PERGAMON

Information Processing and Management 39 (2003) 251–266

INFORMATION PROCESSING & MANAGEMENT

www.elsevier.com/locate/infoproman

Modelling subjectivity in visual perception of orientation for image retrieval

D. Sánchez *, J. Chamorro-Martínez, M.A. Vila

Department of Computer Science and Artificial Intelligence, University of Granada, Avda. Andalucia 38, 18071 Granada, Spain

Abstract

In this paper we combine computer vision and data mining techniques to model high-level concepts for image retrieval, on the basis of basic perceptual features of the human visual system. High-level concepts related to these features are learned and represented by means of a set of fuzzy association rules. The concepts so acquired can be used for image retrieval with the advantage that it is not needed to provide an image as a query. Instead, a query is formulated by using the labels that identify the learned concepts as search terms, and the retrieval process calculates the relevance of an image to the query by an inference mechanism. An additional feature of our methodology is that it can capture user's subjectivity. For that purpose, fuzzy sets theory is employed to measure user's assessments about the fulfillment of a concept by an image.

Keywords: Image retrieval; Visual perception; Fuzzy association rules; Image semantics

1. Introduction

Nowadays, the amount of multimedia libraries all over the world is growing in an impressive way. Consequently, there is a need of efficient techniques for storage, indexing and retrieval of this kind of information, that has motivated an increasing research effort in this area (Chang, 1997; Chang et al., 1999). However, the gap between signal-based descriptions provided by computer vision techniques and conceptual descriptions required to answer library user's queries is far from being filled.

* Corresponding author. Tel.: +34-958-246397; fax: +34-958-243317.

E-mail addresses: daniel@decsai.ugr.es (D. Sánchez), jesus@decsai.ugr.es (J. Chamorro-Martínez), vila@decsai.ugr.es (M.A. Vila).

0306-4573/03/$ - see front matter
PII: S0306-4573(02)00050-X

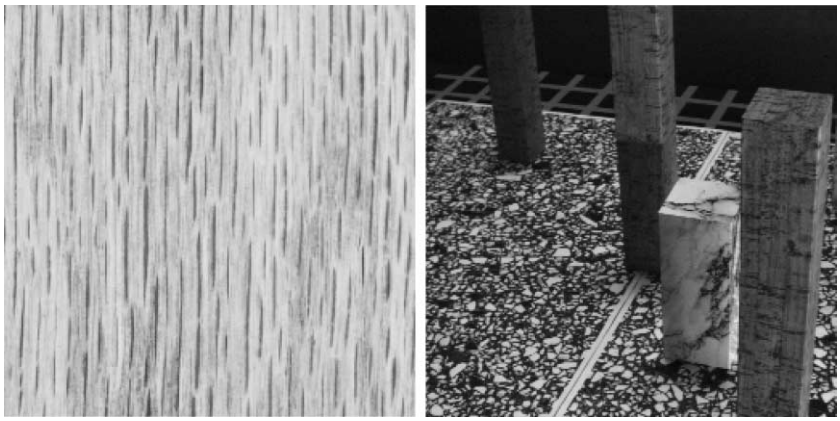

Fig. 1. Two images containing vertical information.

Current image retrieval systems are based mainly in low-level image descriptors, usually called *features*, such as color and shape (Kankahalli, Mehtre, & Huang, 1999; Fuertes, Lucena, de la Blanca, & Chamorro-Martínez, 2001; Gevers & Smeulders, 2000; Korfhage, 1997). In these systems, images are represented by vectors of features, queries are defined as an image or sketch, and the matching between them is performed by measuring the similarity of the corresponding vectors (Del Bimbo, De Marsico, Levialdi, & Peritore, 1998; Del Bimbo & Pala, 1997; Flickner et al., 1995; Pentland, Picard, & Sclaroff, 1996; Smith & Chang, 1996). However, it is widely agreed that this scheme does not address many access needs of users. Indeed, in many occasions the user's idea of similarity cannot be measured in terms of low-level features but on the basis of higher level concepts. For example, consider the images of Fig. 1. These images are similar in that they verify a high-level concept: they contain *vertical information*. This fact can be perceived by humans. Nevertheless, the usual features of color, shape and the like take very different values for both images, so they are seen as different on the basis of similarity metrics. A direct consequence is that we cannot use images or sketches to query for images containing vertical information, or high-level concepts in general.

Another important issue here is subjectivity. Different people can differ in their perception of high-level concepts. Hence, a query such as "give me images containing horizontal information" should take into account the user's perception of orientation. Consider for example the images in Fig. 2. In our experiments we asked three people to provide a verticality degree for those images. The fact is that we have collected very different answers to this question.

Several authors have proposed approaches to model similarity in terms of high-level concepts. One approach is to use metadata attached to the image such as captions, tags and even textual descriptions (Chang et al., 1999; Korfhage, 1997). However, this approach make little use of images, and the generation of such metadata is a hard task that must be performed by humans. Moreover, metadata reflect only the subjectivity of the person who made it.

Another approach is to perform a segmentation of the image, and to store the features for each region (Medasani & Krishnapuram, 1999). Sometimes the vector dimensionality is reduced by using techniques from information retrieval such as SVD (Dumais, 1991), see for example (Fung & Loe, 1999). A similarity metric keeps being the way to match. Hence, though improving the

Fig. 2. Two images where observers have manifested significant differences in their subjective appreciation of verticality.

performance of current systems for image retrieval, this technique has the same problems mentioned before. For example, there are few regions (if any) in images of Fig. 1 whose features were similar.

In this context, we are concerned with learning semantic concepts from images. Humans are able to perform this task, so it seems natural to try to acquire these concepts on the basis of the basic features humans can perceive. Hence, our first attempt will be to learn basic concepts related to the human visual system.

Human vision and visual perception are very complex processes (Hubel & Wiesel, 1959, 1965; Marr, 1982). Since the image is formed on the retina, to our brain interprets "what we see" (in a region called "visual cortex"), the visual information is analyzed at several levels. For this purpose, our visual system uses the responses of a set of cells that are selective to specific visual properties. Thus, there are cells characterized by its selectivity to orientation and size (area V2 in the visual cortex (De Valois, Yund, & Hepler, 1982b; Hubel & Livingstone, 1990; Hubel & Wiesel, 1962)), to color and shape (area V4 (Livingstone & Hubel, 1984)) and to motion (area V5 (Cleland, Dubin, & Levick, 1971)). Therefore, the analysis performed by our brain starts from a study based on simple features (like size or orientation) that afterwards will generate, in the final stages of the interpretation, the association of perceptual properties to sensations like color, motion, size, orientation or shape.

The objective of this work is to characterize images on the basis of these visual features, and to learn basic concepts related to orientation, such as the presence of horizontal, vertical or diagonal information. Specifically, we want to provide a model of such concepts able to represent the subjectivity of the user. In our case, this model consists of a set of rules relating low-level features to concepts. Uncertainty in the rules is used to represent subjectivity.

Modelling such concepts is a powerful tool for image retrieval. They allow to compare images in terms of the semantic information they contain, taking into account the subjectivity of the user. Moreover concepts can be used to formulate a query, hence avoiding the necessity to provide an image or sketch.

2. Methodology

Though in this work we shall focus on learning concepts related to orientation, the methodology we present here is intended to learn concepts related to any other basic feature of the human visual system. Hence, we shall talk of *concepts* and *features* in general. Let us remark that concepts and features are assumed to be related in some way. For example, if we want to learn concepts about orientation, we should use the basic perceptual features for orientation of the human visual system.

Concepts are to be learned from an image database I and user valuations of the images. Let F be a set of features and let C be a set of concepts.

Definition 2.1. Given an image $i \in I$, the representation of i, $r(i)$, is a fuzzy subset of $F \cup C$.

By definition $r(i)$ is a *fuzzy transaction* (Delgado, Marín, Sánchez, & Vila, 2001), characterized by a function

$$\mu_{r(i)} : F \cup C \to [0, 1] \tag{1}$$

where $\mu_{r(i)}(o)$ is the degree to which the image i verifies o. We shall note $r^F(i) = r(i) \cap F$ and $r^C(i) = r(i) \cap C$.

Definition 2.2. The representation of an image database I is the multiset

$$R(I) = \{r(i) | i \in I\} \tag{2}$$

Notice that $R(I)$ is a multiset since two different images in I could have the same representation. By definition, $R(I)$ is an FT-set, i.e., a crisp multiset of fuzzy transactions (Delgado et al., 2001). Also we shall note

$$R^F(I) = \{r^F(i) | i \in I\} \tag{3}$$

$$R^C(I) = \{r^C(i) | i \in I\} \tag{4}$$

An FT-set can be represented in a table. For example, let $I = \{i_1, i_2, i_3\}$ be a set of three images, and let $F = \{f_1, f_2, f_3\}$ and $C = \{c_1\}$. Table 1 represents an imaginary representation for $R(I)$. The representation of i_1 is the fuzzy set $r(i_1) = 0.6/f_2 + 1/f_3 + 1/c_1$, meaning that the image i_1 verify the features f_2 and f_3 with degrees 0.6 and 1 respectively, and the concept c_1 with degree 1. Also, $r^F(i_1) = 0.6/f_2 + 1/f_3$ and $r^C(i_1) = 1/c_1$. Finally, $R^F(I)$ and $R^C(I)$ are represented by the columns 1–3 and 4 respectively.

Table 1
Three fuzzy transactions

	f_1	f_2	f_3	c_1
$r(i_1)$	0	0.6	1	1
$r(i_2)$	0	1	0	1
$r(i_3)$	1	0.4	0.75	0.1

In practice, $R^C(I)$ will be provided by an user, while $R^F(I)$ will be obtained by image analysis. In Section 3 we describe how we analyze features related to the perception of orientation, to be stored in $R^F(I)$.

2.1. Learning high-level concepts

As we shall see, there exists independence between the basic perceptual features that are to be used by our model (Section 3.2.1). This fact has suggested us to use association rule mining to discover local associations between perceptual features and concepts. A set of association rules relating features to a given concept c_k can be interpreted as a model for c_k, see (Blanco, Martín-Bautista, Sánchez, & Vila, 2000; Delgado, Martín-Bautista, Sánchez, & Vila, 2000a).

2.1.1. Association rules

Association rules (Agrawal, Imielinski, & Swami, 1993) are one of the best studied data mining techniques. It is assumed that the basic object of our interest is an *item*, and that data appear in the form of sets of items called *transactions*. Association rules are "implications" that relate the presence of items in transactions. The classical example are the rules extracted from the content of market baskets. Items are things we can buy in a market, and transactions are market basket containing several items. Association rules relate the presence of items in the same basket, for example "every basket that contains bread contains butter", usually noted *bread* $\Rightarrow$ *butter*.

In our case, items are features and concepts, and the set of transactions is $R(I)$, i.e., the set of representations for a given image database I. Specifically, we deal with fuzzy transactions because items and concepts are to a certain degree in the representation of a given image, and hence $R(I)$ is an FT-set. Association rules in FT-sets are called *fuzzy association rules* (Delgado et al., 2001).

We are going to learn concepts by mining for fuzzy association rules in $R(I)$, but with some restrictions in accordance with our problem. Due to their nature, we assume there exists independence between the presence of perceptual features in an image. Hence, we are interested only in rules of the form $f_j \Rightarrow c_k$, with $f_j \in F$ and $c_k \in C$. [1]

2.1.2. Assessing fuzzy association rules

Association rules are assessed by examining their support and accuracy, both based on the support of an itemset (a set of items). The support of an itemset is the probability that the itemset appears in a transaction of R. The support of the association rule $A \Rightarrow B$ in R is $\mathrm{Supp}(A \Rightarrow B) = \mathrm{Supp}(A \cup B)$. Several measures of accuracy have been proposed, the classical one being confidence and defined as the probability of B conditional to A.

Shortliffe and Buchanan's certainty factors (Shortliffe & Buchanan, 1975) have recently been proposed as an alternative to confidence, thereby solving the drawbacks which this has (Berzal, Blanco, Sánchez, & Vila, 2001). The certainty factor of a rule is defined as

[1] To confirm this, we have performed some experiments involving rules with more than one feature in the antecedent, and we have found that all rules of this kind are subsumed by rules with one single feature in the antecedent, as expected.

$$\mathrm{CF}(A \Rightarrow B) = \begin{cases} \frac{\mathrm{conf}(A \Rightarrow B) - \mathrm{supp}(B)}{1 - \mathrm{supp}(B)} & \text{when } \mathrm{conf}(A \Rightarrow B) > \mathrm{supp}(B) \\ \frac{\mathrm{conf}(A \Rightarrow B) - \mathrm{supp}(B)}{\mathrm{supp}(B)} & \text{when } \mathrm{conf}(A \Rightarrow B) < \mathrm{supp}(B) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Certainty factors take values in $[-1, 1]$, indicating the extent to which our belief that the consequent is true varies when the antecedent is also true. A value 1 for $\mathrm{CF}(A \Rightarrow B)$ means maximum increment (i.e., when A is true then B is true). A value -1 means maximum decrement. Among the other properties, certainty factors are an independence test since when A and B are statistically independent, the certainty factor of the rule $A \Rightarrow B$ is 0 (this also holds for the rule $B \Rightarrow A$). Consequently, we can measure the strength of implication between the activation of a given feature and the fulfillment of a given concept by mining for association rules in the database and checking the certainty factor of the rules.

For this purpose we employ the algorithm proposed in Delgado et al. (2001). The calculation of confidence and support in the presence of fuzzy degrees is based on the evaluation of quantified sentences, specifically the method GD (Delgado, Sánchez, & Vila, 2000b). Using this method, the support of an itemset $\mathbb{I}_0$ is the evaluation of the sentence

$$M \text{ of the } R(I) \text{ are } \mathbb{I}_0 \tag{6}$$

where M is a fuzzy relative quantifier (Zadeh, 1983) defined as $M(x) = x$ on [0,1]. The confidence of a rule $A \Rightarrow B$ is the evaluation of

$$M \text{ of the } A \text{ are } B \tag{7}$$

and the certainty factor is calculated from support and confidence by using (5). Among other properties, this method generalizes the measures of support, confidence and certainty factor for crisp data.

Finally, we remark that since we want the search to be exhaustive, we do not employ any support threshold to bound the search. This do not cause any computational problems since we are interested in rules with only one item in the antecedent and the consequent.

2.2. Using semantic concepts for image retrieval

From the learning step we obtain a set of rules for each perceptual concept. Given a concept $c_k \in C$ and a set of images I we note $\mathrm{Rul}(I, c_k)$ the corresponding set of rules. Each rule in $\mathrm{Rul}(I, c_k)$ takes the form $f_j \Rightarrow c_k$ $\mathrm{CF} = x$, with $f_j \in F$. Let us remark that we are only interested in rules with $\mathrm{CF} \in (0, 1]$, since this means that the presence of the feature f_j in an image increase our certainty that the image fulfills the concept. Hence, $\mathrm{Rul}(I, c_k)$ contains only those rules with $\mathrm{CF} > 0$.

Once the model is obtained, we can use it to perform image retrieval in the following way. Suppose we are given an image database I, and we want to retrieve those images in I that fulfill the concept c_k. Let $\mathrm{Cer}(i, c_k)$ be the certainty that image i verifies the concept c_k. Initially, we consider that each image verifies c_k with a certainty factor of $\mathrm{Cer}(i, c_k) = 0$, meaning ignorance. Then we analyze each image i to obtain $r^F(i)$, and we apply the rules $f_j \Rightarrow c_k$ in $\mathrm{Rul}(I, c_k)$ such that $\mu_{r^F(i)}(f_j) > 0$ (i.e. those rules whose antecedent is a feature appearing in i). The application of the rule $f_j \Rightarrow c_k$ yields an increment of our certainty that i verifies c_k. Specifically, let α be the certainty

that i verifies c_k *before* the rule is applied, and let $\delta = \mu_{r^F(i)}(f_j)\mathrm{CF}(f_j \Rightarrow c_k)$. The certainty that i verifies c_k *after* the rule is applied is (Shortliffe & Buchanan, 1975)

$$\mathrm{Cer}(i, c_k) = \alpha + \delta(1 - \alpha) \tag{8}$$

Notice that the process just described is a particular case of the usual inference procedure employed in rule-based expert systems that manage uncertainty by means of certainty factors, like MYCIN (Shortliffe, 1976).

At the end of this process, we have a set of values $\mathrm{Cer}(i, c_k)$ for every $i \in I$. The final set of retrieved images is obtained by one of these procedures:

- Using a threshold minCer provided by the user. This way, every image i with $\mathrm{Cer}(i, c_k) \geqslant$ minCer is retrieved.
- Performing a clustering of the images in I in terms of the values $\mathrm{Cer}(i, c_k)$. The idea is to obtain two clusters, corresponding to retrieved and not retrieved images. This procedure has the advantage that the threshold is computed in an automatic way.

Finally, let us remark that if we store $r^F(i)$ with each image the retrieval process gets faster, because we don't have to calculate it every time a query is performed. This modification does not affect significantly the storage space, since the space needed to store $r^F(i)$ is usually much smaller than for i.

3. Extraction of perceptual features for orientation

In this section, a scheme to represent visual orientation is presented. In order to code "what the observer perceives about orientation", the image analysis will be inspired in features of the human visual system.

Fig. 3 shows a general diagram describing how the data flows in this scheme. In a first stage, the original image is transformed into a spatial frequency representation using the Fourier transform (Fig. 3(B), Section 3.1). Afterwards, a multichannel analysis of the frequency representation is performed in order to extract information about size and orientation (Fig. 3(C) and (D), Section 3.2). For each "channel" (each coloured region in Fig. 3(C)), a degree of "activity" is calculated; using this level of activity, a selection of relevant orientation is performed (Fig. 3(E), Section 3.2.1). Finally, a feature vector which reflects the state of activity at each orientation is generated (Fig. 3(F), Section 3.3). In the following subsections, each of the aforementioned steps will be explained in detail.

3.1. Spatial frequency representation

In the first stage of the image analysis, a Fourier transform is applied in order to obtain a spatial frequency image representation (Fig. 3(B)). This change of domain is based on several physiological studies which suggest that the visual information might be analyzed in terms of Fourier spatial frequency components (Blakemore & Campbell, 1969; Maffei & Fiorentini, 1973);

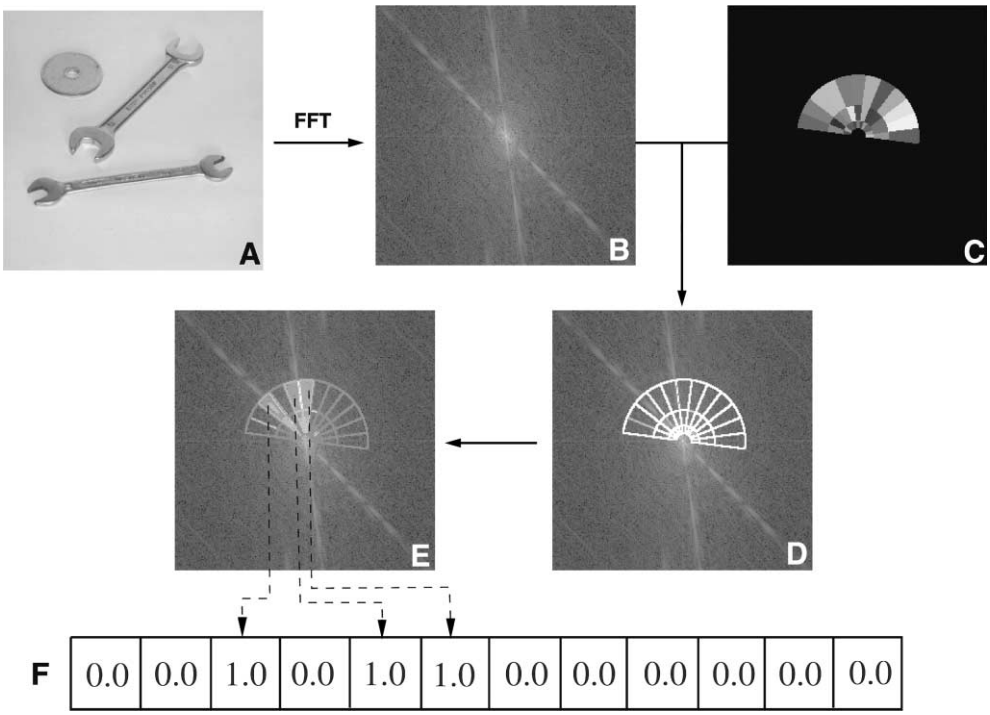

Fig. 3. General diagram showing how an image is analyzed in our model. (A) Original image, (B) spatial frequency representation, (C) set of channels, (D) superposition of B and C, (E) activated orientations, (F) feature vector.

Thus, rather than a representation of the luminance at each point (Fig. 3(A)), we use a representation of the information at each spatial frequency (Fig. 3(B)).

Let i be an image and $i(m,n)$ the grey level intensity at the point (m,n) of i, with $0 \leqslant n \leqslant N-1$ and $0 \leqslant m \leqslant M-1$. Its discrete Fourier transform is defined as a complex function $\Gamma_i(v,v)$ given by the expression (Press, Teukolsky, Vetterling, & Flannery, 1992)

$$\Gamma_i(v,v) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} i(n,m) \mathrm{e}^{-2\pi \mathrm{i}((nv/N)+(mv/M))} \tag{9}$$

with $-((N/2)-1) \leqslant v \leqslant (N/2)-1$ and $-((M/2)-1) \leqslant v \leqslant (M/2)-1$.

This transform converts the input image into a representation based on complex numbers which describes the original image frequency components. The magnitude of these complexes is usually employed to describe the image (Campbell & Robson, 1967). Fig. 3(B) shows the magnitude spectrum for the original image, where each point corresponds to a frequency component (this spectrum has been logarithmically rescaled to show significant frequencies). The situation of a component inside the spatial frequency representation give us information about the orientation of the original patterns: the angle of the component relative to the center is related to the orientation at which these spatial components are perceived (the main spread of the spectrum is orthogonal to the directionality of the original pattern).

3.2. Multichannel analysis

Physiological investigations have shown that the analysis of the frequency information (Fig. 3(B)) is performed by a series of orientation and spatial frequency selective channels (Fig. 3(C)) (Campbell, Cooper, & Enroth-Cugell, 1969; De Valois, Albrecht, & Thorell, 1982a; De Valois et al., 1982b; Graham & Nachmias, 1971; Movshon, Adelson, Gizzi, & Newsome, 1984; Sachs, Nachmias, & Robson, 1971). In this context, a channel can be interpreted as a set of frequency components (points (v, v) in the Fourier domain) that respond with high level of activity only to patterns of a specific orientation and size. In the case of orientations, that means there are separate sets of "frequency analyzers" corresponding to different orientations (for example, vertical and horizontal frequencies are detected by their own and separate sets of channels).

On the basis of the aforementioned property, several computational vision models perform a partition of the Fourier spectrum in order to analyze the responses of different channels (Daugman, 1988; Fdez-Vidal, Rodríguez-Sánchez, Martínez-Baena, & Chamorro-Martinez, 2000; Watson, 1987). In our approach, we propose the use of the partition showed in Fig. 3(C). In this multichannel organization, the spatial frequency plane is divided into 12 different orientations and, for each orientation, the radial axis is divided into three one-octave bandwidth intervals.

In the remainder of this paper, we will note $\{\varsigma_{\rho,\theta}\}_{\rho=1\ldots3,\theta=1\ldots12}$ the set of channels used in our model, where ρ and θ index the radial axis and the orientation respectively.

3.2.1. Channel activity

Graham and Nachmias (1971) showed the independence between channels: activity in one channel do not affect activity in other channels. In this context, the firing of a given channel indicates the presence of information at a particularly size and orientation, but this firing is not affected by the presence of other kind of patterns.

To measure the activity, the magnitude of the frequency components associated to each channel will be used. Given a channel $\varsigma_{\rho,\theta}$, the level of activity is given by the equation

$$w(\varsigma_{\rho,\theta}) = \frac{1}{\mathrm{Card}(\varsigma_{\rho,\theta})} \sum_{p \in \varsigma_{\rho,\theta}} |p| \tag{10}$$

where the frequency component p is a complex number of the form $a + \mathrm{i}b$, and $|p|$ denotes its magnitude given by $\sqrt{a^2 + b^2}$. In order to obtain a normalized weight, a division by the sum of the same radial axis activities is performed. Thus, the final weight associated to a channel $\varsigma_{\rho,\theta}$ will be given by the equation

$$\widetilde{w}(\varsigma_{\rho,\theta}) = \frac{w(\varsigma_{\rho,\theta})}{\sum_{\alpha=1}^{12} w(\varsigma_{\rho,\alpha})} \tag{11}$$

Notice that $\widetilde{w}(\varsigma_{\rho,\theta}) \in [0, 1]$ and $\sum_{\alpha=1}^{12} \widetilde{w}(\varsigma_{\rho,\alpha}) = 1$. Since the Fourier spectrum of natural images falls off as a function of frequency by a factor of approximately $1/k^2$ (Field, 1987), the normalization should be performed using channels of the same spatial frequency (the same "ring" in Fig. 3(C)). A normalization considering all the channels would give priority to those closer to the centre.

3.2.2. Orientation activity

In order to extract information about the relevant orientations presented in the image, a measure of the activity for a given orientation θ is proposed (Fig. 3(E)). For each orientation, we have three weights that must be pooled in an unique value; since each of these weights represents the activity at a particular range of spatial frequencies of this orientation, and taking into account that we only are interested in measuring the presence of relevant orientations (independently of the size), the maximum of the weights for a particular orientation will be used

$$\widetilde{w}_\theta = \max\left\{\widetilde{w}(\varsigma_{\rho,\theta}),\ \rho = 1,\ldots,3\right\} \tag{12}$$

3.3. Feature vector for an image

The normalized weights obtained by means of (12) are calculated by using the activity levels of one image. The absence of a global upper bound for the activity of channels for a generic image keep us from performing a normalization able to make comparable two activity levels of different images. To allow for this comparison, we have associated an activity state (active or non-active) to every orientation. Though this activity state is obtained from the weights of the image, it allows to compare orientations of different ones.

In order to obtain the state of activity, a clustering over the set of weights $\{\widetilde{w}_\theta\}_{\theta=1\ldots12}$ is performed. More precisely, the subset of active orientations is determined using the k-means clustering algorithm (Jain & Dubes, 1988) with two clusters (labeled active and non-active).

Using this state of activity, a feature vector of size 12 is defined. Each position of the vector, corresponding to an orientation, take value 1.0 or 0.0 meaning activity and non-activity respectively (Fig. 3(F)). In the remainder of this paper we will note f_{o_k} the kth feature of the vector.

Let us remark that the feature vector defined above represents information about the global orientation present in the original image. Moreover, this information has been extracted using a methodology inspired in the human visual system.

4. Results

To learn concepts and to measure user's subjectivity, we have employed a database I containing 135 images. This database is very heterogeneous, as can be seen looking at some of the images in Fig. 4. It includes synthetic and real images of different typologies: textures, biomedical images, indoor scenes, etc.

The objective is to learn four concepts related to orientation. These are horizontal (H), vertical (V), diagonal around 45 degrees (D1) and diagonal around 135 degrees (D2). Hence, our concept set is $C_{\mathrm{O}} = \{\mathrm{H}, \mathrm{V}, \mathrm{D1}, \mathrm{D2}\}$. The feature set is $F_{\mathrm{O}} = \{f_{o_k} | k \in \{1, 12\}\}$, corresponding to the orientations in the feature vector (Section 3.1).

4.1. Collecting user judgements

We have asked eight people to provide a fuzzy relevance degree for every image $i \in I$ and every orientation $c \in C_{\mathrm{O}}$. We have added the subindex $_{P_l}$ to the representation of both images and sets

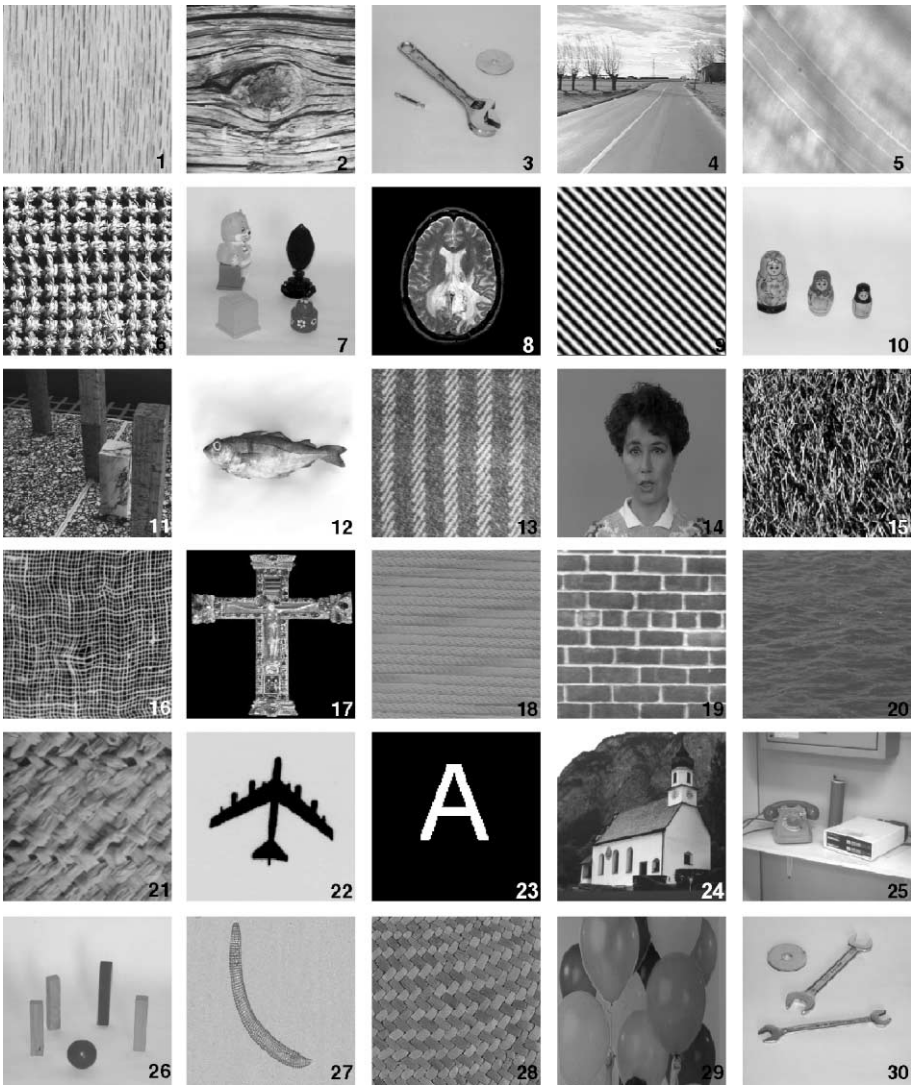

Fig. 4. Some examples of images contained in the database I.

of images for a given individual P_l. For example, we note $r_{P_l}(i)$ and $R_{P_l}(I)$ the representation of an image i and the entire set I, respectively, for individual P_l. The set $R^{F_O}(I)$ was obtained by the procedure explained in Section 3, and it is the same for all individuals, so $R^{F_O}_{P_l}(I) \equiv R^{F_O}(I)$ and $r^{F_O}_{P_l}(i) \equiv r^{F_O}(i)$. The set $R_{P_l}(I)$ is easily calculated since $r_{P_l}(i) = r^{F_O}(i) \cup r^{C_O}_{P_l}(i)$ for all $i \in I$.

Since the final answer to a query is a set of relevant images, perhaps with an associated relevance degree, every user P_l has provided a threshold to calculate the set of relevant images $\mathrm{Rel}(P_l, c)$ for every concept $c \in C_O$. In our experiment, these sets coincide with the sets obtained by clustering images on the basis of relevance degrees for each concept and user. However, this is not always the case.

Table 2
Average similarity between sets of relevant images

H	V	D1	D2
0.70	0.77	0.63	0.67

4.2. Assessing the presence of subjectivity

We performed a first subjectivity test by comparing the sets of relevant images obtained from every user. Similarity between sets has been calculated by using the Jaccard's score, defined for two sets A and B as

$$\mathrm{JS}(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{13}$$

Table 2 shows, for every concept, the average Jaccard's score obtained from the 28 pair comparisons between the eight (one per user) sets of relevant images.

In addition, for each concept we have performed a comparison of the distributions of relevance degrees given by the eight users on the entire database I. In all the cases, the null hypothesis of equality was rejected. Table 3 shows the differences found by the post-hoc least significant difference test. The cell for user P_i and concept c contains the users that differ significantly from P_i in their appreciation of c.

In our opinion, the results above show that both the user valuations and the sets of relevant documents for every concept and user are different enough to assume subjectivity is present.

4.3. Models

We wanted to show an example of the models that can be learned by using our methodology. Fig. 5 shows a graphical representation of the models $\mathrm{Rul}_{P_l}(I,c)$ obtained from the entire database I for some users (for the sake of clarity, only graphics for users P_1–P_4 are shown) and every concept $c \in C_{\mathrm{O}}$. The graphic for the model $\mathrm{Rul}_{P_l}(I,c)$ is a function

$$G_{P_l}^{(I,c)} : F_{\mathrm{O}} \to [0,1] \tag{14}$$

Table 3
Significant differences between users for every concept

	H	V	D1	D2
P_1	P_3, P_5–P_8	P_3, P_4, P_8	P_2, P_4–P_6, P_8	P_4–P_6, P_8
P_2	P_3–P_8	P_4, P_8	P_1, P_5–P_6, P_8	P_4–P_6, P_8
P_3	P_1, P_2, P_5, P_8	P_1, P_4, P_5, P_8	P_4–P_6, P_8	P_4–P_6, P_8
P_4	P_2, P_5, P_8	P_1–P_3, P_7, P_8	P_1, P_3, P_5, P_7	P_1–P_3, P_5–P_7
P_5	P_1–P_4	P_3, P_6–P_8	P_1, P_4, P_6, P_8	P_1–P_4, P_7, P_8
P_6	P_1, P_2, P_8	P_5, P_8	P_1–P_3, P_5, P_7	P_1–P_4, P_7, P_8
P_7	P_1, P_2	P_4, P_5, P_8	P_4–P_6, P_8	P_4–P_6, P_8
P_8	P_1–P_4, P_6	P_1–P_7	P_1–P_3, P_5, P_7	P_1–P_3, P_5–P_7

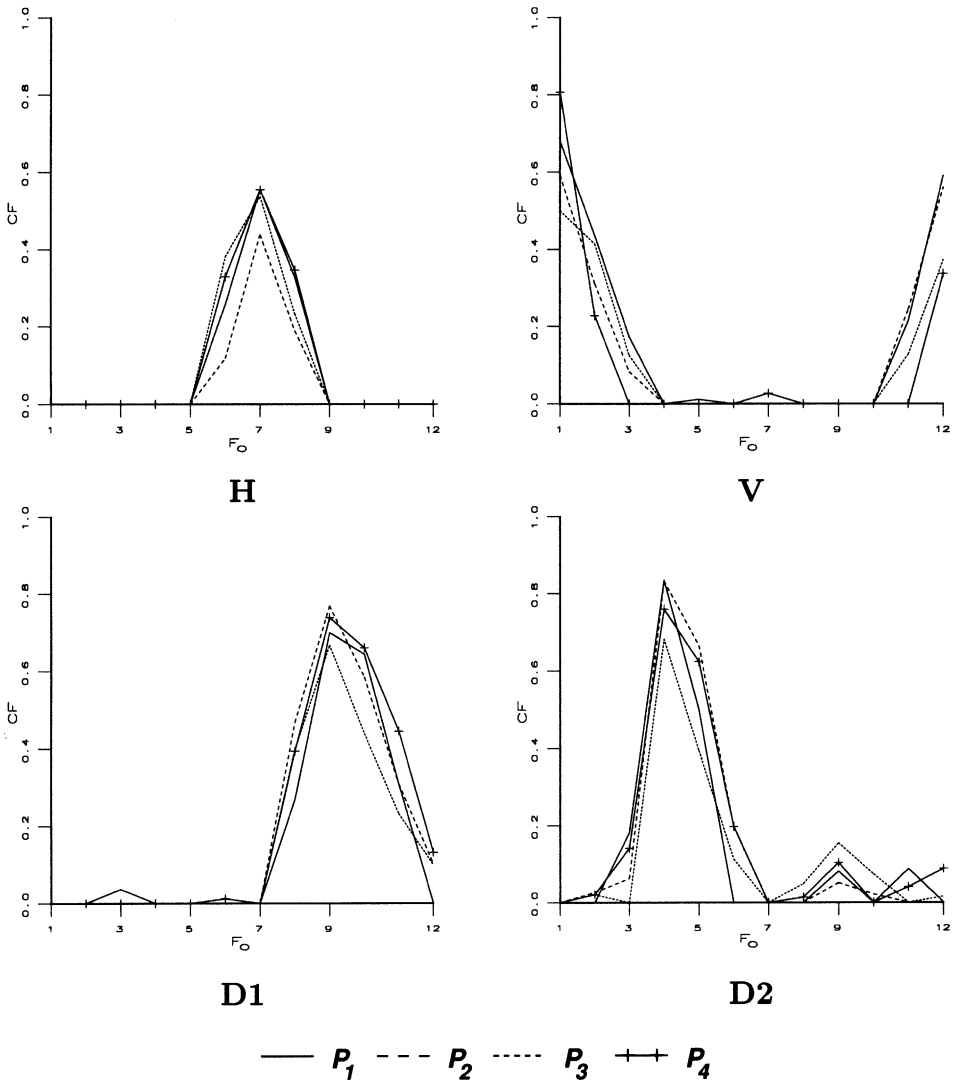

Fig. 5. Graphical representation of the models $\mathrm{Rul}_{P_l}(I,c)$ for every concept c and users P_1–P_4.

such that

$$G_{P_l}^{(I,c)}(f_{o_k}) = \begin{cases} \mathrm{CF}(f_{o_k} \Rightarrow c) & \text{when } (f_{o_k} \Rightarrow c) \in \mathrm{Rul}_{P_l}(I,c) \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

The points have been connected using lines to better appreciate the functions and their differences. Let us remark that, though they are different, all the curves are coherent in some sense with the concepts they are modelling.

Table 4
Average recall/precision for every user and every concept in C_O

	H		V		D1		D2	
	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
P_1	0.97	0.83	0.90	0.71	0.86	0.81	0.86	0.87
P_2	0.96	0.60	0.87	0.68	0.84	0.79	0.88	0.76
P_3	1.00	0.68	0.88	0.58	0.85	0.85	0.82	0.77
P_4	0.94	0.80	0.93	0.72	0.74	0.83	0.71	0.86
P_5	0.85	0.73	0.88	0.79	0.73	0.86	0.72	0.93
P_6	0.85	0.72	0.93	0.69	0.67	0.91	0.77	0.83
P_7	0.97	0.66	0.94	0.72	0.82	0.77	0.86	0.76
P_8	0.84	0.71	0.93	0.76	0.74	0.87	0.80	0.85
Average	0.92	0.71	0.90	0.70	0.78	0.83	0.80	0.82

4.4. Validation: using learned concepts for image retrieval

We have used cross-validation to verify that our methodology is able to learn concepts related to orientation and the subjectivity of the user for image retrieval. We randomly split the database I in two sets I_1 and I_2. For each user P_l and each concept $c \in C_O$ we have obtained two models $\mathrm{Rul}_{P_l}(I_1, c)$ and $\mathrm{Rul}_{P_l}(I_2, c)$. Then we have applied each model for image retrieval in the set of images that was not employed to learn it, and we have measured the precision and recall of the process. Let us remark that the queries we are using in these retrieval processes are not images, but learned concepts. The average precision and recall for each user and each concept is shown in Table 4.

The recall/precision results lead us to some conclusions. First, we think that the performance of the retrieval system is in general satisfactory. Since the set of relevant images was different for every user, as we showed in Section 4.2, this high performance shows that the model is able to capture the subjectivity of the user. Finally, we think that both the feature extraction technique and the learning methodology have proved to be suitable for the problem addressed.

5. Conclusions

We have introduced a general methodology to learn high-level concepts, specifically basic perceptual concepts, on the basis of perceptual features of the human visual system. This methodology make use of fuzzy association rules to describe the relation between features and concepts, representing the strength of such relation by means of certainty factors. The source of knowledge is a representation of an image database containing both features and user judgements about the concepts. These judgements take the form of fuzzy degrees.

A valuable feature of the proposed methodology is that it is able to capture the subjectivity of the user. Moreover, the learned models can be applied for image retrieval, with a feature that, to our knowledge, is unique in image retrieval: we can perform a query without providing an image or sketch, but asking for images verifying a certain concept.

We have tested our techniques for an specific set of concepts related to orientation. For that purpose, we have introduced a mechanism to extract orientation features that is based on physiological studies of visual perception. Our experiments suggest that our proposal is suitable for the task addressed.

A direct application of our techniques is the construction of user profiles for image retrieval. In the immediate future we plan to apply our general methodology to other basic perceptual concepts (like size, motion and color) and to perform experiments in large databases. In addition, on the basis of these basic perceptual concepts, we shall face the definition of more complex high-level concepts, following a process that could be repeated in a hierarchical way.

References

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD Conference* (pp. 207–216).

Berzal, F., Blanco, I., Sánchez, D., & Vila, M. (2001). A new framework to assess association rules. In F. Hoffmann (Ed.), *Advances in intelligent data analysis. Fourth international symposium, IDA'01*. Springer-Verlag (pp. 95–104).

Blakemore, C., & Campbell, F. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal image. *Journal of Physiology, 203*, 237–260.

Blanco, I., Martín-Bautista, M., Sánchez, D., & Vila, M. (2000). On the support of dependencies in relational databases: strong approximate dependencies. *Data mining and knowledge discovery*, submitted for publication.

Campbell, F., Cooper, G., & Enroth-Cugell, C. (1969). The spatial selectivity of the visual cells of the cat. *Journal of Physiology, 203*, 223–235.

Campbell, F., & Robson, J. (1967). Applications of fourier analysis to the visibility of gratings. *Journal of Physiology, 197*, 551–566.

Chang, S. (1997). Content-based indexing and retrieval of visual information. *SPMag, 14*(4), 45–48.

Chang, S.-F., Auffret, G., Foote, J., Li, C.-S., Shahraray, B., Syeda-Mahmood, T., & Zhan, H. (1999). Multimedia access and retrieval (panel session): The state of the art and future directions. In *Proceedings of the seventh ACM international conference on multimedia* (Vol. 1, pp. 443–445).

Cleland, B., Dubin, M., & Levick, W. R. (1971). Sustained and transient neurons in the cat's retina and lateral geniculate nucleus. *Journal of Physiology, 217*, 473–497.

Daugman, J. (1988). Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustic, Speech and Signal Processing, 36*, 1169–1179.

De Valois, R., Albrecht, D., & Thorell, L. (1982a). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research, 22*, 545–559.

De Valois, R., Yund, E., & Hepler, N. (1982b). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research, 22*, 531–544.

Del Bimbo, A., De Marsico, M., Levialdi, S., & Peritore, G. (1998). Query by dialog: an interactive approach to pictorial querying. *Image and Vision Computing, 16*(8), 557–569.

Del Bimbo, A., & Pala, P. (1997). Visual image retrieval by elastic matching of user sketches. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(2), 121–132.

Delgado, M., Marín, N., Sánchez, D., & Vila, M. (2001). Fuzzy association rules: General model and applications. *IEEE Transactions on Fuzzy Systems*, submitted for publication.

Delgado, M., Martín-Bautista, M., Sánchez, D., & Vila, M. (2000a). Mining strong approximate dependencies from relational databases. In *Proceedings of IPMU'2000*.

Delgado, M., Sánchez, D., & Vila, M. (2000b). Fuzzy cardinality based evaluation of quantified sentences. *International Journal of Approximate Reasoning, 23*, 23–66.

Dumais, S. (1991). Improving the retrieval of information from external sources. *Behavior Research Methods, Instruments and Computers, 23*(2), 229–236.

Fdez-Vidal, X., Rodríguez-Sánchez, R., Martínez-Baena, J., & Chamorro-Martinez, J. (2000). Image representational model for predicting visual distinctness of objects. *Proceedings of the 15th International Conference on Pattern Recognition, 1*, 689–694.

Field, D. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*(12), 2379–2394.

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steel, D., & Yanker, P. (1995). Query by image and video content: the qbic system. *IEEE Computing, 28*(9), 23–32.

Fuertes, J., Lucena, M., de la Blanca, N. P., & Chamorro-Martínez, J. (2001). A scheme of colour image retrieval from databases. *Pattern Recognition Letters, 22*, 323–337.

Fung, C. & Loe, K. (1999). Learning primitive and scene semantics of images for classification and retrieval. In *Proceedings of the seventh ACM international conference on multimedia* (Vol. 2, pp. 9–12).

Gevers, T., & Smeulders, A. (2000). Pictoseek: Combining color and shape invariant features for image retrieval. *IEEE Transactions on Image Processing, 9*(1), 102–119.

Graham, N., & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: A comparison of single-channel and multiple-channel models. *Vision Research, 11*, 251–259.

Hubel, D., & Livingstone, M. (1990). Color and contrast sensitivity in the lateral geniculate body and primary visual cortex of the macaque monkey. *Journal of Neuroscience, 10*, 2223–2237.

Hubel, D., & Wiesel, T. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology, 148*, 574–591.

Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106–154.

Hubel, D., & Wiesel, T. (1965). Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *Journal of Neurophysiology, 28*, 229–289.

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice Hall.

Kankahalli, M., Mehtre, B., & Huang, H. (1999). Color and spatial feature for content-based image retrieval. *Pattern Recognition Letters, 20*, 109–118.

Korfhage, R. (1997). *Information storage and retrieval*. Wiley and Sons.

Livingstone, M., & Hubel, D. (1984). Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience, 4*, 309–356.

Maffei, L., & Fiorentini, A. (1973). The visual cortex as a spatial frequency analyzer. *Vision Research, 13*, 1255–1267.

Marr, D. (1982). *Vision*. San Francisco, California: Freeman.

Medasani, S., & Krishnapuram, R. (1999). A fuzzy approach to content-based image retrieval. *Proceedings of the IEEE international fuzzy systems conference* (pp. 1251–1260).

Movshon, J., Adelson, E., Gizzi, M., & Newsome, W. (1984). The analysis of moving visual patterns. In En C Chagas, R. Gattass, & C. Gross (Eds.), *Experimental brain research supplementum II: Pattern recognition mechanisms* (pp. 117–151). New York: Springer-Verlag.

Pentland, A., Picard, R., & Sclaroff, S. (1996). Photobook: content-based manipulation of images databases. *International Journal on Computer Vision, 18*(3), 233–254.

Press, W., Teukolsky, S., Vetterling, W., & Flannery, B. (1992). *Numerical recipes in C: The art of scientific computing*. Cambridge University Press.

Sachs, M., Nachmias, J., & Robson, J. (1971). Spatial-frequency channels in human vision. *Journal of the Optical Society of America A, 61*, 1176–1186.

Shortliffe, E. (1976). *Computer-based medical consultations: MYCIN*. New York: Elsevier.

Shortliffe, E., & Buchanan, B. (1975). A model of inexact reasoning in medicine. *Mathematical Biosciences, 23*, 351–379.

Smith, J., & Chang, S. (1996). Visualseek: a fully automated content-based image query system. *Proceedings of ACM multimedia 96* (pp. 87–98).

Watson, A. (1987). The cortex transform: rapid computation of simulated neural images. *Computer Vision, Graphics and Image Process, 39*, 311–327.

Zadeh, L. A. (1983). A computational approach to fuzzy quantifiers in natural languages. *Computing and Mathematics with Applications, 9*(1), 149–184.