Research Article

On the Perceptual Organization of Image Databases Using Cognitive Discriminative Biplots

Christos Theoharatos,[1] Nikolaos A. Laskaris,[2] George Economou,[1] and Spiros Fotopoulos[1]

[1] Electronics Laboratory, Department of Physics, University of Patras, 26500 Patras, Greece
[2] Artificial Intelligence and Information Analysis Laboratory, Department of Informatics,
  Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

A human-centered approach to image database organization is presented in this study. The management of a generic image database is pursued using a standard psychophysical experimental procedure followed by a well-suited data analysis methodology that is based on simple geometrical concepts. The end result is a *cognitive discriminative biplot*, which is a visualization of the intrinsic organization of the image database best reflecting the user's perception. The discriminating power of the introduced cognitive biplot constitutes an appealing tool for image retrieval and a flexible interface for visual data mining tasks. These ideas were evaluated in two ways. First, the separability of semantically distinct image classes was measured according to their reduced representations on the biplot. Then, a nearest-neighbor retrieval scheme was run on the emerged low-dimensional terrain to measure the suitability of the biplot for performing content-based image retrieval (CBIR). The achieved organization performance when compared with the performance of a contemporary system was found superior. This promoted the further discussion of packing these ideas into a realizable algorithmic procedure for an efficient and effective personalized CBIR system.

1. INTRODUCTION

The notion of image similarity has been exhaustively studied through the last decades in the field of computer vision. Its application includes recognition, classification, retrieval, and database organization, while its formulation is usually based on low-level attributes. Although such features (e.g., color, shape, texture, or combination of them) are considered to contribute to human judgment for image similarity, the idiosyncrasies of human perception are not fully considered at the algorithmic stage of feature extraction. The only exception is the recently developed research directions of relevance feedback [1, 2] and active learning procedures [3], in which the user is engaged to some iterative procedures aiming to alter the relative importance among the bunch of preselected features such that the modified machine-vision procedure matches his/her perception. Although significant progress has been reported since the early years [4], none of the existing methodologies are able to entirely encapsulate the semantic concepts necessary for expressing a high-level (i.e., cognitive) similarity between images.

Information regarding visual perception is highly desirable and related knowledge can be acquired from diverse research fields. In general, advanced perceptual systems cope with the problem of providing, as much as possible, an optimal association between the outside world and the internal representation of human inspection. Evidence concerning perceptual similarity is commonly provided through psychophysical experimental paradigms followed by suitable quantitative analysis [5, 6]. Such investigations on visual perception attempt to bridge the gap between the high-level semantics that people actually use to estimate image similarity and the low-level visual attributes that are widely popular in computer-vision community (e.g., [6–8]). The principal information dimensions of the human visual system include shape, size, orientation, intensity, color, and texture. In addition, several other variables, such as elevation, shadows, projection, motion, and depth, play a subtler role [9]. In order to interpret an image and/or extract properties before representing its similarity, individuals combine the previous set of features, known as semantic indicators or cues [8], in a very complex and task-dependent manner. A number of theories exist in the field of cognitive neuroscience about how brain carries out so efficiently the task of perceptual representations and comparisons [10–14]. For instance, the philosophical foundation of shape representation is known

to be isomorphism: structural or metric information stored in the brain reflects corresponding properties of shapes in the world, on one-to-one basis [15]. In consequence, the perception of shapes as instances of objects categories is considered to emerge via the comparison of incoming visual stimuli (the images of objects) with the stored representations in the human brain [11, 16].

The purpose of this work was to investigate the issue of image similarity from the perspective of human observer, regarding the specific tasks of handling and querying visual databases. Our objective was the structural organization of a given database that would reflect the user's intrinsic perceptual characteristics/strategies. In this enterprise, we meant to surpass the stage of feature selection that required strong assumptions about the character of perception. To achieve this, we adopted a well-founded psychophysical tactic consisting of an easy-to-run experimental procedure and accompanied with a battery of data-learning techniques that provided an ingenious representation of results in the form of a *biplot* [17]. This was a 2D point diagram reflecting the perceptual comparisons between all the pairs of database images. To validate the emerged structure, we contrasted it with the one reflecting machine perception (i.e., a mathematically defined similarity measure that compares features extracted from the images). The surprising success of biplots, in the sense that they not only provided a faithful representation of perceptual differences, but also acted as an invaluable interface for visual data mining, and motivated the further development of our approach to an easy-to-implement algorithmic paradigm for *personalized* organization and tracing of visual databases.

Regarding methodology, the concept of biplots is the mathematical cornerstone of this paper. Biplots have been introduced recently [17], as an extension to multidimensional scaling (MDS) techniques, which is a well-developed branch of multivariate statistics dealing with relational data (i.e., data in the format of a distance matrix) [18]. A "*biplot*" is defined as a low-dimensional (usually, but not necessarily, 2D) plot that simultaneously represents, in a meaningful way, the objects under study, the relationships between these objects, and the variations in the data ensemble. The basic rule for interpreting these point diagrams is that similarity between objects is reflected by their closeness on this diagram. Their main characteristic is that the explicit (i.e., using features) representation of the encountered objects is not required. In our case, the ensemble of images in the visual database is the set of objects to be represented as a point sample in the biplot. The structure emerging in this point diagram will reflect all the perceptual comparisons, and therefore portrays the human-centered organization of the given database.

To ease the presentation, we avoided the standard division of the presented material to methods and experimental results. We preferred instead to present all the individual steps employed in our analysis and to directly exemplify them by providing the corresponding outcomes simultaneously. At the end, we summarize what we have learned from this exercise and conclude by suggesting easy-to-implement algorithms for user-tailored content-based image

retrieval (CBIR) and friendly visual data mining from image databases.

In our experimentations, $N_0 = 750$ photographic images were selected, which have been formerly preassigned by experts into $C = 15$ distinct classes. Each class contained $N = 50$ perceptually similar images. Using user-defined (perceptual) image similarities and *vector median* [19] estimation, a representative image is originally selected from each category and considered as class prototypical image. Then the organization of the utilized database starts by comparing the class prototypes and continues by feeding the results of perceptual comparisons into an MDS technique. The resulting biplot shows the 15 class prototypes organized in a low-dimensional (reduced) space. This visualized result is compared with the one corresponding to similarities estimated using an algorithmic procedure from the recent computer vision literature [20]. The entire image database is finally organized using the *appending technique* [21], an efficient geometrical scheme that incorporates any given image directly into the precomputed reduced space of prototypes in a way that geometrical relationships reflect original perceptual similarities. By measuring the similarity (based on perceptual judgment) of each image to be appended with only the class prototypes, robust and economical comparisons are achieved. The evaluation of the perceptually defined biplot is carried out using two different data analysis perspectives. First, the class-discrimination power of the resulting mapping is assessed using the multivariate *Wald-Wolfowitz test* (*WW-test*) [22]. Then, a nearest-neighbor retrieval scheme is built in order to estimate the precision performance when the emerged cognitive terrain is used as interface for database searches. For comparison purposes, results are also provided regarding indexing and retrieval based on a competitive machine-vision scheme that was recently developed by the authors [20, 23].

The rest of this paper is organized as follows. In Section 2, some aspects of perceptual similarity in association with visual recognition and categorization are described. The perceptual organization of the utilized image database is provided in Section 3. In Section 4, the induced database organization is evaluated. A simplified algorithm for user-dependent CBIR is suggested and demonstrated in Section 5. Finally, a concluding discussion is provided in Section 6.

2. PERCEPTUAL SIMILARITY AND CATEGORIZATION

Visual similarity of images is considered to be an explicit metacognitive process that is directly related to psychophysical and physiological mechanisms of information processing, taking place in a distributed network of distinct brain regions. Judging the amount of image similarity is not considered to be a primary function of our cognitive system. Instead, some neuroscientists hypothesize that it arises as a byproduct of image classification and categorization [11, 16]. Given the initial sets of visual features from a pair of incoming stimuli (i.e., objects or images), their comparison might include a series of *pattern recognition processes*. If the appearance of visual objects is not affected (or altered) by any

extrinsic factors, the visual recognition module simply compares the objects via template matching. Otherwise, the correspondence problem has to be solved first [10, 16]. The output of this similarity judgment depends on the accuracy of all intermediate steps.

According to reported psychological studies [15], similarity relations could be approximated through a geometrical modeling that considers the images as points in a coordinate system, such that the observed dissimilarities between images correspond to the metric distances between the respective points. However, it was proven that visual similarity does not appear to fully accord with predictions of a metric model, since—at least—symmetry does not hold [24]. To include such a behavior, the original model was modified in such a way that a comparison between the intrinsic visual information of the viewed image and some internally stored representations of image categories was encountered before the similarity estimation. That is, in order to estimate the amount of similarity between two given images, the visual system is required to originally determine the category in which each image belongs. For example, an image containing a tiger walking across a river is automatically classified by an individual into some "Tiger category." The images having no tiger-related pattern at all and which are to be compared with the previous image are considered to be totally irrelevant to that. Such a memory-based line of thinking can be directly related with the process of *categorization*, a task that requires the system to make sense of distinct visual primitives that are reliable class identifiers (i.e., such as those used in shape representation [15]).

A straightforward way to conceptualize the above ideas is to associate the images with points in a properly parameterized feature space and define a distance function within, so that each cluster of points corresponds to a distinct class of images. The relative location of the corresponding image point with respect to a specific cluster reflects the degree of image similarity to some predefined class category. Originally, the visual system carries out a huge number of measurements in a high-dimensional and, probably, rich-structured space. However, the estimation of pairwise image similarity is channeled through the formal categorization of the involved images into the stored representations in the human brain. In this sense, a similarity measure becomes equivalent to determining location differences (of the images to be compared) within a lower-dimensional space defined by the class prototypes [16]. It was exactly this doctrine that motivated the present study and guided the individual steps discussed in the following sections. In a nutshell, prototypical-images were selected and the corresponding perceptual similarities were defined and used in revealing the structure of the space in which these prototypes resided. A low-dimensional sketch of the prototypes' space (i.e., biplot) was calculated and further used to incorporate the whole set of database images.

3. PERCEPTUAL DATABASE ORGANIZATION

Image database organization is the procedure of classifying semantically relevant images having (probably, but not

TABLE 1: Categories of the utilized experimental image database.

ID	Category name	ID	Category name	ID	Category name
1	Airplanes	6	Fireworks	11	Mountains
2	Balloons	7	Flowerpots	12	Sea
3	Butterflies	8	Flowers	13	Sunrises
4	Cars	9	Grassland	14	Tigers
5	Eagles	10	Horses	15	Views

necessarily) similar visual attributes into discrete perceptual groups. In classical CBIR, the systems need not understand image semantics, but merely assign each image to a specific category using similarity measures based on low-level characteristics [4]. It is only the last few years that research is directed towards similarity measures realized based on high-level visual attributes [6–8]. Ideally the organization of an image database would align with human perception the database structure reflecting all the perceptual similarity relations among images. In what follows, we describe our attempts to organize a given database in fully accordance with human perception.

The ensemble of utilized images was a heterogeneous subset of the Corel gallery, including $N_0 = 750$ still color images of 24 bpp each, given in portable pixel map format of size either $[192 \times 128]$ or $[128 \times 192]$ pixels. The dataset contained a wide variety of images, from animals and plants, to views and natural images. Originally, it was formed by pre-assigning the images into $C = 15$ distinct classes (e.g., cars, eagles, flowers, etc.) as introduced in the SCHEMA reference system [25]. These are presented in Table 1. By incorporating the coherent opinion of five individuals, $N = 50$ semantically relevant images were kept for each category to form the entire data collection. The specific database was created in a way that is considered standard in CBIR-research community, in the sense that could facilitate the subsequent estimation of validation indices (see Section 4), which were necessary for the quantitative evaluation and comparison of the obtained results [4].

3.1. Representative images: the space of class prototypes

Prototyping was the first stage of our experimental procedure. It was dictated by the theoretical considerations described in the previous sections and aimed at the definition of anchor points in the perceptual space. Apart from this theoretical motivation, practical reasons (related with the extent of the subsequent psychophysical stage) made the extraction of representative images necessary. To make this selection as accurate as possible, we adopted the following scheme with the scope of selecting a single prototypical image from each distinct class based on perceptual similarity judgment. The display configuration for the selection process is shown schematically in Figure 1. Each time, a pair of images from a specific class was presented in randomized order on the screen, denoted as C_i and C_j, with $i, j = 1, 2, \ldots, N$ and $j \neq i$. The users could insert the degree of similarity for each pair of
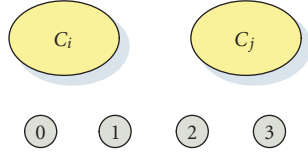
FIGURE 1: The display configuration used in our experiments for judging image similarity between images belonging to the same category, using perceptual attributes.

displayed images, via the four buttons shown at the bottom of the screen. The extreme-left button indicated the minimum degree of similarity (i.e., zero (0)), while the rightmost button the maximum (i.e., three (3)). Using paired comparisons to measure image similarity, all couples of stimuli were finally compared. Thus, for each pair of images, the users assigned a number proportional to perceived similarity in the 4-point scale of Figure 1.

To reinforce the experimental procedure, five subjects participated in the similarity estimation process and the reported results were obtained by averaging the computed measurements. By calculating all the pairwise similarity indices, $C = 15$ similarity matrices of sizes $N \times N$ ($N = 50$ in our case) were obtained, $S_k(i, j)$, with $i, j = 1, 2, \ldots, N$ and $k = 1 : C$. In order to estimate the prototypical image P_k from each distinct category, we followed a procedure that is reminiscent of the vector median algorithm [19]. We identified the image (among the N belonging to that category) that could be considered as the class centroid. Algorithmically, this was accomplished readily by first taking the sum across all rows of each proximity matrix S_k and then identifying the entry of maximum in the produced vector. The image corresponding to this maximal sum of similarities was selected as our class prototype:

$$P_k \longrightarrow j_k^{\text{sel}} = \left\{ \arg\max \left(\sum_{i=1}^{N} S_k(i, j) \right) \right\}, \tag{1}$$
$$j = 1, 2, \ldots, N, \quad k = 1, 2, \ldots, C.$$

Figure 2 presents the 15 representatives extracted from all $C = 15$ image categories. This set of images, despite the fact of being a very sparse sampling from the multidimensional perceptual space, can be considered as a meaningful coarse outline. It is denoted hereafter as the set of *class prototypes*. In short, the selected prototypes (using the previously described metacognitive process) appear to be those members that most reflect the redundancy structure in their own categories. Taking into consideration that the preassignment of database images into semantic classes was implemented so as to maximize the common visual information, the representative images of Figure 2 can be thought of as the reference points (semantic landmarks) of the perceptual space. Perceptual similarities between any pair of database images could therefore be approximated after investigating the perceptual similarities between these prototypes. The problem is now transformed to that of learning the structure in the set of class prototypes.

3.2. A sketch of image-perception space: the cognitive biplot of prototypes

According to Edelman's theory of representation [11, 15], to make sense of a stimulus means to locate it in a low-dimensional psychological space. Psychophysical experiments are employed in order to learn and parameterize this space. Very often subjects are asked to register their estimation about stimuli similarities during paired presentation of different stimuli spanning the space to be explored. MDS techniques are then applied to the collected data as a means of approximating the inaccessible space. In our case, the different prototypical images played the role of the indicative set of stimuli, the subjects denoted the amount of perceptual similarity between any pair of prototypes, and MDS provided the sketch of perceptual space. MDS is defined as any procedure that, given a dissimilarity matrix corresponding to a set of objects (here the perceived stimuli), configures points in a low-dimensional space (usually 2D) as images of the objects in a way that the interpoint distances approximate as much as possible to the original pairwise dissimilarities [18].

To arrange the set of prototypes $\{P_k\}_{k=1:C}$ in an r-dimensional space ($r < C$), a similarity matrix of C-prototypes needs to be created, based on the perceptual judgment among these class representatives. Using the display setup shown in Figure 1, the C-prototypes ($C = 15$ in our case) were presented in pairs and the users were asked to register the perceived similarities by selecting a value in the range [0–3]. All the pairs were portrayed in randomized order and the values were collected, after averaging across all subjects, in order to form the $[C \times C]$ similarity matrix $\mathbf{S}$. These values were weighted by integer numbers in the range [0–14] conveying the information regarding the relative ordering of similarity. This modulation was considered necessary, since during the initial round of experimentation we noticed that the similarity level of zero dominated the measurements. Taking into consideration that the perceived similarity depends on the general context of the image pairs to be contrasted, we introduced row vectors W_i, $i = 1, 2, \ldots, C$, with each one corresponding to a specific prototype and having values $W_i(j)$, $j = 1, 2, \ldots, C$ a certain permutation of the set $\{0, 1, \ldots, C - 1\}$. Putting it in words, since the "tiger" image P_{14} is considered to be perceptually closer to the "horse" image P_{10}, the weight $W_{14}(10) = 1$ should be defined, and since the "grassland" prototype comes next, the weight $W_{14}(9) = 2$ should be defined accordingly, and so on (while $W_{14}(14) = 0$). Using these modulating weights, the similarity matrix became $\mathbf{S}' = \mathbf{S} \cdot^* \mathbf{W}$ (with "$\cdot^*$" denoting array multiplication). Finally, the entries of $\mathbf{S}'$ matrix were normalized (by division with the maximum) so as to lie in the range [0-1].

The above matrix was converted to a dissimilarity matrix $\mathbf{D}$, $D(i, j) = 1 - S'(i, j)$, which in turn underwent a transformation step, via the "minimum-rule" $D(i, j) \leftarrow \min(D(i, j), D(j, i))$ to become symmetric. The original asymmetric nature of $\mathbf{D}$ was due to the fact that perceptual similarity between prototype images P_i and P_j was estimated by interchanging each time the reference image. The estimated $[C \times C]$ dissimilarity matrix was next entered to

1: Airplanes 2: Balloons 3: Butterflies 4: Cars 5: Eagles

6: Fireworks 7: Flowerpots 8: Flowers 9: Grassland 10: Horses

11: Mountains 12: Sea 13: Sunrises 14: Tigers 15: Views

FIGURE 2: The selected class prototypes.

a routine performing the classical (metric) MDS algorithm [26] (e.g., using the Matlab implementation of *cmdscale* routine). The output was given by the formula $\mathbf{Y}_{[C \times r]} = \text{MDS}_r(\mathbf{D})$, defining a set of C-vectors lying in an r-dimensional space ($r < C$, is a user-defined parameter) [18, 26]. In this way, the $C = 15$ class prototypes were embedded as a r-dimensional point set, with interpoint distances being analogous to perceptual dissimilarities between images. In general, the higher the dimension r is, the more reliable this representation is. However, the use of small values for r (i.e., $r < 4$) enables a visualization that can provide direct insight into the structure of perceptual space. The creditability of this low-dimensional diagram can be evaluated using the following index:

$$E = \frac{\sum_{i,j>i}^{C} D(i,j) - \|y_i - y_j\|_{L_2}}{\sum_{i,j>i}^{C} D(i,j)}. \tag{2}$$

Using $r = 2$, we obtained $E \approx 0.91$, a result with twofold significance. It indicates that the image-perception space was indeed low-dimensional as predicted by theory (see Section 2), and moreover that a trustworthy representation of the space of image prototypes could be provided in a convenient 2D format.

In Figure 3, the organization of class prototypes reflected in the MDS-based point diagram is given in the form of a 2D biplot in order to ease interpretation (the computed 2D coordinates have been used to place the corresponding *thumbnails*). The displayed biplot not only approximates the per-

ceptual space, but also provides a self-explanatory representation of the structure in the set of class prototypes.

Not surprisingly, the above 2D visualization reflects the influence of color. This is in agreement with the current belief in CBIR, according to what color-related features are sufficient characteristics for organizing color-image databases. In addition, color information has been generally recognized as the most important indicator of the general "mood" of an image and seems to correlate well with semantics [6–8]. In Figure 3, for example, images with prominent blue color are located at the bottom-left region of this biplot. Moreover, the categories of "Airplanes" and "Eagles" tend to gather together and deviate from "Views," "Mountains," and "Sea" classes. On the other hand, images with green hue appear at the bottom-right corner; among them, the "Tigers" and "Horses" categories tend to cluster together due to their close semantic meaning. Thus, the prototypical images seem to be distributed into different color-related regions in the derived diagram. Apart from this, the semantic knowledge of the pre-assigned classes is also reflected in the shown geometrical arrangement of selected class prototypes.

3.3. A sketch of machine perception based on derived image similarities

To fully justify the previous biplot of Figure 3, we compared it with the corresponding one arising from the pairwise comparisons of class prototypes using a standard image-similarity measure from machine-vision literature. Although
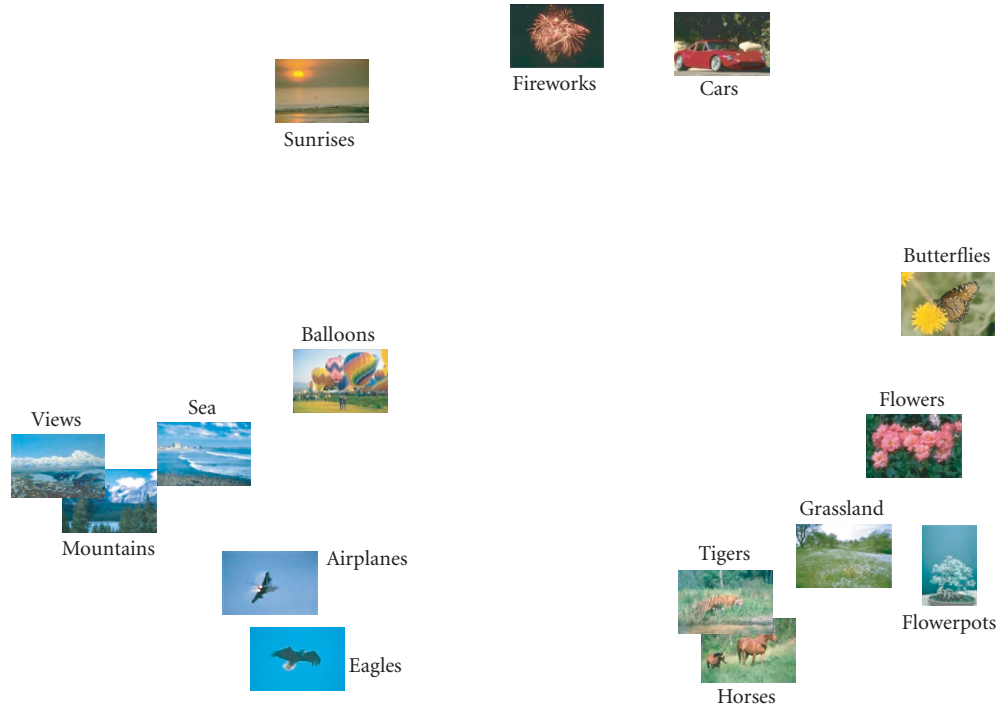
Figure 3: 2D biplot reflecting the perceptual differences between the class prototypes.

information derived from high-level semantics may greatly facilitate image similarity matching, the current state of computer-vision systems have not yet reached to a generally admissible accuracy level, not to mention that semantic information is not easily and automatically extracted from general-purpose photographic images [6]. Since color is actually an important cue for organizing our database (images belonging to Corel dataset contain profound color information in contrast to texture and shape attributes), we relied on a recently introduced algorithm for expressing color-image similarity [20, 23]. The algorithm relied on the extraction of a set of low-level characteristics (expressing the color spatial layout) from each image to be compared and the subsequent statistical comparison using the multivariate WW-test [20, 22]. To align well with the human color perception, the whole algorithmic procedure was implemented in the perceptually uniform CIE-Lab color space. Dissimilarities were computed for all pairs of class prototypes and tabulated in a $[C \times C]$ distance matrix that entered the MDS routine as previously (Section 3.2). The produced geometrical arrangement of the class prototypes is given in Figure 4 (for $r = 2$, the $E \approx 0.73$).

The positive influence of color differences in the pairwise comparisons of class prototypes is clearly apparent in the biplot of Figure 4. However, there are several semantically irrelevant classes clustered together. For instance, the "Flowerpots" category is grouped together with the "Sea," "Mountains," and "Views" categories. The "Cars" and "Balloons" classes also tend to mix in this diagram. In addition, other class prototypes that subjects perceived as close enough in

perceptual space (see Figure 3) are arranged as totally irrelevant in the specific machine-vision-based biplot of Figure 4 (e.g., the "Flowerpots," the "Flowers," and the "Grassland" prototypes).

3.4. Perceptual organization of image database: the appending technique

The biplot of Figure 3 provides only a rough representation of the perceptual space geometry regarding the given database, since its computation involved merely the $C = 15$ class prototypes. Despite being very informative regarding the perceptual variations in the database, it does not provide any information about the detailed structure of the whole image collection. In principle, such information could be obtained by repeating the procedure of Section 3.2 for the whole set of $N_0 = 750$ images. This would result in an extended biplot, that is, a point diagram of 750 vectors. Since the class prototypes were selected among the 750 images, this extended biplot should contain the original one of Figure 3, making clear how semantic classes are formed. However, there was a practical problem in implementing this idea. The amount of necessary pairwise comparisons $((1/2) \times 750 \times 749)$ was prohibitive. To surmount this problem, we adopted the consideration that in order to locate a stimulus in perceptual space, it is enough to define its similarity with respect to some stored prototypes (see Section 2). Therefore, only 750×15 comparisons should be registered instead. Fortunately, there was available in the literature [21] an algorithmic procedure for augmenting a precomputed
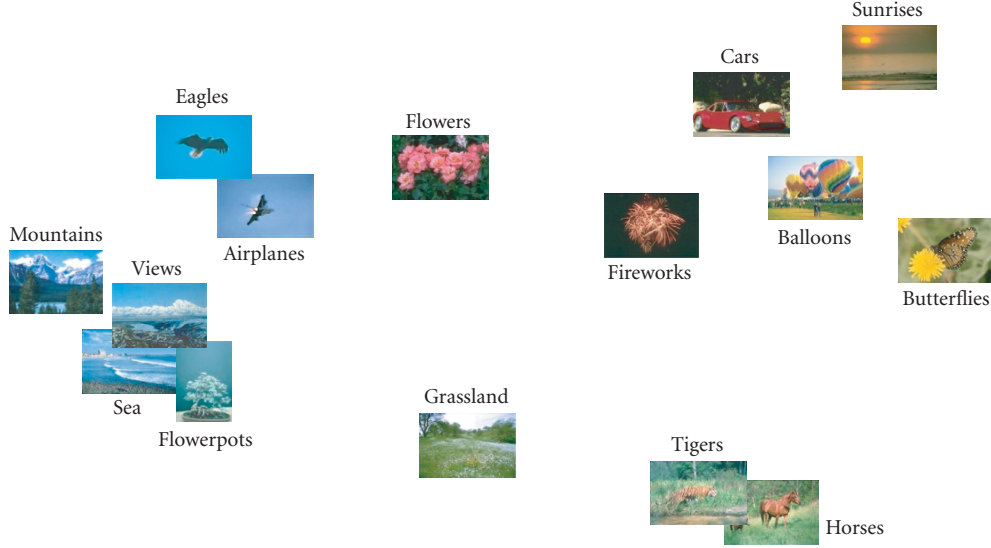
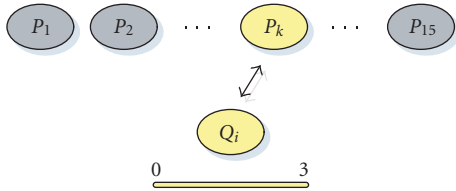FIGURE 4: 2D biplot reflecting the machine perception, regarding the class prototypes.

FIGURE 5: The experimental display for expressing perceptual similarity between a query image Q_i and the set of class prototypes P_k.

MDS-based point diagram. This enabled us to incorporate the "projection" of the entire image database in the biplot of Figure 3.

At the experimental stage of this procedure, we used the setup shown in Figure 5. All database images were compared against the $C = 15$ class prototypes P_k, which were presented simultaneously on the top of the display. In a randomized order, a database image was selected and displayed at the bottom. For each "query"-image Q_i, the subjects were asked to assign the degree of its similarity with each one of P_k. This resulted in a $[C \times 1]$ similarity vector, which was then modulated by weights denoting the rank ordering of prototypes (as described in Section 3.2). These N_0-vectors were gathered in a $[C \times N_0]$ matrix, the entries of which were normalized (by division with the maximum) so as to lie in the range $[0 - 1]$. Similarities were then transformed to dissimilarities and tabulated in matrix $\mathbf{L}$.

After estimating the perceptual dissimilarities between the database images and the class prototypes, it became feasible to add in the biplot of the latter the points corresponding to the former. This task was accomplished via Gower's technique [21] employed in a batch mode [26]. In this way, the geometry of the augmented point diagram reflected all the

pairwise relations even among images that had not been directly compared by users.

Summarizing the so-called *appending-technique* [26], we remember that $Y_{[C \times r]}$ are the coordinates of class prototypes $\{P_k\}_{k=1:C}$ in the original biplot of Figure 3 (produced via MDS with $r = 2$). Using all the interset dissimilarities between $\{P_k\}_{k=1:C}$ and $\{Q_i\}_{i=1:N_0}$ tabulated in $\mathbf{L}$, we computed the coordinates of Q_i on the biplot (i.e., the appended points) via the following equation:

$$Y^{\text{app}}_{[N_0 \times r]} = \frac{1}{2}\left(\left(Y^T Y\right)^{-1} Y^T dL\right)^T,$$

$$dL_{[C \times N_0]} = \text{diag}(B) \cdot \vec{1}_{[1 \times N_0]} - \mathbf{L}, \quad \vec{1} = [1,\dots,1],$$

(3)

where $B = -(1/2)HD^*H$ is a matrix related to the centering operator $H_{[C \times C]}$ ($H(i,j) = \delta_{ij} - 1/C$) that is applied to the dissimilarity matrix of the class prototypes after squaring its elements, that is, $D^*(i,j) = D(i,j)^2$.

The whole image database could therefore be presented in a point diagram in which the coordinates of the class prototypes are the same as in the biplot of Figure 3. The rest of the images are expected to scatter around these reference points reflecting the detailed perceptual organization of the database. Figure 6 displays the augmented biplot. To efficiently visualize this result, different labels (i.e., symbols and colors) have been assigned to images from distinct classes. It is clear that all images coming from the same class are grouped around their initially selected prototypes (shown via filled black diamonds). The high concentration of the different categories in the 2D projection space is indicative of the accurate classification performance reported in the previous sections and the consistency of the visualization provided via the appending technique. Naturally, there exist some coupled categories that tend to mix together in the meaningful terrain
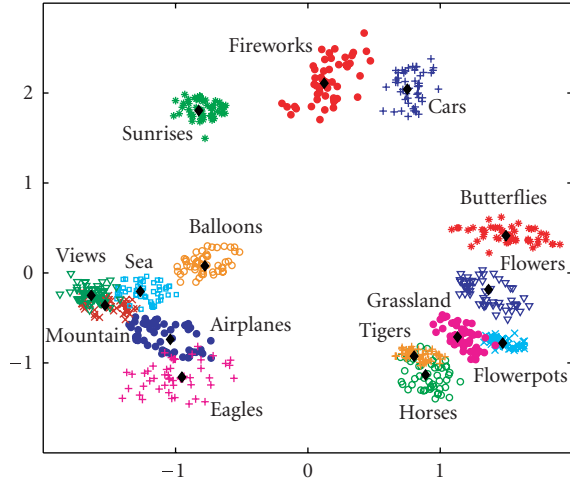
FIGURE 6: Perceptual organization of the entire database based on the efficient approximation facilitated via the appending technique.

of Figure 6, such as the "Tigers versus Horses," "Airplanes versus Eagles," "Mountains versus Views," and "Mountains versus Sea" categories.

A principal advantage of the encountered approach is that the entire database can be organized using the 2D interface provided via the appending technique. Perceptual similarities between images that had never been estimated could be easily inferred by the geometry of the biplot in Figure 6. In this biplot, semantics have been embedded during the definition of landmarks. We consider prototyping a highly cognitive task that—by defining reference points in perceptual space—influences the subsequent estimation of perceptual similarities. These semantics can be further augmented easily by incorporating the biplot in a GUI environment. The later could facilitate efficient tracing of the database and visual data mining as well. It is the scope of Section 4 to provide a quantitative validation of our belief that the employed perceptual charting of the database constitutes a useful tool for the efficient handling of image collections. In what follows, we have restricted ourselves to measure the gain in organization performance.

4. EVALUATION OF THE PERCEPTUAL DATABASE ORGANIZATION

The suggested cognitive organization of the utilized image database was characterized regarding class separability and effectiveness in a simple CBIR scheme. To enable comparisons, we included also the results from the application of the machine-vision methodology discussed in Section 3.3.

4.1. Testing class separability using the Wald-Wolfowitz test

In order to fully justify the term "discriminative," we considered it necessary to measure the degree of separability between the different image categories in the resulted cognitive biplot. To this end, we encountered the multivariate *Wald-Wolfowitz* test (*WW-test*) [22], a nonparametric test dealing with the multivariate *two-sample problem*. The specific test 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) [27]. The output of this test, which is expressed by the indices R and W as described in the appendix, can be expressed as the probability that the two point samples are coming from the same distribution. Its great advantage is that no a priori assumption about the distribution of points in the two samples is a prerequisite [20, 23].

WW-test was applied to pairs of image classes. For each pair of classes to be compared, the corresponding points from the biplot in Figure 6 were the vectorial inputs. The output was a value of the W-index. The more positive the W-value is, the higher the overlap between the corresponding point distributions is, and therefore the lower the class separability. Ideally, each individual image category should not mix together in the original perceptual space, since our cognitive judgment clearly separates them apart (as realized by the five individuals at the beginning of our psychophysical experiment). However, in the 2D cognitive biplot of Figure 6, there are some classes that seem to overlap. A few examples are provided in Figure 7, where the images belonging to overlapping categories are compared using the WW-test. In each single diagram, two sets of vectors corresponding to the 100 images (comprising the two different categories to be contrasted) are compared in the 2D space. The aggregate MSTs are constructed using Prim's algorithm [28] and the nodes corresponding to different categories are labelled differently (e.g., in Figure 7(a) the brown crosses are associated with the "Tiger images," while the green cycles are associated with the "Horse-images"). It can be easily noticed that the points corresponding to different samples constitute, roughly, different subtrees in the overall MST. The measured indices provided by the multivariate WW-test are (a) $R = 22$ and $W = -5.8431$, (b) $R = 14$ and $W = -7.4541$, (c) $R = 27$ and $W = -4.8351$, and (d) $R = 7$ and $W = -8.8653$. This clearly indicates the difference in the distribution governing each distinct image class.

The lowest class separability is detected between the "Mountain" and "Views" categories (Figure 7(c)). By looking at the prototype images of the corresponding classes (depicted in Figure 2), one may notice that they seem to be—semantically—very closely related. Such observations might prove significant in the original preassignment of database images in different categories. The previous two categories could be considered as a unified image class, reducing further the overall processing time of the appending technique as well as computational complexity.

To fully justify the previously reported class separability results, which were measured in the approximated perceptual space, we carried out the corresponding measurements in a space approximating the machine-vision perception (as this is emerging from the methodology in Section 3.3). Dissimilarities were estimated for all the pairs of images in the
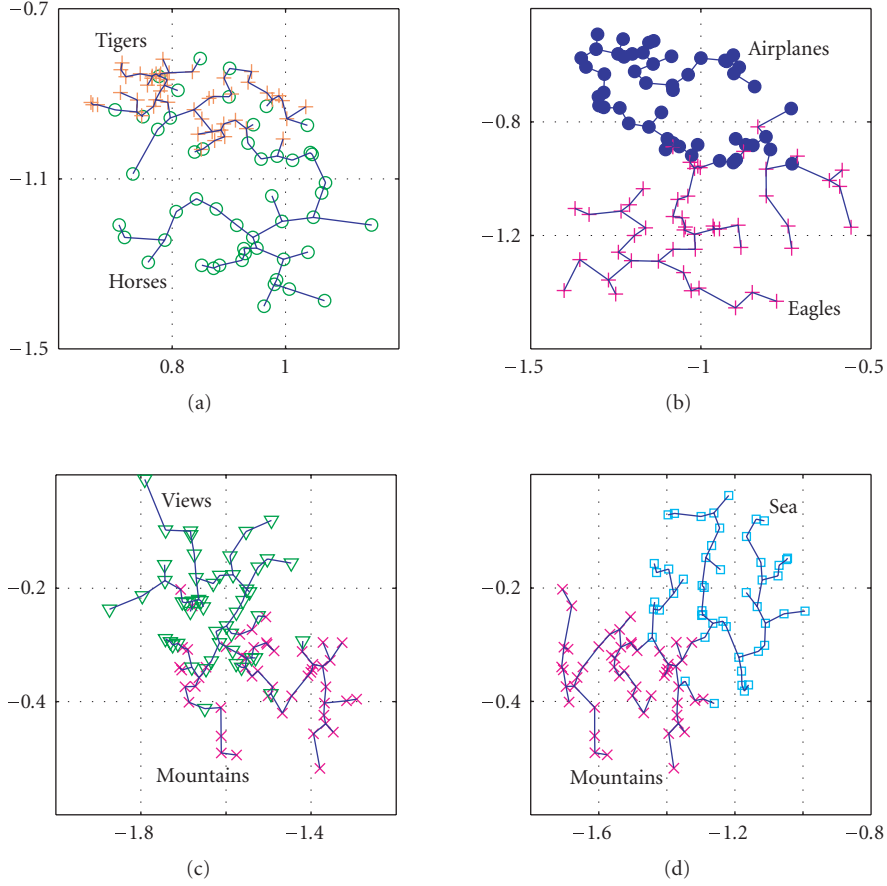
FIGURE 7: WW-test related measurement of class separability characterizing the cognitive biplot of Figure 6. (a) Tigers versus Horses: $W = -5.8431$, (b) Airplanes versus Eagles: $W = -7.4541$, (c) Mountains versus Views: $W = -4.8351$, and (d) Mountains versus Sea: $W = -8.8653$.

database. The application of MDS (using $r = 2$, E was estimated ~ 0.73) resulted in the point configuration shown in Figure 8 (left panel). For visualization purposes, two coupled pairs of categories are depicted in Figure 8 using the same symbols as in the corresponding diagrams of Figures 6 and 7 (while all the remaining images are depicted as a swarm of black dots). Apparently, the distinct image categories are relatively mixed together in the 2D reduced space, having low inter-cluster density. By applying the WW-test to the two depicted pairs of classes (see Figure 8, right panel), the measured indices were found to be (a) $R = 44$ and $W = -1.4098$ for the "Tigers versus Horses" couple and (b) $R = 46$ and $W = -1.0072$ for the "Airplanes versus Eagles" couple. As it was mentioned earlier, the bigger value of the R- and the higher value of the W-parameters indicate the lower discrimination power of the machine-vision approach (contrasted to Figure 7). By systematically comparing (all pairs of different categories) the results of WW-test for both database organization methodologies (i.e., cognitive biplots and the machine-vision-based scheme), we verified that a much higher class separability was achieved with the new approach. These outcomes motivated the inclusion of term *discriminative* in the name "cognitive biplot" so as to fully

characterize the performance of the computed organization terrain shown in Figure 6.

4.2. Using the approximated perceptual space for searches in the database

The high class separability in the produced biplot motivated us to further investigate its potential use for efficient image retrieval purposes. In the field of CBIR, query-by-example search engines are highly popular, since they can be constructed readily. The definition of a proper image-similarity measure is practically all that is needed. Using this measure, a comparison of the query-image with all the images in the database is performed and the most similar ones are retrieved (i.e., a search for nearest neighbors in the feature space implied by the similarity function). In this study, we found it interesting to measure the performance of a search engine built over the cognitive discriminative biplot. The perspective would be that after organizing a database as above, a query image could be easily compared with the semantic prototypes (in the setting of Figure 5), and using the appending technique could be placed on the discriminative terrain of Figure 6. There, the geometrical relationships actually
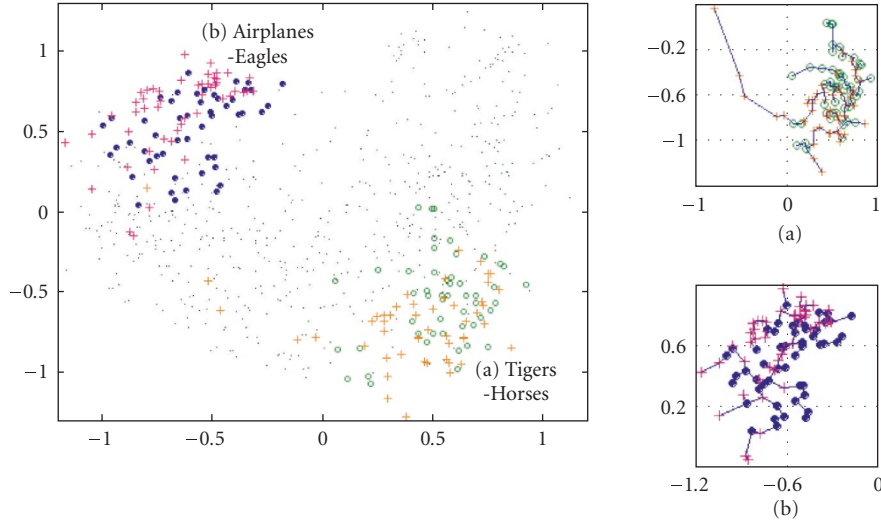
FIGURE 8: Left: MDS-based approximation of the perceptual space for a machine-vision system (see Section 3.3). Right: measuring class separability using WW-test, (a) Tigers versus Horses ($W = -1.4098$), (b) Airplanes versus Eagles ($W = -1.0072$).

extrapolate the measured perceptual comparisons to the whole database, and therefore the points close to the "projection" of the query image would correspond to the images to be retrieved from the database. To fully justify the above idea, we used an experimental procedure that is common practice in current CBIR research, that is the construction of *precision* (*Pr*) and *recall* (*Re*) curves for the corresponding query-by-example search engine [29].

In short, each image in turn was considered as the query image, and therefore its appended "projection" on the biplot of Figure 6 was the starting point of the search. We estimated all the $N_0 - 1$ distances between the query-related point and the rest of the "projections" using the Euclidean norm (i.e., all computations involved a subset of distances $D_2(i, j) = \|Y_i^{\text{query}} - Y_j^{\text{app}}\|_{L_2}$, $i, j = 1, 2, \ldots, N_0$), and performed the ranking of them. The sorted list reflected an order for the database images that enabled the selection of the first T images as the most similar ones, where T is a user-defined parameter controlling the size of the selected list of images returned by the engine (e.g., $T = 10$ refers to the first top 10 images). In the evaluation of the search results, we exploited the fact that images had been assigned to distinct semantic classes. Let N be the number of actually similar images (i.e., $N = 50$ in our case) and CR the number of images among the T selected ones that actually come from the same class (i.e., correct results). The well-known indices Pr and Re were calculated as usual, that is, $Pr = CR/T$ and $Re = CR/N$ [29]. The procedure was repeated for all the N-images in each class and the individual indices were finally averaged (within each class). For the sake of comparison, we reported the two indices for the corresponding search engine employing the machine-vision similarity of Section 3.3.

Figure 9 contrasts the retrieval performance of the two search engines showing the precision index as a function of the number of the selected list, T ($T = 5 : 5 : 50$), for all the $C = 15$ image categories separately. The depicted characteristic symbols used in each corresponding curve are in exact analogy to those used in Figure 6 to visualize the images coming from different categories. It is clearly evident that the cognitive-biplot-based approach outperforms significantly (roughly 40%) the machine-vision one in terms of accuracy in retrieval. An interesting remark is that the "Sunrises" class depicted by the green-star symbol holds the absolute retrieval score of 100% as shown in Figure 9(a). The specific category is well organized and clearly separated in the projection space of Figure 6. In Figure 9(b), the corresponding Pr index shows much lower values (and this is one of the "best-behaving" classes).

Finally, the retrieval performance of the two different schemes, as methods for accessing image databases, was evaluated following the standard procedure of constructing the *precision-versus-recall* diagram.

Based on the depicted curves of Figure 10, it is clear that the perception-dependent scheme outperforms, in terms of retrieval accuracy, the machine-vision one at an average rate of 50%. For example, taking the $T = 10$ most similar images (provided by the second label point in each curve), the cognitive biplot maintains a precision of 0.9, while on the other hand the machine vision reaches approximately 0.375. Moreover, the Pr index follows a stable slope with the increment of the selected list T. One should bear in mind that the retrieval process has been implemented for both approaches in a 2D reduced coordinate space. Specifically, for the machine-vision scheme this was quite far from the optimal condition (this would be expected from the measured E-index which was ~0.73 for the point diagram of Figure 8). Using more dimensions to approximate the space of machine perception, a better performance could be achieved for the corresponding
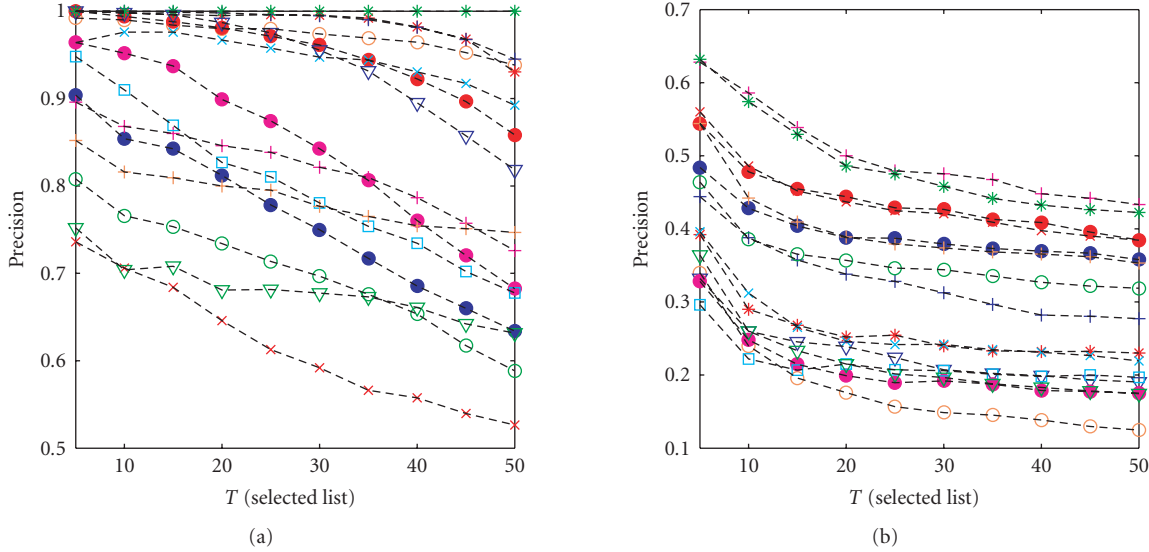
(a)

(b)

FIGURE 9: The Precision index of all the preassigned image categories (following the same class labeling as provided in Figure 6), as a function of the selected list T of images for (a) the cognitive dependent scheme and (b) the machine-vision-based methodology.
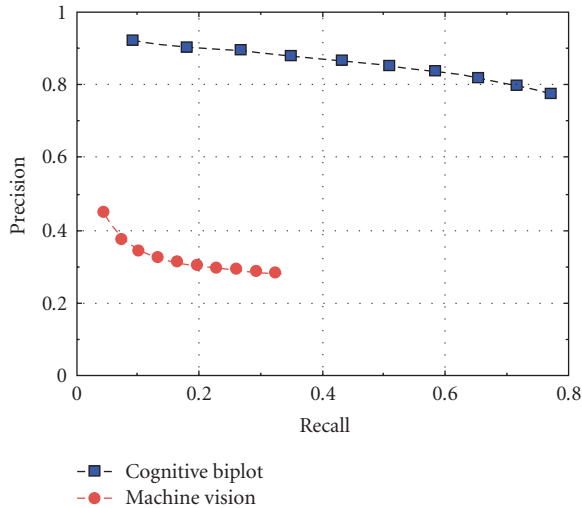

FIGURE 10: Comparison of the retrieval performance via the precision versus recall diagram for the two different methodologies.

search engine. The precision index of the computer-vision-based methodology was slowly rising with the increment of the number of dimensions. By residing in a 3D projection space the Pr index exhibited an increment of $\sim$7.5% (e.g., for $T = 10$, the Pr score was 0.45), in 4D the increment was $\sim$13.5% (e.g., for $T = 10$ the Pr score was 0.51), and so forth. However, its maximum performance (when the number of space dimensions coincided with the number of database images) was $\sim$11% lower than the performance of the cognitive-biplot-based scheme.

## 5. THE PERSPECTIVE: A PRACTICAL ALGORITHM FOR USER-DEPENDENT DATABASE ORGANIZATION

We are concluding by proposing a more realistic algorithm for organizing image databases. The scope of this section is not only to provide an upshot of this work, but also to demonstrate how the presented ideas can be implemented under the pragmatic scenario of an unknown database needed to be explored by an individual user. In such a case, semantically distinct classes will not be available beforehand (even the existence or the exact number of classes would be undefined), and therefore prototyping cannot be realized based on vector-median algorithm.

Since humans are natural experts in analyzing visual information, we expected that a user could declare readily some prototypes when faced with the variation of the image database presented in an orderly manner. Taking into account that perceptual space (at least regarding the perception of a given image database) is low-dimensional, we predicted that it could be recovered even from prototypes selected in a very approximate way. Under this perspective, we experimented as follows.

First, a very small portion of images (7.5%) was selected—via random sampling—from the database. A biplot representation of machine-vision perception was computed using this sample of 100 images (similarities were estimated based on histogram intersection, which is the simplest, fastest, and most popular technique [30]). The corresponding thumbnails were then incorporated giving rise to a display in which the user could locate easily a restricted number (e.g., 5 or 10) of prototypes well suited for the given database. In the utilized application, the user was not only

(a)
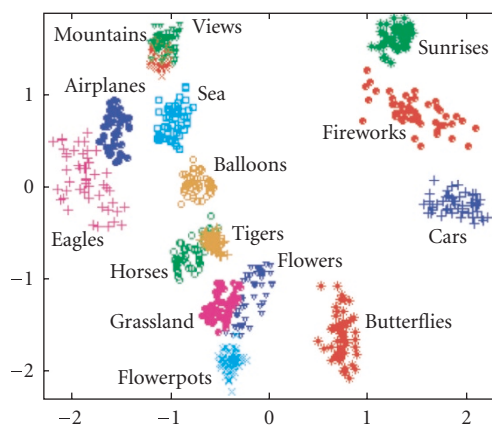
(b)
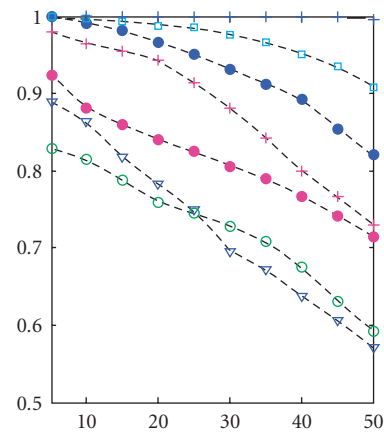
(c)

FIGURE 11: (a) The preliminary biplot (based on machine perception) used for prototyping. Based on the 10 highlighted representative images, a cognitive-biplot approximating user's perceptual space was estimated; (b) the perceptual organization of the whole image database; (c) the precision curves of a search engine build over the discriminative terrain shown in (b).

naive with respect to the database content, but also freely allowed to select the number of prototype images that best reflected his/her intuition. After this selection, the perceptual organization of the whole database proceeded as described in Sections 3.2 and 3.4. Figure 11 demonstrates these steps and includes the precision diagram associated with the new cognitive biplot. In the depicted organization scheme of Figure 11(b) as well as the precision curves presented in Figure 11(c), we have used the same labels (i.e., symbols and colors) for representing the distinct image categories as those

used in all earlier stages of our work, in order to accomplish a straightforward comparison with what has been demonstrated so far.

Despite the encountered approximation (i.e., only ten prototypes were chosen in a "fast mode"), a search engine of high performance was built over the new biplot. Interestingly, some categories exhibit higher retrieval accuracy when compared with the accuracy measured in Figure 9(a) (e.g., "Airplanes" depicted with blue bubbles, "Sea" symbolized using cyan squares, "Eagles" depicted by magenta

crosses), while others exhibit slightly lower retrieval (e.g., "Flowers" represented with the blue triangles). However, we should underline a main difference between Figure 9(a) and Figure 11(c), that is, the latter is based on single-subject similarity evaluation. The results actually confirmed our original predictions and showed that the presented approach deserves further consideration for building efficient querying tools that can readily be adapted to the input database and also get tailored to user's demands.

6. DISCUSSION

Our inspiration to pursue this study was drawn from Edelman's work on visual representation and recognition [11, 15, 16]. The existence of a low-dimensional perceptual space revolving around distinct category prototypes was an appealing consideration, regardless of its neuroscientific significance. The marriage of a standard psychophysical procedure with the theory of biplots was attempted as a means of recovering this space and further exploiting the internalized structure for the visual organization of image databases.

Our experiments with an image database (typical of those used in CBIR research) showed that a directly interpretable approximation of the build-in structure could be achieved and used to predict the database organization regarding user's perception. We then proposed a practical algorithm comprising three consecutive steps: (1) prototyping using a machine-perception biplot (the cognitive stage), (2) MDS-based approximation of perceptual space using the selected prototypes, and (3) embedding of the whole database in the reduced perceptual space with known topography (the discriminative stage). The latest is related with the technique of *anchoring*, a simple technique that represents objects by their distances to a few well-chosen landmarks or vantage points, revealing reduced feature dimensionality and lower computational complexity for image similarity comparisons [31].

The novelty of the proposed algorithm is that it goes beyond the current machine-vision schemes in which the feature selection and/or the definition of similarity measure are designed so as to match the user's perception. Here, the user expresses directly similarity judgments for a restricted number of images. The rest perceptual comparisons are predicted via "extrapolation" based on the structure underlying the biplot computations. It is worth mentioning that once the database has been organized, the search for similar images given a previously unseen query image can be cast in a very efficient way. It is only a visual comparison with $\sim$10 prototypes and the application of appending technique (in which simple matrix operations are encountered) that is required.

Our efforts are currently concentrated in speeding up the stage of database organization, in which each image is contrasted with the selected prototypes. This stage can take considerable time in the case of huge databases. This can be alleviated in two ways. First, by modifying the interactive environment (Figure 5) so as perceptual similarity can be cast in a more convenient way. For example, Borgne et al. [32] organized a psychophysical experiment for image categorization, so as to determine a human perception space for identifying

perceptual categories. In their work, one image was presented along with eight randomly selected images and the subject was asked to quickly select the most similar image to the reference one, as well as to indicate its proximity among four different levels. Some straightforward extensions for testing the robustness of our organization methodology could be, for example, by randomly selecting a prototype inside each predetermined class, or even systematically omitting a prototype/class when constructing the biplot. The second methodological way to operate is by modifying the corresponding algorithmic step so as to work at multiple resolution levels organized in a pyramidal manner.

We believe that the ideas discussed in this paper deserve further consideration, since the cognitive discriminative biplots can serve as extremely powerful interfaces for user-friendly database browsing. Embedded in simple GUI environment, they can facilitate flexible navigation in image databases. Of extreme interest is their user-adaptive character, as well as the potential for becoming tailored to application-specific image databases. Their operation by domain experts could be equivalent with knowledge-based semantic indexing and browsing. An important issue that needs further consideration is the asymmetric nature of the originally estimated distance matrix **D**. In applications where human similarity judgment is investigated, the resulted dissimilarity matrix is nonsymmetric due to the fact that perceptual similarity between two prototype images P_i and P_j is estimated by interchanging each time the reference image [32]. The sequential symmetric operations that take place cuts off some useful semantic information of nonsymmetries, since nonmetric data are forcefully embedded into a vector space [33, 34]. The specific matter is of crucial importance and remains open for further investigation and future research directions.

Finally, it should be mentioned that regarding the engagement of users in expressing perceptual similarities, a similar in-spirit approach has been carried out recently [5, 6, 8, 32]. We consider this direction as a potential alternative to relevance-feedback and expect to attract more researchers in the near future.

APPENDIX

Let $\{X_i\}_{i=1:m}$ and $\{Y_i\}_{i=1:n}$ be two multidimensional point samples of independent random vectors with distributions F_x and F_y, respectively. The classical two-sample problem tests the hypothesis H_0 whether they are coming from the same multivariate distribution:

$$H_0 : F_x(i) = F_y(i), \quad \text{for every } i \in R^p, \qquad \text{(A.1)}$$

where the distribution functions F_x and F_y are unknown. In the multivariate case, the *Wald-Wolfowitz test* (*WW-test*) is built as follows. At first, the sample identity of each point is not encountered and the overall MST of the two separate distributional samples is constructed. The edges for which the defining nodes originate from different samples are afterwards removed. Then, based on the sample identities of the points, a test statistic R is computed. R is 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. Rejection of H_0 is for small values of R. The null distribution of the test statistic has been derived, based on combinatorial analysis [22].

Now, consider samples of size m and n, respectively, from distributions F_x and F_y, both defined in $\mathbf{R^P}$. Let $N = m + n$, let C be the number of edge pairs of MST sharing a common node, and let d_i be the degree of the ith node. Under $\mathbf{H_0}$, the mean and variance of R can be computed as follows:

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

It has been shown that the quantity

$$W = \frac{R - E[R]}{\sqrt{\mathrm{Var}[R]}} \qquad (A.3)$$

approaches (asymptotically) the standard normal distribution while the mean $E[R]$ and variance $\mathrm{Var}[R|C]$ of R depend on the sizes m and n of the two point-samples [22]. This enables the computation of the *significance level* (and *p-value*) for the acceptance of the hypothesis $\mathbf{H_0}$.

ACKNOWLEDGMENTS

REFERENCES

[1] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: a power tool for interactive content-based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 644–655, 1998.

[2] Z. Su, H. Zhang, S. Li, and S. Ma, "Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 924–937, 2003.

[3] E. Chang and S. Tong, "SVMActive - support vector machine active learning for image retrieval," UCSB Technical Report, University of California, Santa Barbara, Santa Barbara, Calif, USA, November 2001.

[4] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.

[5] B. E. Rogowitz, T. Frese, J. R. Smith, C. A. Bouman, and E. B. Kalin, "Perceptual image similarity experiments," in *Human Vision and Electronic Imaging III*, vol. 3299 of *Proceedings of SPIE*, pp. 576–590, San Jose, Calif, USA, January 1998.

[6] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N. Yianilos, "The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 20–37, 2000.

[7] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: semantics-sensitive integrated matching for picture libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 9, pp. 947–963, 2001.

[8] A. Mojsilović, J. Gomes, and B. Rogowitz, "Semantic-friendly indexing and quering of images based on the extraction of the objective semantic cues," *International Journal of Computer Vision*, vol. 56, no. 1-2, pp. 79–107, 2004.

[9] J. Bertin, *Semiology of Graphics*, University of Wisconsin Press, Madison, Wis, USA, 1983.

[10] S. Ullman, *High-Level Vision: Object Recognition and Visual Cognition*, MIT Press, Cambridge, Mass, USA, 1996.

[11] S. Edelman, *Representation and Recognition in Vision*, MIT Press, Cambridge, Mass, USA, 1999.

[12] H. S. Seung and D. D. Lee, "The manifold ways of perception," *Science*, vol. 290, no. 5500, pp. 2268–2269, 2000.

[13] E. T. Rolls and G. Deco, *Computational Neuroscience of Vision*, Oxford University Press, New York, NY, USA, 2002.

[14] L. M. Chalupa and J. S. Werner, *The Visual Neuroscience*, MIT Press, Cambridge, Mass, USA, 2004.

[15] S. Edelman, "Representation is representation of similarities," *Behavioral and Brain Sciences*, vol. 21, no. 4, pp. 449–498, 1998.

[16] S. Duvdevani-Bar and S. Edelman, "Visual recognition and categorization on the basis of similarities to multiple class prototypes," *International Journal of Computer Vision*, vol. 33, no. 3, pp. 201–228, 1999.

[17] J. C. Gower and D. J. Hand, *Biplots*, Chapman & Hall, London, UK, 1996.

[18] J. B. Kruskal and M. Wish, *Multidimensional Scaling*, Sage, Beverly Hills, Calif, USA, 1978.

[19] J. Astola, P. Haavisto, and Y. Neuvo, "Vector median filters," *Proceedings of the IEEE*, vol. 78, no. 4, pp. 678–689, 1990.

[20] C. Theoharatos, N. A. Laskaris, G. Economou, and S. Fotopoulos, "A generic scheme for color image retrieval based on the multivariate Wald-Wolfowitz test," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 808–819, 2005.

[21] J. C. Gower, "Adding a point to vector diagrams in multivariate analysis," *Biometrika*, vol. 55, no. 3, pp. 582–585, 1968.

[22] J. H. Friedman and L. C. Rafsky, "Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests," *Annals of Statistics*, vol. 7, no. 4, pp. 697–717, 1979.

[23] C. Theoharatos, N. A. Laskaris, G. Economou, and S. Fotopoulos, "Combining self-organizing neural nets with multivariate statistics for efficient color image retrieval," *Computer Vision and Image Understanding*, vol. 102, no. 3, pp. 250–258, 2006.

[24] A. Tversky, "Features of similarity," *Psychological Review*, vol. 84, no. 4, pp. 327–352, 1977.

[25] V. Mezaris, H. Doulaverakis, R. Medina Beltran de Otalora, S. Herrmann, I. Kompatsiaris, and M. G. Strintzis, "A testbed for region-based image retrieval using multiple segmentation algorithms and the MPEG-7 eXperimentation model: the schema reference system," in *Proceedings of the 3rd International Conference on Image and Video Retrieval (CIVR '04)*, pp. 592–600, Dublin, Ireland, July 2004.

[26] N. A. Laskaris and A. A. Ioannides, "Semantic geodesic maps: a unifying geometrical approach for studying the structure and dynamics of single trial evoked responses," *Clinical Neurophysiology*, vol. 113, no. 8, pp. 1209–1226, 2002.

[27] C. T. Zahn, "Graph-theoretical methods for detecting and describing gestalt clusters," *IEEE Transactions on Computers*, vol. 20, no. 1, pp. 68–86, 1971.

[28] R. C. Prim, "Shortest connection networks and some generalizations," *Bell Systems Technology Journal*, vol. 36, pp. 1389–1401, 1957.

[29] V. Castelli and L. D. Bergman, *Image Databases: Search and Retrieval of Digital Imagery*, John Wiley & Sons, New York, NY, USA, 2002.

[30] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[31] A. Natsev and J. R. Smith, "A study of image retrieval by anchoring," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '02)*, vol. 2, pp. 421–424, Lausanne, Switzerland, August 2002.

[32] H. Le Borgne, N. Guyader, A. Guérin-Dugué, and J. Herault, "Classification of images: ICA filters vs human perception," in *Proceedings of 7th International Symposium on Signal Processing and Its Applications (ISSPA '03)*, vol. 2, pp. 251–254, Paris, France, July 2003.

[33] J. Laub and K.-R. Müller, "Feature discovery in non-metric pairwise data," *Journal of Machine Learning Research*, vol. 5, pp. 801–818, 2004.

[34] J. Laub, V. Roth, J. M. Buhmann, and K.-R. Müller, "On the information and representation of non-Euclidean pairwise data," *Pattern Recognition*, vol. 39, no. 10, pp. 1815–1826, 2006.

Christos Theoharatos received the B.S. degree in physics in 1998, the M.S. degree in electronics and computer science in 2001 and the Ph.D. degree in image processing and multimedia retrieval in 2006, all from the University of Patras (UoP), Greece. He has actively participated in several national research projects and is currently working as a Postdoctoral Researcher at the Electronics Laboratory (ELLAB), Electronics and Computer Division, Department of Physics, UoP. Since the academic year 2002, he has been working as Tutor at the degree of Lecturer in the Department of Electrical Engineering of the Technological Institute of Patras. His main research interests include pattern recognition, multimedia databases, image processing and computer vision, data mining, and graph theory.

Nikolaos A. Laskaris received both the M.S. degree in medical physics (1995) and Ph.D. degree in biomedical signal processing (1998) from Patras University, Greece. He joined the Laboratory for Human Brain Dynamics in the Brain Science Institute (BSI) of RIKEN in Japan, for four years (1999–2003), where he worked mostly as a Neuroengineer. He received a grant from the Greek GSRT (ENTER Program) for working as an Associate Researcher at Electronics Laboratory of Physics Department in Patras University (2003-2004), where he was involved in the development of neuromorphic signal/image processing techniques. Currently, he is a Faculty Member of Informatics Department in AUTH. His current research interests include computational intelligence, machine vision, soft computing, data mining, nonlinear dynamics, and their applications in biomedicine and neuroscience.

George Economou received the B.S. degree in physics from the University of Patras (UoP), Greece in 1976, the M.S. degree in microwaves and modern optics from University College London in 1978, and the Ph.D. degree in fiber-optic-sensor-systems from the University of Patras in 1989. He is currently an Associate Professor at Electronics Laboratory (ELLAB), Department of Physics, UoP, where he teaches at both undergraduate and postgraduate levels. He has published papers on nonlinear signal and image processing, fuzzy image processing, multimedia databases, data mining, and fiber-optic sensors. He has also served as a referee for many journals, conferences, and workshops. His main research interests include signal and image processing, computer vision, pattern recognition, and optical signal processing.

Spiros Fotopoulos is Professor at the Department of Physics of the University of Patras. He is Director of the M.S. course on electronics and information processing. He is working in the digital signal and image processing area. His research activities include nonlinear digital filters, multichannel filters, fuzzy image processing, neural-networks techniques, graph-theoretic approaches, and applications to satellite images and biomedical signals.