparametric test statistics, information-theory divergences and ground distance ones. In the context of image indexing and retrieval, the different variants of the color histogram-related methodology have provided satisfactory results, especially in practical situations in

which the feature extraction step needs to be accomplished as simply and promptly as possible. Soon it became popular, since it was very simple to implement and exhibits fast retrieval response time, making it a good candidate for real-time applications. However, the performance of this technique was not found to be high enough, mainly due to the necessary trade-off during the binning procedure. An adequate compromise could be achieved via the use of an adaptive binning procedure (e.g. Leow & Li, 2004), in which the histogram bins would adapt to the actual distribution of colors in images. Apart from the facts that bin-adaptation can be a computational demanding task and, in general, is still considered an open issue in the field of image processing, existing systems adopt fixed-binning histograms since most dissimilarity measures are unable to cope with histograms build over different sets of bins (Rubner et al., 2001).

In order to overcome the above limitation, an attempt was made recently by Rubner et al., (2000) to combine the benefits from the use of a distribution distance with a flexible description of color-content that adapts its resolution to individual images. The innovative work mentioned above introduced the Earth Mover's Distance (EMD), a computational demanding task based on the solution of the well-known transportation problem. In summary, a representation scheme suitable for color distributions and based on Vector Quantization (VQ) preceded the computation of EMD between pairs of distributions. In this scheme, after the complicated k-d trees algorithmic procedure for cluster analysis, each distribution was represented by means of a number of cluster-centroids and the corresponding proportions of image pixels with colors within the identified groups. The EMD-related technique was shown to be more robust than histogram-matching techniques, since it could operate on variable-length representations of the distributions that were avoiding quantization problems related with the binning procedure. In short, higher performance was achieved at the expense of computational efficiency. However, the integrated representation design is not related directly to the reliability of color distribution. Although the efficiency of k-d tree algorithm is generally recognized, their effectiveness for clustering data of complex distributions or data with high correlations among variables is questionable. Moreover, there is lack of supporting evidence in the field of statistics that EMD is indeed an appropriate measure for comparing multivariate distributions, apart from the theoretical benefit that correlates with perception when applied in the CIE-Lab space.

In the context of textural features, these are also represented using histogram-based methodologies. Indexing, similarity and retrieval of compressed images have recently become a very active research area, since the great amount of digital images provided on the WEB are stored in JPEG format (Wallace, 1991). In particular, the JPEG compression standard applies DCT transform in order to achieve a large amount of compression, significantly reducing the image size. Such compression is suitable for Internet-based applications, reducing the storage space while increasing the downloading speed. Thus, measuring image similarity directly in the compressed domain becomes more and more beneficial, compared to the pixel-based one. To bridge the gap between compressed- and pixel-space, where the majority of image processing algorithms are developed, recent research is now starting apace to develop content feature extraction algorithms working directly in the compressed domain (e.g. Zhong & Jain, 2000; Ngo et al., 2001; Jiang et al., 2004). Since the inverse DCT (IDCT) is an embedded part of the JPEG decoder and the DCT itself is one of the best filters for feature extraction working directly on the DCT domain, it has proven to be a well-promising area for image similarity in the compressed domain. DCT

has, to a certain extent, unique scale invariance and zooming characteristics, which can provide insight into objects and texture identification (Ngo et al., 2001). In addition, it exhibits a set of good properties such as energy compaction and image data decorrelation and, therefore, is naturally considered to be a potential domain in mining visual information. Thus, direct feature extraction from DCT domain can provide better solutions in characterizing the image content, apart from its advantage of eliminating any necessity of decomposing an image and detecting its features in the pixel domain (Jiang et al., 2004).
The rest of the presentation is organized as follows. Section 3 provides an overview of the proposed distributional-'metric' for comparing multivariate data, including the graph-theoretic framework of MST and the multivariate WW-test. Color image similarity is presented in Section 4, using the Neural-Gas network for expressing the image content-signature. In Section 5, visual similarity in the compressed domain is analysed by extracting color and texture attributes directly from the DCT-space. Finally, conclusions are drawn in Section 6, along with an outline of our future research objectives in Section 7. Throughout our study, image similarity is evaluated via a query-by-example image retrieval scheme

## 3. The Distributional 'Metric' for Comparing Multivariate Data

A *nonparametric* test dealing with the *"Multivariate Two-Sample Problem"* (Friedman & Rafsky, 1979) is proposed for measuring image similarity in a reliable and more sophisticated way. The specific test is a multivariate extension of the classical statistical test of Wald and Wolfowitz and compares two different samples of vectorial observations (i.e. two sets of points in $\mathbf{R}^p$). The output of the test can be expressed as the probability that two point-samples are coming from the same distribution. Its great advantage is that no a-priori knowledge about the distribution of points in the two samples is a prerequisite (Theoharatos et al., 2005). This model-free assumption stems from the graph-theoretic origin of the WW-test, which is actually based on the concept of the MST-graph (Zahn, 1971). For this reason, a compact description of MST is preceded first.

### 3.1 MST-Graph Representation

Given the establishment of a systematic procedure for extracting low-level characteristics from a color (or multivariate, in general) image that are individually represented as vectors in a predetermined space, one can rely on graph theory to provide a collective perspective that captures the essence of the visual content of the image under study. Graph theory, by putting emphasis on the structural relationships between the extracted characteristics, provides robust descriptions against noise degradation widely and randomly spread over the field and simple transformations like image scaling. Specifically the MST-graph appears as an extremely useful condensation of the bulk of information conveyed within the ensemble of image characteristics. In addition, the MST provides a compact description of a point set. It contains the '*nearest neighbor*' information about each point and the '*shortest linkage*' information about subsets of points (Laskaris & Ioannides, 2001). In his study, Zahn (1971) established another advantage of MST, the *determinacy*, meaning that the results from the application of a method working with MST-graph do not depend on random choices or the order in which points are scrutinized, but are affected solely by the point set provided as input. Overall, the MST structure is unchanged under transformations like translation,

rotation and non-linear ones, preserving the ordering of edge lengths (Theoharatos et al., 2005).

Graph theory sketches the MST structure with the following definitions (Zahn, 1971). A *graph* $G(V,E)$ is a mathematical structure for representing pairwise relationships among data. It consists of a set of points called *nodes* $V=\{V_i\}_{i=1:N}$ (or *vertices*) and a set of links $E=\{E_{ij}\}_{i\neq j}$ between nodes called *edges* (or *lines*). An edge links two nodes defining it, when it is incident on both of them. The *degree* $d_i$ of a node is the number of edges incident to it. When a weight $e_{ij}$ is assigned to each link, a weighted-graph is formed and in the particular case that $e_{ij}=e_{ji}$ this graph is called *undirected weighted graph*. A *connected graph* has a path between any two distinct nodes and a *tree* is a connected graph with no cycles. A *subgraph* of a given graph is a graph with all of its nodes and edges in the given graph. A *spanning tree T* of a (connected) weighted graph $G(V,E)$ is a connected subgraph of $G(V,E)$ such that: (i) it contains every node of $G(V,E)$, and (ii) it does not contain any cycle. The *MST* is a spanning tree containing exactly $N-1$ edges, for which the sum of edge weights is minimum.

Suppose now that $N-$ pixels are randomly selected from an image and the corresponding RGB vectors are represented as an ensemble of points in the feature space. The specific points are used as the nodes of the original (fully-connected) graph, while the interpoint Euclidean distances as the weights of the corresponding edges. Using a standard algorithm (Prim, 1957), the MST is evolved from the original graph, offering a parsimonious description of the low-level information in an image. Given a second image, the color content of which is to be compared with the content of the first one, we can proceed with the selection of pixels as previously and transform the comparison between feature-contents into a comparison between the corresponding MST-graphs (Theoharatos et al., 2005). To perform such a comparison, a well-defined statistical test is available in the literature of multivariate statistics.

### 3.2 The Multivariate WW-Test

Consider samples of size *m* and *n* respectively from distributions $F_x$ and $F_y$, both defined in $\mathbf{R}^{\mathbf{p}}$. The hypothesis $H_0$ to be tested is whether they are coming from the same distribution, thus $F_x=F_y$. We are interested in the rejection of the original hypothesis, which is the alternative hypothesis $F_x\neq F_y$. In the univariate case ($p=1$), the WW-test begins by sorting the $N=m+n$ univariate observations in ascending order. Friedman and Rafsky (1979) proposed the use of MST as a multivariate generalization of the univariate sorted list, introducing in this way a methodology to define the two-sample test statistics based on the MST in analogy with those based on the sorted list.

In the multivariate case, the hypothesis $H_0$ to be tested is whether two multidimensional point samples $\{X_i\}_{i=1:m}$ and $\{Y_i\}_{i=1:n}$ are coming from the same multivariate distribution. In this general case, the WW-test can be summarized with the following steps (Friedman & Rafsky, 1979): (i) Consider samples of size *m* and *n* respectively from distributions $F_x$ and $F_y$, both defined in $\mathbf{R}^{\mathbf{p}}$, (ii) Construct the overall MST without encountering the sample identity of each point (iii) Delete the edges for which the defining nodes originate from different samples. Then, based on the sample identities of the points the test statistic $R$ is

computed, defining the total number of *runs*, while a *run* is defined as a consecutive sequence of identical sample identities. $R$ can be also defined as the number of disjoint subtrees that finally result. In order to illustrate the WW-test for ease in understanding, two randomly selected samples of size $m=5$ and $n=8$ are used in the 2-D of Fig. 1. After deleting those edges coming from different distribution, the number of disjoint subtrees is calculated and found equal to $R=5$. It must be pointed out here that, the MST possesses two significant properties which make it appropriate for application to the multivariate two-sample problem (a) it connects all the $N-$nodes with $N-1$ edges, which comes from the fact that the MST is a spanning tree and (b) the node pairs defining the edges represent points that tend to be close together, which stems from the requirement that the sum of the edge weights is minimum.

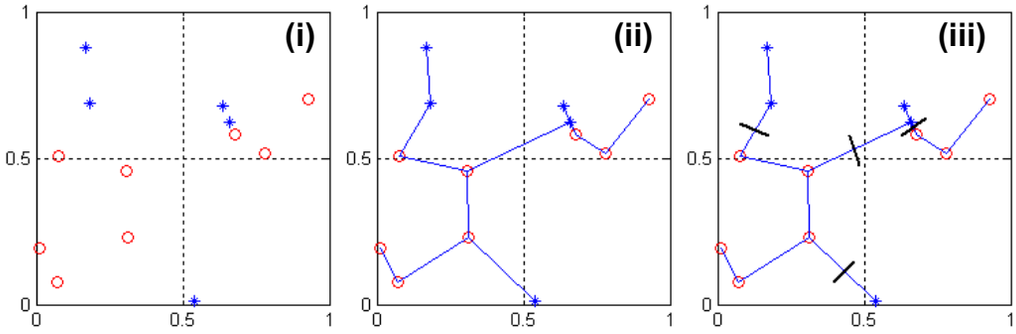


Fig. 1. Visual configuration of the multivariate WW-test algorithmic procedure for two randomly sampled distributions: (i) consider the two sample-distributions, (ii) construct the overall MST and (iii) delete the edges of the nodes originating from different distributions.

The null distribution of the test statistic is derived, based on the combinatorial analysis given by Friedman and Rafsky (1979). Let $N=m+n$, $C$ be the number of edge pairs of MST sharing a common node, and $d_i$ be the degree of the $i^{th}$ node. Then, $C=\frac{1}{2}\sum_{i=1}^{N} d_i(d_i-1)$. Number the $N-1$ edges of the MST arbitrarily and define $Z_i$, $1\le i \le N\text{-}1$, as:

$$Z_i=\begin{cases}1 & \text{if the } i\text{-th edge links nodes from different samples}\\ 0 & \text{otherwise}\end{cases}$$

Then, $R=\sum_{i=1}^{N-1} Z_i+1$. Under $H_0$, the mean and variance of $R$ can be calculated using a systematic analysis as follows:

$$E[R]=\frac{2mn}{N}+1,\ Var[R\,|\,C]=\frac{2mn}{N(N-1)}\times\left\{\frac{2mn-N}{N}+\frac{C-N+2}{(N-2)(N-3)}[N(N-1)-4mn+2]\right\}$$

It has been shown that the quantity:

$$W=\frac{R-E[R]}{\sqrt{Var[R]}} \tag{2}$$

approaches (asymptotically) the standard normal distribution while $E[R]$ and $Var[R]$ are given in closed form based on the size of the two samples (Friedman & Rafsky, 1979). The

importance of the previous is that by using simple formulae, the *significance level* (and *p-value)* for the acceptance of the hypothesis $H_0$ can be readily estimated.

## 4. Comparing Color Distributions via a self-organizing algorithm

Regarding the plethora of methods and feature extraction techniques, image indexing and similarity is associated with different levels of image understanding. Provided that a number of feature vectors are given, the resulted feature space may not be uniformly occupied. Sheikholeslami et al. (2002) studied the way clustering individuates the sparse and dense pixel-areas in the image, revealing the underlying distribution of the feature space. In addition, a vector quantization scheme realizes a concise representation of the input data regardless of the actual meaning and significance of the clusters (Gdalyahu et al., 2001). The resulting codebook vector can be considered as a compact description of the data distribution (e.g. the color information of database images), providing effective and alternative ways to portray image content.

To avoid missing the generality of the approach and at the same time propose its efficient/intelligent version, the use of two sequential modules is illustrated in the specific domain of color image information management, which considers the RGB-vectors corresponding to individual pixels (i.e. points in $\mathbf{R}^3$). In a nutshell, using Neural-Gas based prototyping a data-summary will be produced, which constitutes a meaningful sampling from the underlying color distribution of each image. With the subsequent application of WW-test to compare samples of color prototypes, robust and economical comparisons regarding color content will be achieved (Theoharatos et al., 2006a).

### 4.1 Representation of Color Distributions via Self-Organizing Networks

Summarizing data distributions via prototypes has roots in the theory of VQ, which is a powerful strategy for data compression and can be accomplished via different techniques (Gray, 1984). Briefly, a vector quantizer encodes a data manifold $V \subseteq \mathbf{R}^p$ utilizing only a finite set of reference or "codebook" vectors $O_j \in \mathbf{R}^p$, $j=1,\ldots,k$, which are also called cluster centers. Each data vector $X \in V$ is described by the best-matching reference vector $O_{j(X)}$ for which the distortion error $d\left(X, O_{j(X)}\right)$, usually measured via the squared Euclidean distance, is minimal. The main core of the procedure depends on the division of the original manifold V into a number of subregions $V_i$ called Voronoi polygons or Voronoi polyedra, out of which each data vector $X$ is described by the corresponding reference vector $O_j$. The efficient application of VQ mainly depends on the codebook design, i.e. the proper selection of reference vectors. For this critical step, the use of traditional clustering algorithms like the k-means had been originally proposed. However, it was experimentally verified later that these algorithms often lead to a suboptimal choice of reference vectors $O_j$ in the case of nontrivial data distributions, as well as in the case of an inappropriate selection for the number of reference vectors. Such a suboptimal solution can have a significant impact on the subsequent encoding of the data and even result to highly distorted representations.

The tremendous development of neural theory of unsupervised learning and the related algorithms of Self-Organizing Neural Networks (SONNs) revitalized the field of VQ. The ability to efficiently deduce prototypes from the data, common in many SONNs like the Kohonen's feature map (Kohonen, 1997) and the Neural-Gas (Martinez et al., 1993), could be

exploited in the reliable codebook design. For a thorough treatment of SONNs and their applications related with VQ, the interested reader can refer to the seminal study of Martinez and Schulten (1994).

Stochastic presentation of the input data, competition among the neural nodes (to which weight vectors $A_j \in \mathbf{R}^p$ have been assigned) and a 'soft max' adaptation rule are the common characteristics of these networks that guarantee the fast convergence to a set of weight vectors (i.e. prototypes), which can serve as a high-fidelity codebook. The resulting codebook vectors are allocated according to the probability distribution of data vectors over the manifold V, and in such a way that the average distortion error is minimized. The main difference between the SONN-algorithms compared to other traditional clustering methodologies is that not only the best-matching reference vector $O_{j(X)}$ is adjusted every time a data vector $X$ is presented, but also the reference vectors adjacent to it are updated accordingly. Among the SONNs, Kohonen's feature map is the most popular mainly due to the accompanying visualization scheme that enables the projection of the input data nonlinearly onto a lower dimensional lattice (Kohonen, 1997; Haykin, 1999). Inspired by the possibility that some high level organization in the brain may be created during learning through self-organization, Kohonen (1997) presented a self-organizing learning algorithm that presumably produces feature maps similar to those occurring in the human brain. In this way, the self-organizing map (SOM) forms a nonlinear regression of the ordered set of reference vectors into the input space. The reference vectors constitute a low-dimensional network that follows the original data distribution; for this reason, it is also referred to as '*self-organizing semantic map*'. However, to obtain efficient quantization results with Kohonen's feature map algorithm, the topology of the lattice has to match the topology of the data manifold V that is to be represented. Since the primary interest in our study lies in the precise quantization of the data and not in dimensionality reduction, we avoided the use of Kohonen's network. Instead, we resorted to Neural-Gas network, which had been proven to quickly converge to distortion-errors lower than the ones achieved using Kohonen's algorithm or other classical clustering algorithms (Martinez et al., 1993).

### 4.2 The Neural-Gas Algorithm for Vector Quantization

For the purposes of vector quantization, the Neural-Gas algorithm is presented in this step and utilized in the dual segregation algorithmic procedure for our efficient image similarity methodology. It is a neural network algorithmic procedure that sustains specific properties that make it appropriate as a feature extraction scheme: (1) it converges quickly to low distortion errors, (2) it reaches a distortion error much lower than the corresponding using the K-means clustering and other traditional techniques or the one resulting from the SOM-approach, and (3) it obeys a gradient descent on an energy surface, in contrast to the Kohonen's feature map network (Martinez et al., 1993).

In the Neural Gas network algorithm, a stochastic sequence of incoming data vectors $X(t)$, $t = 1,2,\ldots,t_{\max}$, which is governed by the distribution $P(X)$ over the manifold V, drives the adaptation step for adjusting the weights of the $k$ neurons $\{A_j\}_{j=1:k}$ (i.e. the reference vectors)

$$\Delta A_j = \varepsilon\, h_\lambda\left(f_j\left(X(t),\{A_i\}_{i=1:k}\right)\right)\left(X(t)-A_j\right),\ j=1,\ldots,k,\ \forall\, t=1,\ldots,t_{\max} \tag{1}$$

The function $h_\lambda(y)$ in the above equation has an exponential form $e^{-y/\lambda}$ and $f(X,\{A_i\})$ is an indicator function that determines the 'neighbourhood-ranking' of the reference vectors according to their distance from the input vector $X$. For both parameters $\varepsilon$ and $\lambda$, an exponentially decreasing schedule is followed, with $t_{\max}$ being the final number of adaptation steps that can be defined from the data based on simple convergence criteria (for analytical details refer to Martinez et al. (1993), see also Martinez & Schulten (1994)).
Martinez et al. (1993) mathematically proved that the asymptotic density distribution of the codebook vectors $P(A)$ was proportional to the data density $P(A) \propto P(X)^{\underline{d}/(\underline{d}+2)}$, where $\underline{d} \le d$ is the intrinsic dimension of the input data. This theoretical proposition along with the accompanying experimental evidence, showing that the Neural-Gas network is indeed capable of representing successfully data-manifolds with even intricate intrinsic geometries (Martinez & Schulten, 1994), motivated our conjecture that the designed codebook could serve as a faithful representation of the vectorial distribution in color-space. Therefore, it could be utilized in the subsequent comparisons regarding color content.
Fig. 2 illustrates the color-content representation through Neural-Gas prototypes, which clearly evidences that the distribution of the codebook vectors follows very closely the corresponding color distribution (Theoharatos et al., 2006a). In the depicted figure, three images are included (two of which "look similar" to each other), while their RGB distributions corresponding to all the pixels are shown in the left column along with their representations using the associated codebooks. In addition, the entire set of pixels comprising each RGB-image is presented as a black dot-swarm. It is clearly evident that the distribution of the codebook vectors follows very closely the corresponding color distribution. Therefore, each codebook can be thought of as a properly "down-sampled" version of the original RGB-distribution (Laskaris & Fotopoulos, 2004). Aiming at higher computational efficiency, an intermediate step of subsampling has been introduced between embedding an image in RGB-space and Neural-Gas based prototyping. Within this step, only a small portion (~5 %) of the pixels in the image is selected using uniform random sampling, and the associated vectors are used as input data to the neural network. The comparison of the codebooks designed with (right column) and without (left column) the subsampling step, shows only slight differences.

### 4.3 Comparing Color Signatures using the WW-test

In order to assess the similarity between two color images, the WW-test is utilized as follows. Provided the two color codebooks $\{A_i\}_{i=1:k}$ and $\{B_i\}_{i=1:k}$ extracted from a pair of images, the WW-procedure follows, with the extracted prototypes playing the role of the input point-samples $\{X_i\}_{i=1:m}$ and $\{Y_i\}_{i=1:n}$ respectively. $W$ is computed based on the involved codebook vectors and used as a similarity measure in a way that the more positive its value is, the more similar the color distributions in the two images are (Theoharatos et al., 2005). The $W-$quantity computed between pairs of images plays the role of a "distributional distance" and therefore inherits interesting invariant-characteristics. In the past, a few other statistical indices have been proposed, as well, as means of measuring similarity between color distributions. These distances, for instance the Kolmogorov-Smirnov distance (KS), the chi-square test ($\chi^2$-statistic), etc. (Rubner et al., 2001), measure

how unlikely it is that one distribution is drawn from the population represented by the other.

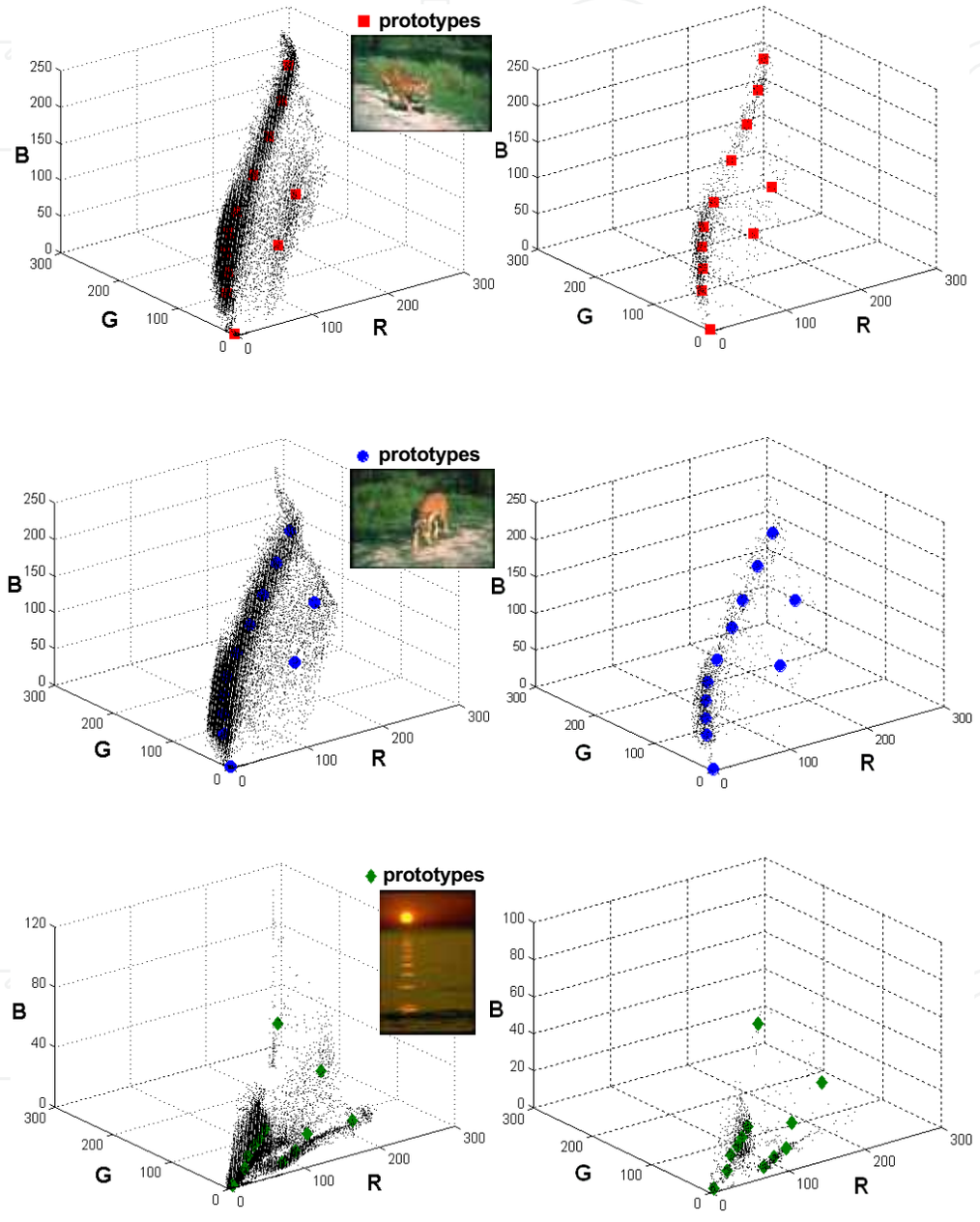


Fig. 2. Codebook color representation based on the Neural-Gas algorithm. For each image, a point-distribution is formed using the RGB-vectors corresponding to all (left panel) or a small portion (right panel) of the pixels, summarized through 12 prototypes.

Fig. 3 provides a demonstration of the test performance using the images and the codebooks presented in Fig. 2. In both panels different labels are associated with each of the two images to be compared. The $k = 12$ color prototypes extracted from each image are located in the RGB-space and the points indicating their position have been labeled according to the image they are coming from. By contrasting the two MSTs, it becomes evident that in the case of similar images (Fig. 3a) there are many edges having different labeled nodes as endpoints, while only a few in the case of dissimilar ones (Fig. 3b).

The unique benefit of WW-approach is that since it engages "distributional distance" acting on samples of image constituents, the emerging similarity measure possesses desirable invariant characteristics, such as rotation and translation invariance. Part of the flexibility is due to the statistical nature of the core procedure, the WW-test, and specifically its multivariate orientation. Theoharatos et al. (2005, 2006a, 2006b) have shown that not only different image characteristics can - in principle - be combined naturally in one type of query (i.e. color plus texture features), but also different types of queries can evolve independently and their results can be compared across types, as in the case of an image retrieval system. The latter is a direct consequence of the fact that the measured W-index relates directly to significance level and therefore can be used as an absolute measure to rank among the results of different types of query. Under these perspectives, the WW-test can be directly incorporated in retrieval processes from large image libraries, with the great advantage of being suitable for dealing with multivariate distributions.

### 4.4 Experimental evaluation via a query-by-example image retrieval scheme

In order to demonstrate and validate the effectiveness of the proposed methodology, a query-by-example image retrieval system was built. The image database included in the retrieval scheme contains a subset of $D = 1000$ color images from the Corel Collection. The utilized image-set was formed by pre-assigning the images into 20 distinct classes of $S = 50$ semantically similar images. A subset of $Q = 60$ query images from this heterogeneous set was also included in our retrieval system (three images per category). For the evaluation of the retrieval results, the *precision* (Pr) and *recall* (Re) indices (Castelli & Bergman, 2002) were adopted.

In the introduced methodology, the results are coming from different settings of the involved parameters. For different codebook sizes, the Precision was computed as a function of the number of RGB-vectors randomly sampled from each image and used as input data to the Neural-Gas network. The graphs obtained in this way showed that the Pr-index approached a relatively high value very soon (~1% of the pixels) and remained practically constant beyond the number of approximately 5% of the pixels (Theoharatos et al., in press a), which was the typical value chosen used throughout our evaluation study. In addition, by experimenting with the size of codebook vectors *k* that need to be drawn from each color image, extensive measurements have confirmed that after extracting $k = 25$ prototypes the Pr-index remained almost constant (Theoharatos et al., in press a). These results show that our method reaches the maximum performance for a moderate size of codebooks ($k \approx 25$) and therefore a more detailed representation of color distribution is unnecessary. This observation is very important for finding the best trade-off between effectiveness and efficiency when applying our algorithmic procedure.

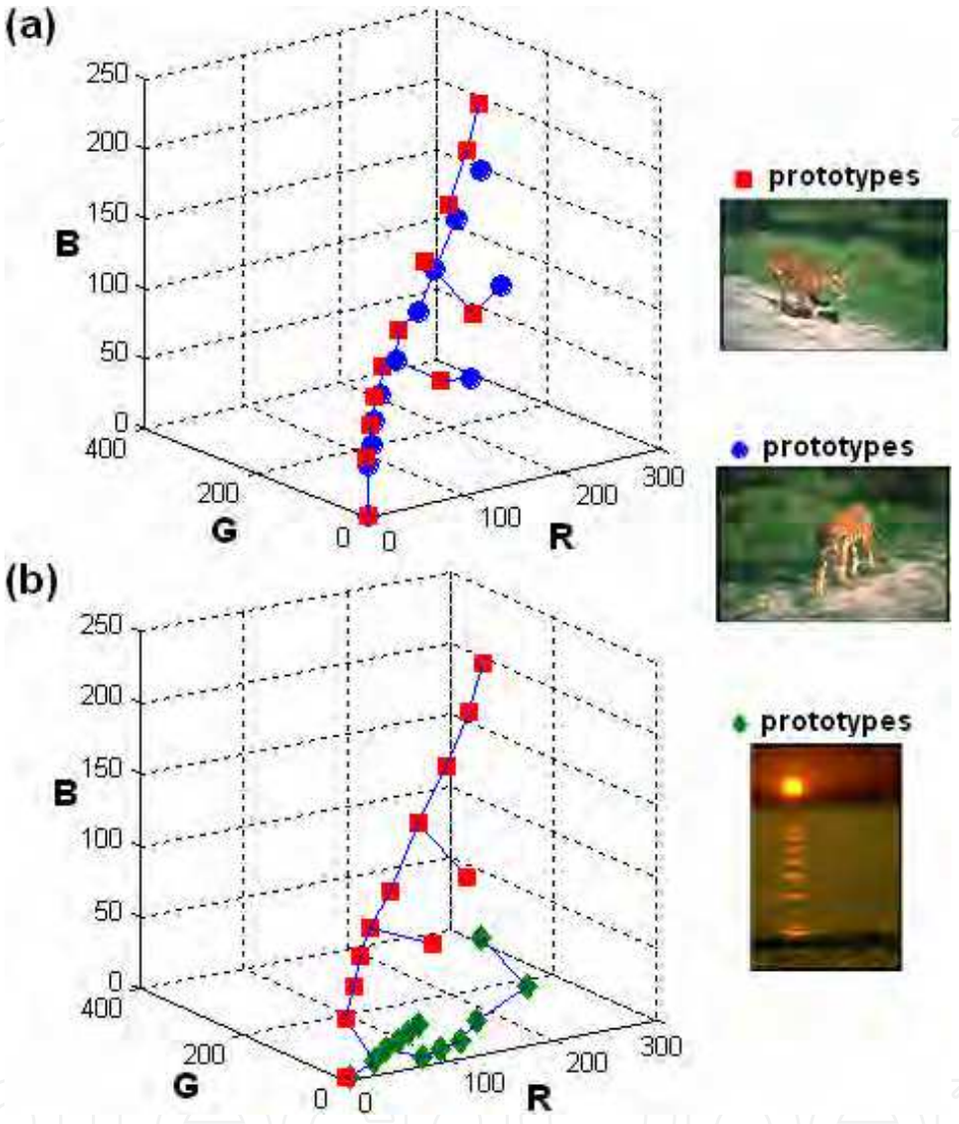


Fig. 3. WW-test for a pair of similar images (a) and dissimilar images (b), based on the k=12 color prototypes shown in Fig. 2. In the top panel, there are 19 edges having differently labeled nodes as endpoints and therefore splitting the overall MST into 20 subgraphs, thus R=20 (W=2.6523). On the contrary there are only 2 such edges in the bottom panel, thus R=3 (W=-4.8361).

For the full justification of our proposal, precision measurements regarding query-by-example search in the specific database are included. A plot of Pr-index as a function of the codebook size $k$ is presented in Fig. 4, for the $T = 10$ top retrieved images of the selected list. The performance of the Neural-Gas based WW-test is compared to the one using the EMD-

metric (Rubner et al., 2000) when applied upon the corresponding Neural-Gas based color-signatures. Although the depicted curves follow - as theoretically expected - a relatively similar trend, the WW-test outperforms the EMD-measure; for $k=10$ a satisfactory improvement of ~5% is apparent, while for $k\geq 15$ a significant increase in performance (~10%) is depicted. In addition, the general trend of the depicted curve is very interesting. The Pr-index reaches a plateau pretty soon and remains almost constant above the codebook size of $k=25$. This observation is of great importance regarding the involved computational load of our method and will be discussed in the last Section. The slightly decreasing trend that becomes apparent after the size of $k=40$ is a by-product of the fact that the number of extracted codebook vectors is increased without increasing the number of sampled vectors from the image. Therefore the Neural-Gas network attempts a detailed representation that is adapted to the idiosyncrasies of the random sample and tends to capture stochastic variations as delicate data-structure (a common-place problem in neural networks, usually referred to as over-training). By experimenting with greater sample sizes (~10% of the pixels), this trend is drastically reduced.

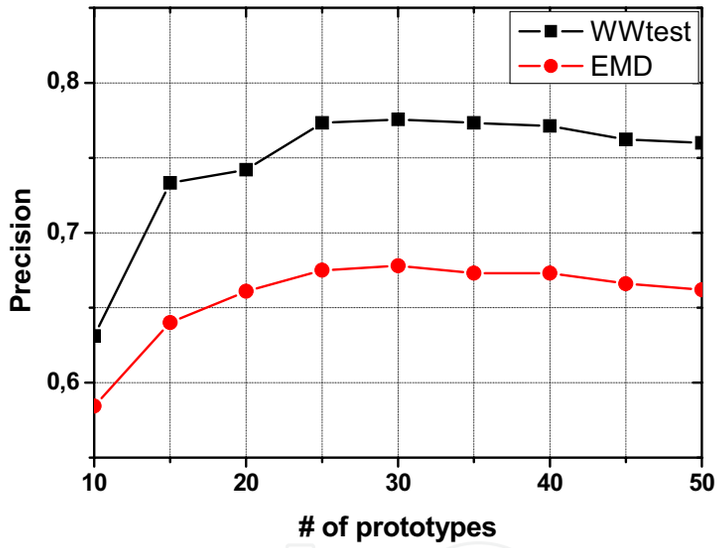


Fig. 4. Precision measurements of the WW-test and the EMD-related metric based on the same color codebooks, as a function of codebook size k.

The performance of the hybrid methodology as a method for accessing image databases was also evaluated following the standard procedure of constructing the Precision vs. Recall diagram. The Pr- and Re- indices were first evaluated for different sizes $T$ of the selected list (for $T=5:5:30$), and the computed values were used in the plot of Fig. 5. The corresponding diagrams for other dissimilarity measures (HI, $\chi^2$-test and JD using color histograms introduced in Rubner et al. (2000) and EMD applied on color signatures presented in Rubner et al. (2001)) have also been included in the same figure, enabling the direct comparison of the different approaches. It is clearly obvious that the WW-engine significantly outperforms all other methodologies.

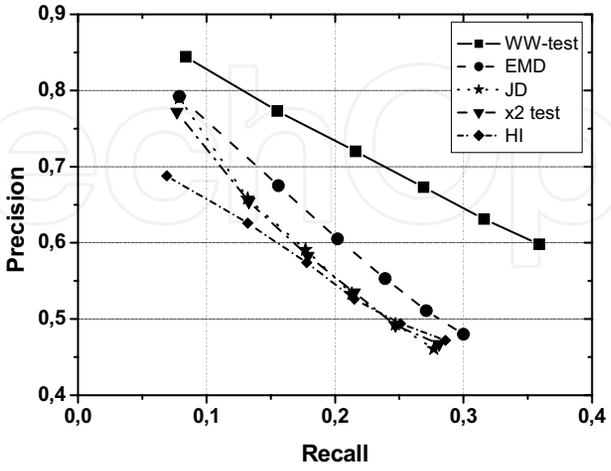


Fig. 5. Precision vs. Recall diagrams for the new hybrid approach, in comparison to other related techniques.

## 5. Visual similarity in the compressed domain

The flexible character of the WW-methodology relies on the multivariate flavour of the core statistical procedure. By altering the feature-extraction implementation, complementary ways to portray the image content appear without scaling effects or different cardinalities of the feature sets. An attempt is described here to adopt our methodology so as to work in compressed image domains that have recently gained high popularity (e.g. Zhong & Jain, 2000; Ngo et al., 2001; Jiang et al., 2004). This is expected not only to increase the efficiency of WW-based similarity scheme - by avoiding image decompression -, but also to constitute it suitable for novel applications like searching and retrieval in the World-Wide-Web - since the images of the Web are mostly included in a standard compressed format - (Jiang et al., 2004). Within this part we focus specifically on images from the standard JPEG compression scheme (Wallace, 1991). Competent ways to extract feature vectors directly from the zig-zag DCT-coefficients of the images are explored and their effectiveness is studied when exploited within the general framework of WW-methodology (Theoharatos et al., 2006b).

Color and texture features are utilized directly from the DCT-domain in the form of an ensemble of feature vectors represented in the YCrCb tri-chromatic model, in line with the JPEG standard (Wallace, 1991). In order to represent color information from each $N \times N$ pixel-block of a given image, all DC components are separately extracted and used as input vectors in the WW-engine to form a 3-D vector space. Texture features, on the other hand, can be defined as the spectrum energies in different localizations of a local block. Since the DC coefficient $F_{0,0}$ represents the average grayscale value of each $N \times N$ macroblock, it is not considered to carry any texture information. The remaining AC coefficients can be considered to characterize image texture and be used as texture features. Zhong and Jain (2000) pointed out that even though the DC component is used for color feature characterization and the remaining AC components for texture features, color and texture attributes are mixed together in the $(N \times N) - 1$ coefficients contained inside a pixel-block.

Most of the times, it is extremely hard to draw an absolute line between color and texture attributes, since color variation results in color texture. In this way, color is expected to be present at several AC coefficients, packing most of its spectral energy in the fewest number of low-frequency coefficients at the upper left corner of the macroblock. Zhong and Jain (2000) proposed to compute the absolute values of the AC coefficients, selecting those *M* lowest-frequency features carrying most of the energy. By rotating an image-block, the absolute values of the set of contained DCT coefficients remains unaltered, but their position along each zig-zag line is changed. However, by computing the distance between the corresponding matrices for the initial block and its rotated version, a totally false alarm is resulted in accordance with their perceptual similarity.

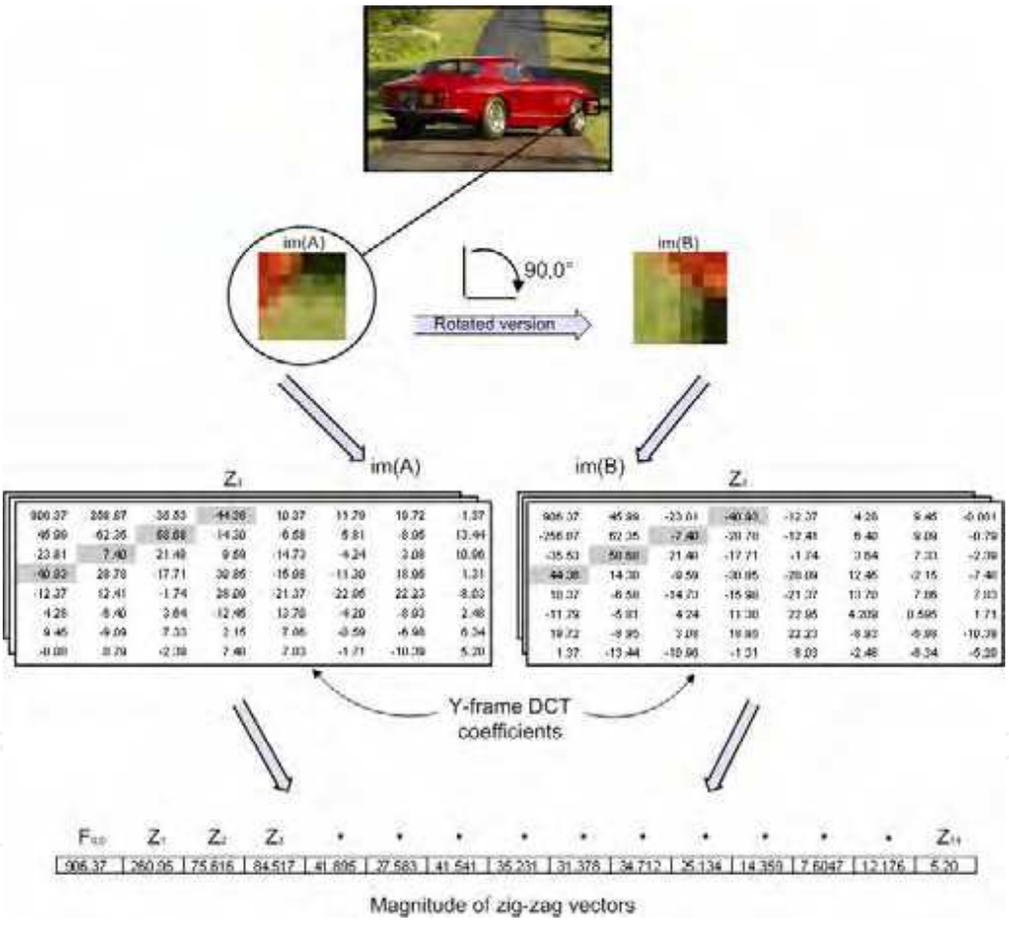


Fig. 6. DCT coefficients in the case of an image macroblock im(A) and its rotated version im(B). Each diagonal line of the zig-zag scheme is considered as a vector. The corresponding vectors (gray-shaded) contain AC coefficients having equal absolute value, although they are located at different positions. However, their magnitudes are the same, as provided at the bottom panel.

In this section, an efficient indexing method is outlined (Theoharatos et al., 2006b). Primarily, $k-$vectors are extracted from the diagonal zig-zag coefficients of each block, where a vector is defined by the AC components contained inside each diagonal line of the zig-zag scheme. The $k-$magnitudes $V_k$, $k=1,2,\ldots,2N-2$ of the corresponding zig-zag vectors are computed in the sequel, from $Z_1$ to $Z_k$ (in the case of $8\times8$ image block, $k=14$ as presented in Fig. 6). This representation has been proven to be robust to image geometric transformations. That is, by applying the DCT transform to an image block and its rotated version, the set of the absolute values of the DCT coefficients is identical, whereas their positions in the zig-zag ordering scheme are different (Theoharatos et al., 2006b). This obvious advantage is illustrated in the example of Fig. 6, where an image block of size $8\times8$ is extracted along with its $90^o$ right-rotated version and labeled *im(A)* and *im(B)* respectively. By applying the DCT transform to both initial and rotated image block, the set of the absolute values of the DCT coefficients is identical, but their positions are different in the zig-zag ordering scheme (depicted by the shaded lines in both matrices). Estimating the simple Euclidean distance between the corresponding zig-zag vectors of *im(A)* and *im(B)* using the proposed methodology, it was apparently found to be zero.

A critical issue that has to be solved is the number of selected AC components that need to be extracted from each image block, so as to represent effectively and efficiently the color and texture attributes. Owing to the very nature of the DCT, the set of AC coefficients generated for each $N\times N$ block are considered approximately uncorrelated. For an $N\times N$ pixel-block, the general intention is to choose those $M-$features out of the total number of $N^2$ DCT coefficients (except from the DC component that is always chosen as color attribute) that capture most of the spectral energy, while in our case, to select $k-$vectors out of the $2N-1$ ones estimated inside an image block. The number of selected texture and color-texture features must be extracted separately from each image channel. By testing with several JPEG images and using standard statistical methods (Duda et al., 2001) such as entropy estimation, the number of extracted zig-zag vectors was approximately found to be $k=3$, therefore using the first three zig-zag vectors. It should be noticed here that the extraction of color and texture features (i.e. the DC component and the $k-$zig-zag vectors) from each chromatic frame, increases the dimensionality of the derived feature space. However the computational complexity is not increased due to the fact that the WW-test is a function of the number of input vectors and not of their dimensions. On the other hand, the similarity measure is optimized by the higher number of extracted image features. Additionally, the optimal number of extracted vectors from the Y- frame was experimentally found to be 8, increasing the dimensionality of the feature space to 16 (8-dimensions for the luminance Y-frame and 4-dimensions for each of the two chrominance channels).

The performance of the proposed indexing scheme was evaluated on the same query-by-example retrieval system, using the WW-test as the similarity measure and following the standard procedure of constructing the Precision vs. Recall diagram. The Pr- and Re-indices were first evaluated for different sizes $T$ of the selected list ($T=5:5:50$) and the computed values were used in the plot of Fig. 7. The corresponding curves for the other techniques presented earlier have also been included in the same figure. In all curves, the same number of color and color-textures features has been extracted from each image macroblock, enabling the direct comparison of the different approaches. As we can perceive, the

proposed methodology outperforms all other techniques, having in all cases of the selected list T of retrieved images significantly higher precision rate.

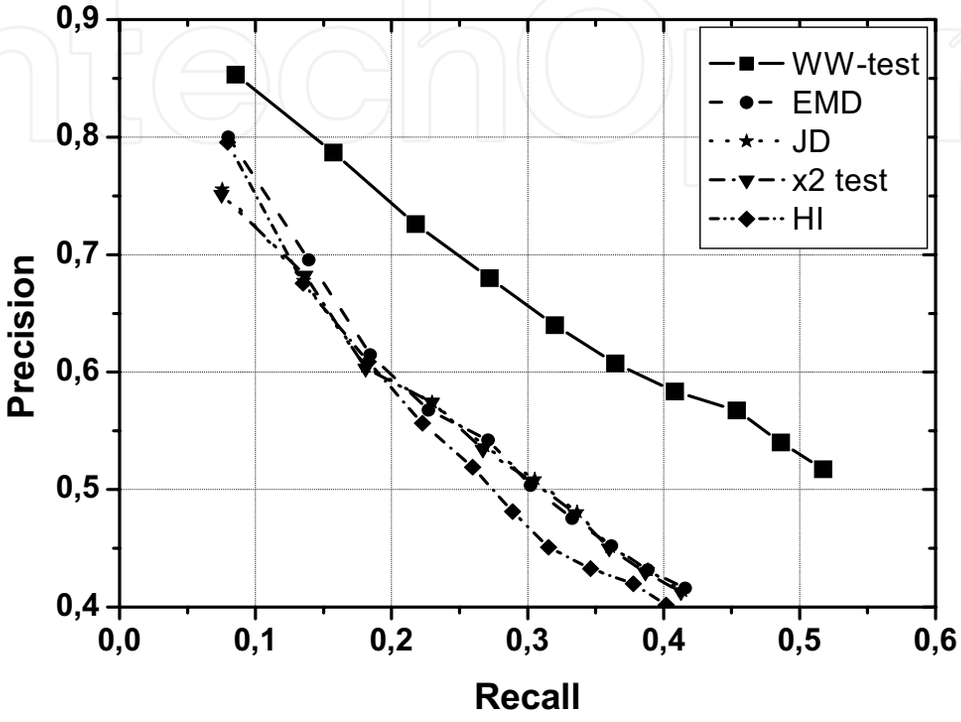


Fig. 7. Precision vs. Recall diagrams for the proposed compressed-domain retrieval scheme, in comparison to other related techniques also applied using the same indexing scheme in the compressed domain.

## 6. Conclusions

An intelligent strategy to visual information similarity is introduced based on the use of the nonparametric multivariate Wald-Wolfowitz statistical test. Our approach relies on a dual segregation-integration algorithmic step. The set of low-level characteristics is extracted in the form of an ensemble of feature-vectors and then 'set-differences' are computed between pairs of image representations. The new method is built on firm mathematical concepts, providing us with all the practical advantages of employing a suitable distributional distance. Its intelligent character stems from the fact that by altering the involved visual information, we can modify the flavour of formulated queries. By measuring the performance of the proposed distributional measure using some pre-defined feature extraction procedures, we show that it outperforms previously related ones that are considered as classical approaches for image similarity. The suggested methodology is evaluated within a query-by-example image retrieval scheme.

The only seemingly weak point of the proposed scheme is that it relies on the formation of MST, which is known to be a computationally demanding procedure. To provide some

insight about the complexity, the MST construction requires computational time $O(N^2)$ using a standard algorithm (Laskaris & Ioannides, 2001), while the test statistic can be evaluated in time $O(N)$, where $N$ is the number of involved data points. The selection of a small number of input feature vectors can alleviate the computational load of the WW-engine and is fully justified by the presented experimental results. These results show that our method reaches the maximum performance for moderate size of visual attributes and therefore a more detailed distributional representation is unnecessary (Theoharatos et al., 2006a; Theoharatos et al., 2006b). Apart from this experimental fact, it should be noticed that, nowadays, the theory of randomized algorithms (Motwani & Raghavan, 2000) provides alternative fast approximations to the MST construction problem. Using such algorithms, the efficiency of the presented method might improve further.

## 7. Future trends

Future research remains to examine different/advanced representations of image content so as to be embedded in the WW-engine. For instance, blob representations of images emerging from a context-dependent segmentation algorithmic procedure could be incorporated in the retrieval scheme, also done within the EMD-framework (Greenspan et al., 2004). In this way, we will able to compare images that are considered to be semantically more relevant and which require the identification of specific types of objects and scenes. This can be accomplished by modifying the visual attribute-extraction process from that of *primitive* features (such as color, texture, shape or spatial location of image elements) to that of *logical* features (such as the identity of the objects depicted in an image). The most appealing and simultaneously straightforward adjustment is definitely the engagement of the recently proposed neuromorphic training scheme (Laskaris & Fotopoulos, 2004) that leads to image content representations that are highly relevant to human visual perception.
The problem of modelling image semantics needs to be systematically examined, so as to be incorporated in the standard WW-framework. In this way, techniques that capture the semantic meaning of images have to be studied for perceptual categorization and WW-based similarity of color images, using low-level descriptors derived from high-level semantic primitives. Recent research focuses on implementing perceptually motivated feature extraction algorithms into real-working environments. In their work, Mojsilovic and Rogowitz (2004) performed several subjective experiments in order to understand important semantic categories that drive our visual perception and, afterwards, extracted meaningful low-level descriptors from these semantic categories in order to perceptually characterize the database images. By integrating these features into our WW-engine, enhanced retrieval results and better organization of image databases can be achieved (Theoharatos et al., 2007). Finally, other intelligent methodologies (Eakins, 2002) can be directly adopted in our system in order to improve the matching process and also provide the significance level of perceptual image similarity using semantically relevant visual attributes.

## 8. Acknowledgements

This work was financed by the European Social Fund (ESF), Operational Program for Educational and Vocational Training II (EPEAEK II), and particularly the Program "New graduate programs of University of Patras".

## 9. References


Castelli, V. & Bergman, L.D. (2002). *Image databases: Search and retrieval of digital imagery,* New York: John Wiley & Sons, ISBN 0-471-32116-8, USA.

Duda, R.O.; Hart, P.E. & Stork, D.G. (2001). *Pattern Classification*, 2nd Edition, New York: John Wiley & Sons, ISBN 0-471-05669-3, USA.

Eakins, J.P. (2002). Towards intelligent image retrieval. *Pattern Recognition,* Vol. 35, No. 11, pp. 3-14.

Friedman, J.H. & Rafsky, L.C. (1979). Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *Annals of Statistics,* Vol. 7, No. 4, pp. 697-717.

Gdalyahu, Y.; Weinshall, D. & Werman, M. (2001). Self-organization in vision: stochastic clustering for image segmentation, perceptual grouping, and image database organization. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 23, No. 10, pp. 1053-1074.

Gray, R.M. (1984). Vector quantization. *IEEE ASSP Magazine,* Vol. 1, No. 2, pp. 4-29.

Greenspan, H.; Dvir, G. & Rubner, Y. (2004). Context-dependent segmentation and matching in image databases. *Computer Vision and Image Understanding,* Vol. 93, No. 1, pp. 86-109.

Haykin, S. (1999). *Neural networks – a comprehensive foundation,* 2nd Edition, Canada: Prentice Hall Inc., ISBN 0-13-908385-5, USA.

Jiang, J.; Armstrong, A. & Feng, G.C. (2004). Web-based image indexing and retrieval in JPEG compressed domain. *Multimedia Systems,* Vol. 9, No. 5, pp. 424-432.

Kohonen, T. (2001). *Self-organizing maps,* 3nd Edition, Berlin: Springer-Verlag, ISBN 3-540-67921-9, USA.

Laskaris, N.A. & Ioannides, A. (2001). Exploratory data analysis of evoked response single trials based on minimal spanning tree. *Clinical Neurophysiology,* Vol. 112, pp. 698-712.

Laskaris, N.A. & Fotopoulos, S. (2004). A novel training scheme for neural-network based vector quantizers and its application in image compression. *Neurocomputing*, Vol. 61, pp. 421-427.

Leow, W.K. & Li, R. (2004). The analysis and applications of adaptive-binning color histogram. *Computer Vision and Image Understanding,* Vol. 94, No. 1-3, pp. 67-91.

Martinez, T.M.; Berkovich, S.G. & Schulten, K.J. (1993). "Neural-Gas" network for vector quantization and its application to time-series prediction. *IEEE Trans. on Neural Networks,* Vol. 4, No. 4, pp. 558-569.

Martinez, T.M. & Schulten, K.J. (1994). Topology representing networks*, Neural Networks,* Vol. 7, No. 3, pp. 507-522.

Mojsilovic, A. & Rogowitz, B.E. (2004). Semantic metric for image library exploration. *IEEE Trans. on Multimedia,* Vol. 6, No. 6, pp. 828-838.

Motwani, R. & Raghavan, P. (1995). *Randomized Algorithms,* U.K.: Cambridge University Press, ISBN 0-521-47465-5, USA.

Ngo, C.-W.; Pong, T.-C. & Chin, R.T. (2001). Exploiting image indexing techniques in DCT domain. *Pattern Recognition,* Vol. 34, No. 9, pp. 1841-1851.

Paschos, G.; Radev, I. & Prabakar, N. (2003). Image Content-based Retrieval using Chromaticity Moments. *IEEE Trans. on Knowledge and Data Engineering,* Vol. 15, No. 5, pp. 1069-1072.

Prim, R.C. (1957) Shortest connection networks and some generalizations. *Bell Sys Tech J*, Vol. 36, No. 6, pp. 1389-1401.

Rubner, Y.; Tomasi, C. & Guibas, L.J. (2000). The earth mover's distance as a metric for image retrieval. *Int. Journal of Computer Vision,* Vol. 40, No. 2, pp. 99-121.

Rubner, Y.; Puzicha, J.; Tomasi, C. & Buhmann, J.M. (2001). Empirical evaluation of dissimilarity measures for color and texture, *Computer Vision and Image Understanding,* Vol. 84, No. 1, pp. 25-43.

Schettini, R.; Ciocca, G. & Zuffi, S. (2001). A Survey on methods for colour image indexing and retrieval in image databases. In: *Color Imaging Science: Exploiting Digital Media*, Luo, R.; MacDonald, L. (Ed.), New York: John Wiley & Sons.

Sheikholeslami, G.; Chang, W. & Zhang, A. (2002). SemQuery: Semantic clustering and querying on heterogeneous features for visual data. *IEEE Trans. on Knowledge and Data Engineering,* Vol. 14, No. 5, pp. 988-1002.

Smeulders, A.W.M.; Worring, M.; Santini, S.; Gupta, A. & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 22, No. 12, pp. 1349-1380.

Stricker, M. & Orengo, M. (1995). Similarity of Color Images, *Proceedings of SPIE Storage and Retrieval for Image and Video Databases III*, pp. 381-392, San Jose, CA, USA, 9 Feb. 1995.

Swain, M.J. & Ballard, D.H. (1991). Color indexing. *Int. J. Computer Vision,* Vol. 7, No. 1, pp. 11-32.

Theoharatos, Ch.; Economou, G. & Fotopoulos, S. (2005). Color edge detection using the minimal spanning tree. *Pattern Recognition,* Vol. 38, No. 4, pp. 603-606.

Theoharatos, Ch.; Laskaris, N.A.; Economou, G. & Fotopoulos, S. (2005). A generic scheme for color image retrieval based on the multivariate Wald-Wolfowitz test. *IEEE Trans. Knowledge and Data Engineering,* Vol. 17, No. 6, pp. 808-819.

Theoharatos, Ch.; Laskaris, N.A.; Economou, G. & Fotopoulos, S. (2006a). Combining self-organizing neural nets with multivariate statistics for efficient color image retrieval. *Computer Vision and Image Understanding,* Vol. 102, No. 13, pp. 250-258.

Theoharatos, Ch.; Pothos, V.K.; Laskaris, N.A.; Economou, G. & Fotopoulos, S. (2006b). Multivariate image similarity in the compressed domain using statistical graph matching. *Pattern Recognition, Special Issue on Similarity-based Pattern Recognition,* Vol. 39, No. 10, pp. 1892-1904.

Theoharatos, Ch.; Laskaris, N.A.; Economou, G. & Fotopoulos, S. (2007). On the perceptual organization of image databases using cognitive discriminative biplots. *EURASHIP Journal on Applied Signal Processing*, *Special Issue on Image Perception*, Volume 2007, Article ID 68165, 15 pages, in press.

Wallace, G.K. (1991). The JPEG still picture compression standard. *Communication ACM,* Vol. 34, No. 4, pp. 31-45.

Zahn, C.T. (1971). Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Trans. on Computers,* Vol. C-20, No. 1, pp. 68-86.

Zhong, Y. & Jain, A.K. (2000). Object localization using color, texture and shape. *Pattern Recognition,* Vol. 33, No. 4, pp. 671-684.


**Vision Systems: Segmentation and Pattern Recognition**
Edited by Goro Obinata and Ashish Dutta

ISBN 978-3-902613-05-9
Hard cover, 536 pages
**Publisher** I-Tech Education and Publishing
**Published online** 01, June, 2007
**Published in print edition** June, 2007

Research in computer vision has exponentially increased in the last two decades due to the availability of cheap cameras and fast processors. This increase has also been accompanied by a blurring of the boundaries between the different applications of vision, making it truly interdisciplinary. In this book we have attempted to put together state-of-the-art research and developments in segmentation and pattern recognition. The first nine chapters on segmentation deal with advanced algorithms and models, and various applications of segmentation in robot path planning, human face tracking, etc. The later chapters are devoted to pattern recognition and covers diverse topics ranging from biological image analysis, remote sensing, text recognition, advanced filter design for data analysis, etc.



**InTech Europe**
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

**InTech China**
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen