Expert Systems with Applications 40 (2013) 2398–2409

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Self organizing natural scene image retrieval

José Félix Serrano-Talamantes [a], Carlos Avilés-Cruz [b], Juan Villegas-Cortez [b,*], Juan H. Sossa-Azuela [c]

[a] Escuela Superior de Cómputo – Instituto Politécnico Nacional (ESCOM–IPN), Av. Juan de Dios Bátiz y Othón de Mendizábal s/n, Col. Lindavista, 07738 México, DF, Mexico
[b] Universidad Autónoma Metropolitana, Unidad Azcapotzalco, Departamento de Electrónica, Av. San Pablo 180, Col. Reynosa Tamaulipas, 02200 México, DF, Mexico
[c] Centro de Investigación en Computación – Instituto Politécnico Nacional (CIC–IPN), Av. Juan de Dios Bátiz y Othón de Mendizábal s/n, Col. Lindavista, 07738 México, DF, Mexico

ARTICLE INFO

Keywords:
Image analysis
Image processing
Content-based image retrieval (CBIR)
Feature extraction
Indexed database

ABSTRACT

In this work we describe a new statistically-based methodology to organize and retrieve images of natural scenes by combining feature extraction, automatic clustering, automatic indexing and classification techniques. Our proposal belongs to the content-based image retrieval (CBIR) category. Our goal is to retrieve images from an image database by their content. The methodology combines randomly extracted points for feature extraction. The describing features are the mean, the standard deviation and the homogeneity (from the co-occurrence matrix) of a sub-image extracted from the three color channels (HSI). A K-means algorithm and a 1-NN classifier are used to build an indexed database. Three databases of images of natural scenes are used during the training and testing processes. One of the advantages of our proposal is that the images are not labeled manually for their retrieval. The performance of our framework is shown through several experimental results, including a comparison with several classifiers and comparison with related works, achieving up to 100% good recognition. Additionally, our proposal includes scene retrieval.

1. Introduction

Nowadays, image retrieval is a very active research field because most of the information in Internet are images approximately, 73% (Bimbo, 1998). It is worth mentioning that this image related information is not well organized. Most of the known techniques and systems for image retrieval use some method of adding metadata such as captioning, keywords, or descriptions to the images, so that retrieval can be performed with the annotation words. Manual image annotation is, however, time-consuming, laborious and expensive. To address this issue, a large amount of research on automatic image annotation has been done. In content-based image retrieval (CBIR), also known as query by image content (QBIC), the term "Content" refers to colors, shapes, textures, or any other information derived from the image itself (Datta, Joshi, Li, & Wang, 2008; Liu et al., 2007). CBIR techniques are desirable because most web-based image searching engines relying purely on metadata produce a lot of garbage results. Also having humans manually entering keywords for images in a large database is inefficient, expensive and may not capture every keyword that describes an image. Techniques and systems able to filter images based on their content would provide better indexing and return more accurate results. One of the key issues for a successful CBIR system is choosing the features that represent the images accurately and uniquely. These features have to be discriminative and sufficient in describing the content of an image. Usually, any CBIR system will use three basic types of features: color, texture and shape. Years of research have demonstrated that it is very difficult to achieve satisfactory retrieval results using only one of these feature types (Datta et al., 2008; Liu et al., 2007). Based on the above discussion, in this paper we describe a CBIR methodology for the retrieval of images of natural scenes. To attain the goal of efficiently retrieving images of natural scenes, in this paper, we propose to combine color and texture features for image organization and image retrieval. We have evaluated our methodology using three databases:

- Vogel and Shiele (VS) Vogel et al. (2006). 700 Images classified as: 144 coast, 103 forest, 179 mountain, 131 prairie, 111 river/lake, and 32 sky/cloud.
- Oliva and Torralba (OT) Oliva and Torralba (2001). 1472 Images classified as: 360 coast, 328 forest, 374 mountain and 410 prairie.
- Fei-Fei, Fergus and Perona (FP) Fei-Fei et al. (2004). 373 Images classified as: 128 Bonsai, 60 Joshua Tree, 85 Sun Flower, 64 Lotus, and 36 Water Lily.

We will refer to these image databases as: VS, OT, FP (Caltech-101) respectively. We compare our results against (Vogel & Schiele,

* Corresponding author. Tel.: +52 55 53189551; fax: +52 55 53946894.
E-mail addresses: jfserrano@ipn.mx (J.F. Serrano-Talamantes), caviles@correo.azc.uam.mx (C. Avilés-Cruz), juanvc@correo.azc.uam.mx (J. Villegas-Cortez), hsossa@cic.ipn.mx (J.H. Sossa-Azuela).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.10.064

2007; Oliva & Torralba, 2001 & Fei-Fei et al., 2004), achieving up to 100% good recognition with our proposal.

The performance of the proposal is tested with several classifiers (*K*-NN, Bayesian and ANN backpropagation type). The rest of the paper is organized as follows. In Section 2 we present a state of the art review of the most important works related with the problem of image retrieval, mainly CBIR systems. In Section 3, each of the stages of the proposal (training and testing) is detailed. In Section 4 several experiments to test the performance of our methodology are presented. In Section 5, we show the comparison between our results and related works (Vogel et al., 2006; Oliva & Torralba, 2001; Fei-Fei et al., 2004). Finally the conclusions and further research are detailed in Section 6.

2. State of the art

Nowadays, considering the availability of large storage spaces, a huge number of images can be found over the Internet. With this huge distributed and heterogeneous image database, people want to search and make use of the images contained there. Thus a great challenge emerges: finding out accurate ways of searching images. Basically, images can be retrieved in two ways, first, text-based, secondly, content-based or query by example-based. Text-based retrieval techniques are very well-known and widely used. In this case users are provided with a text area to enter the key words (usually the image file name) on the basis of which image searching is done. It is widely used in the Google web-based image searching technique.

Because of the availability and the rapid growth of storage devices, a huge number of images, digital photographs and Internet access, content-based image retrieval (CBIR), has become increasingly used in recent years around the world. With this huge distributed and heterogeneous image database, people want to search in and make use of the images there contained. Systems that can automatically analyze, categorize and search image databases have been developed in research labs and with commercial concerns.

The term CBIR seems to have originated in the earlier 2000s (Feng, Siu, & Zhang, 2010; Vogel & Schiele, 2007; Liu et al., 2007; Hiremath, 2007; Sumana, Islam, Zhang, & Lu, 2008; Diaz & Sturm, 2009). CBIR includes research on: Automatic Feature Extraction (Bimbo, 1998; Datta et al., 2008), Automatic Feature Extraction with a Semantic Content (Vogel & Schiele, 2007; Rui, Huang, & Chang, 1999; Liu et al., 2007 & Li & Wang, 2006) and data representation (Li & Wang, 2006). CBIR techniques use low-level features such as texture, color and shape to represent images. Among those low level image features, texture features have shown very effective and subjective results (Feng et al., 2010).

CBIR technique in Vogel et al. (2006), Vogel and Schiele (2007), describes a computational image representation with the aim to reduce the semantic gap between image understanding by humans and computers. Our proposal allows a semantic description, understanding, and modeling of natural images. CBIR technique in Serrano et al. (2009) allows image retrieval from a natural scenario database. This technique combines fixed and random extracted points for feature extraction. The advantage of the proposal is that it is not needed to manually label images such as (Vogel & Schiele, 2007, 2006; Vogel, Schwaninger, Wallraven, & Bulthoff, 2006) done to achieve image retrieval. CBIR technique in Serrano, Avilés, Sossa, Villegas, and Olague (2010) allows image retrieval from a database with similar lighting conditions.

The image indexing to handle large volumes of information is another technical consideration to be taken into account to integrate modules for feature extraction, physical image storage, similarity measures, query procedures, user interface and system architecture (Liu et al., 2007).

Gonzalez-Garcia (2007) carry out image retrieval using Daubechies wavelet transform, which uses histograms for color extraction. For classification features uses a multilayer perception. Image retrieval is based on color.

Schmid (2004) proposes implementing image retrieval using generic descriptors such as invariant to rotations and applied them to each pixel. Images are in gray levels. This proposal uses *K*-means algorithm to form clusters. Use Euclidean distance for the comparison between the descriptors.

Sumana et al. (2008) propose a new image feature based on so-called curvelet transform. They apply discrete curvelet transform on texture images and compute the low order statistics from the transformed images. Images are then represented using the extracted texture features. Through experiments, authors show that their proposal significantly outperforms the Gabor texture feature (Manjunath & Ma, 1996).

Gonzalez (2006) combine wavelet-based descriptions and artificial neural networks for image representation and retrieval. They tested performance in three different ways to obtain wavelet coefficients. For training, authors used 120 color images of airplanes; for testing they used 240 images. The best efficiency of 88% was obtained with the third description method.

Subrahmanyam, Maheshwari, and Balasubramanian (2012) propose an image indexing retrieval system for CBIR. The extracted characteristics are based on color histogram and SOT (spacial orientation tree), which defines the spatial parent–child relationship among wavelet coefficients in multi-resolution wavelet sub-bands, and these characteristics are disposed as vector points. This proposed method was tested on Corel 1000, Brodatz texture image and MIT VisTex databases of natural images.

Yildizer, Balci, Hassan, and Alhajj (2012) propose a CBIR system which uses Multiple Support Vector Machine Ensemble. The characteristic vector is conformed using the Daubechies wavelet transform over the images. This method is oriented for the classification of big image databases and provides fast results over natural-scene images of Corel image database.

The image classification techniques in Bosch, Muñoz, and Martí (2007) present different approaches such as low-level modeling and semantic modeling. The low-level modeling uses low-level features (color, texture) and information from the histogram to determine directly the type of scene.

Su, Chen, and Lien (2010) proposed to combine two well known methods, relevance feedback (RF) and region-based image retrieval (RBIR) in order to enhance the performance of the CBIR systems. This method considers human vision aspects of quantity-visual perception (unit and groups, e.g. ‘single flower' and ‘bouquets of flowers'), in order to provide subjectively feedbacks in accordance with particular interests. This method develops a modified group biased discriminant analysis (GBDA) considering regions of the image, and applying a similarity measure between images constructed on the basis of the region-based relevance feedbacks, providing heuristic pre-clustering results. The method is tested using the Corel image database in several categories with low resolution (120×80 pixel or 80×120 pixel).

Image databases have been studied for several years. The first approaches to index large volumes of images were performed using keywords, but the construction of the index would become costly and subjective. The Photobook or the QBIC (Liu et al., 2007; Rui et al., 1999) among others have been some of the early works for classification of image databases in which the authors were concerned with the visual properties of the image and its characteristics of shape, area and texture, implementing an image retrieval system that uses visual operators. This last method has the common sense of visual perception such as (Su et al., 2010).

ElAlami (2011) introduces an unsupervised retrieval framework based on a rule base system algorithm. The images and their classes pass through a relevant feedback phase followed by a clustering refined model. The retrieval phase is according to the rules, such as the query image match with the query image features. The system rules are proposed in the context of data mining in the form If-Then one as follows: if <*Conditions*> then <*Class*> (Zhu & Huang, 2008).

Finally, Long et al. present a complete survey of CBIR fundamental theories in Feng et al. (2010).

3. Proposal

In this section we describe each stage of the proposed methodology for image retrieval of natural scenes from a database. As shown in Figs. 1 and 2, the methodology has two basic stages as follows:

3.1. Training stage

This stage is divided into four main phases: (see Fig. 1)

1. Feature extraction
2. Feature clustering (*K*-means)
3. Feature classification (*K*-NN)
4. Indexed database building

3.1.1. Feature extraction (see Fig. 1)

Given a database composed of *n* images, during this phase, the *n* images (training set) in RGB format are first read. Each image is then converted to HSI format.

For each of these *n* images, 300 pixels are uniformly selected at random (see Fig. 3(a)). For each of these 300 points, a square window of 10×10 pixels is opened. Fig. 3(b) shows several examples. For each of these 300 windows, the following describing features are extracted: the mean, the standard deviation (Fukunaga, 1990) and the homogeneity obtained from the co-occurrence matrix (Marceau, Howarth, & Gratton, 1990). This process is performed for each of the 300 windows in three channels: Hue (H), Saturation (S) and Intensity (I) of an image. The corresponding describing vector for each window of the image thus has nine components, three for the H channel, three for the S channel and three for the I channel.

3.1.2. Feature clustering (see Fig. 1)

In order to cluster the resulting $n \times 300$ describing vectors (300 for each of the *n* images) a *K*-means algorithm (Fukunaga, 1990) was applied to organize these $n \times 300$ features into 10 clusters ($k = 10$ image components). At the end of this phase we will have the $n \times 300$ vectors clustered into the ten classes (image components). An image component (*IC*) is a representative part of an image. In the case of an image of a natural beach scene an *IC* could be the sky region of the image or its sand region, etc.

To verify the correct functioning of the *K*-means algorithm, we took one image at random from the *n* images of one of the three databases and it was marked with 300 seeded random points over the whole image. This is depicted in Fig. 4, which has its 300 points clustered in similar areas.

3.1.3. Feature classification

During this phase (see Fig. 1), the describing vectors obtained during the training phase along with the *IC*s to which these vectors belong are used to train a selected classifier. The describing features are: the mean, the standard deviation and the homogeneity

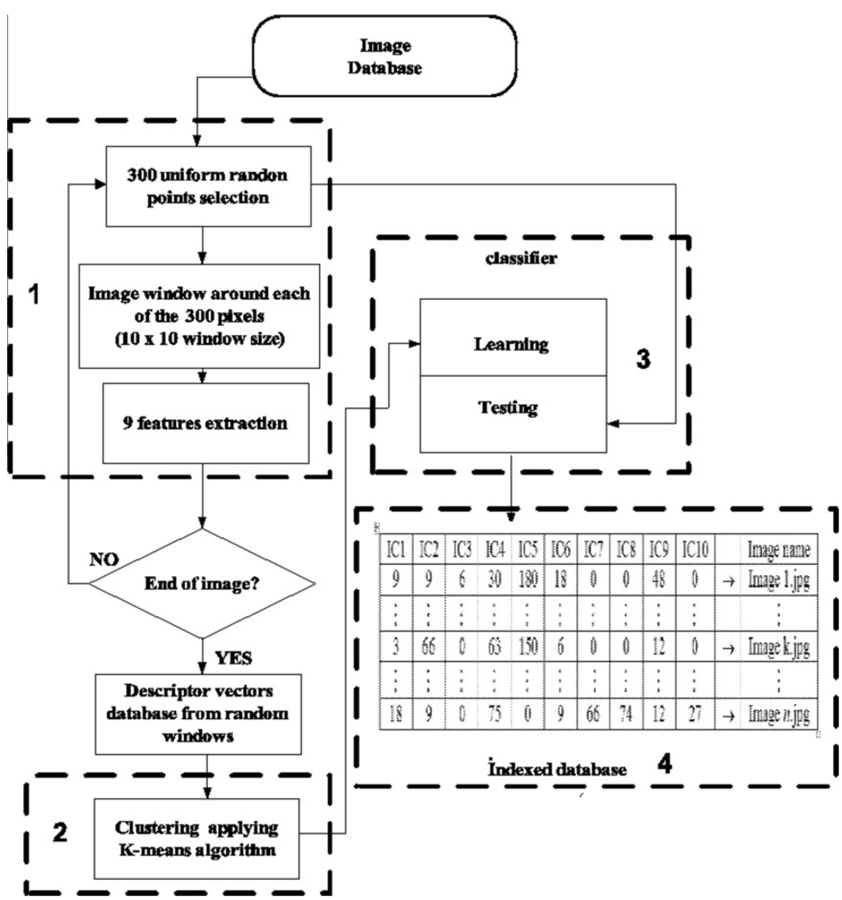


Fig. 1. Flow diagram for training stage.

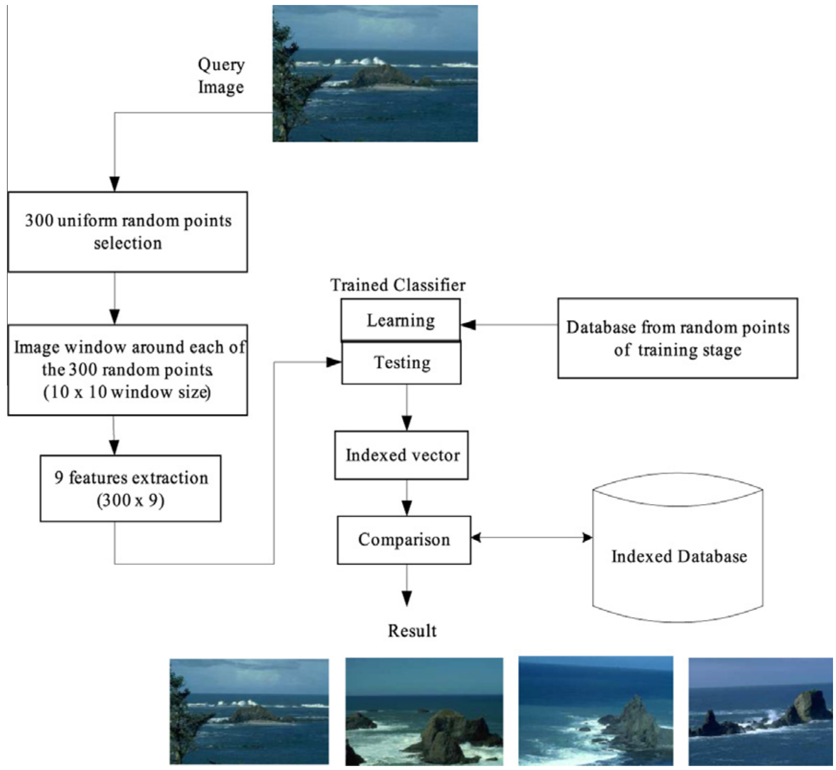


Fig. 2. Flow diagram for testing stage.

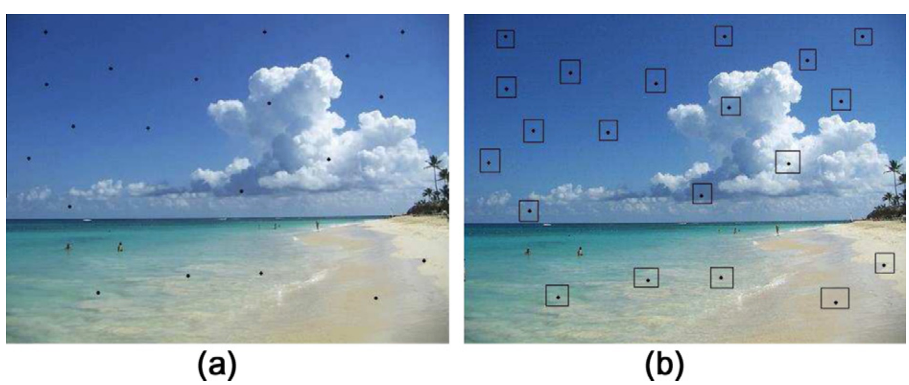


Fig. 3. Example of selected points over an image at random (only 20 points are presented). (a) Uniforme random points, (b) their corresponding window of M × N pixels.

(from the co-occurrence matrix) of a sub-image extracted from the three color channels (HSI). So the descriptor vector dimension is equal to 9, and forms one database of n vectors describing 9 features each. The goal of applying $K = 10$ in the K-means algorithm is to organized the database into 10 possible classes of objects that a natural scene can contain, such as: sky, water, rock, clouds, grass, foliage, sky/water, sky/grass, sky/clouds, and sky/rock. The classifier input is the vector (9 describing features) from the previously built database and its output is the association of every describing feature per input vector to one class ($IC1 \sim IC10$), and after all the whole training process we have 10 IC clusters obtained from the K-means algorithm.

Fig. 4. Clusters formed with the 300 random seeded points on an image of a natural scene after K-means application algorithm. We can see that the black diamonds signed points fall into the grass component of the image, the add symbols group together into the sky component of the image, and the boxes cluster together into the foliage image component.

3.1.4. Indexed database building

To build the indexed database we proceeded as follows. For each of the n images of the selected database, we took its 300 describing vectors (those obtained during the first training phase) and presented at the input of the already trained classifier. At the

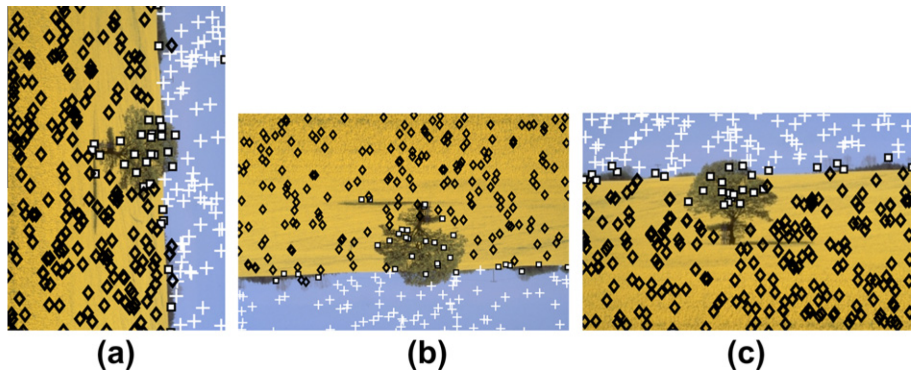


Fig. 5. Clusters formed after *K*-means application algorithm on image transformations and scaling: (a) 90° rotation, (b) 180° and (c) scaling change at 50%. Note that in all cases, the dots cluster are very similarly.

Table 1
Distribution of the 210,000 describing features among the 10 image components for Vogel–Shiele database.

Image component number	Number of the describing features per image component
1	22,086
2	23,267
3	23,899
4	16,127
5	23,926
6	24,506
7	30,262
8	10,708
9	10,967
10	24,252
	Total = 210,000

Table 2
Distribution of the 441,600 describing features among the 10 image components for Oliva–Torralba database.

Image component number	Number of the describing features per image component
1	50,482
2	26,978
3	21,988
4	45,213
5	54,105
6	52,257
7	29,643
8	48,114
9	52,957
10	59,863
	Total = 441,600

Table 3
Distribution of the 111,900 describing features among the 10 image components for Fergus–Perona (Caltech-101) database.

Image component number	Number of the describing features per image component
1	9326
2	8328
3	9324
4	8867
5	14,657
6	12,320
7	8941
8	10,072
9	17,769
10	12,296
	Total = 111,900

Table 4
Indexed database structure for the Vogel–Shiele database.

*IC*1	*IC*2	*IC*3	*IC*4	*IC*5	*IC*6	*IC*7	*IC*8	*IC*9	*IC*10	→	Image name
9	9	6	30	180	18	0	0	48	0	→	image1.jpg
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3	66	0	63	150	6	0	0	12	0	→	image k.jpg
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
18	9	0	75	0	9	66	74	12	27	→	image n.jpg

Table 5
Indexed database structure for the Oliva–Torralba database.

*IC*1	*IC*2	*IC*3	*IC*4	*IC*5	*IC*6	*IC*7	*IC*8	*IC*9	*IC*10	→	Image name
65	74	9	10	30	0	0	1	107	4	→	image1.jpg
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	130	7	29	1	1	55	7	69	1	→	image k.jpg
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	41	28	35	28	64	0	26	7	71	→	image n.jpg

Table 6
Indexed database structure for the Fergus–Perona (Caltech-101) database.

*IC*1	*IC*2	*IC*3	*IC*4	*IC*5	*IC*6	*IC*7	*IC*8	*IC*9	*IC*10	→	Image name
40	46	21	0	47	8	3	77	7	51	→	image1.jpg
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
13	17	129	13	14	0	0	91	16	7	→	image k.jpg
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
31	22	24	1	113	3	0	4	72	30	→	image n.jpg

output of the classifier we obtained the class (*IC*) to which each describing vector (sub-image) belonged. Each time a describing vector was classified into an *IC*, a vote was accumulated to a case in a 10 element register.

We took the same image and applied to it several image transformations to verify if the clustering results was kept. Fig. 5(a)–(c) show several image transformations; 90° image rotation, 180° image rotation and scaling at 50%. In all cases, image components clusters are very similar.

As an example, for the VS database using ($n = 700$) images for training, the Table 1 shows how many of these 210,000 (700×300) vectors fall into *IC* number one, how many vectors fall into *IC* number two, and so on until *IC* number ten. This gives us somehow the probability that a given *IC* belongs to the 700 images. From Table 1, we can see that *IC* number 7 is the most representative, while *IC*s 8 and 9 are the less representative. For the OT database using ($n = 1472$) images for training, the Table 2 shows how many of these 44,100 (1472×300) vectors fall into *IC* number one, how many vectors fall into *IC* number two, and so on until *IC* number ten. This gives us somehow the probability that a given *IC* belongs to the 1472 images. For the FP (Caltech-101) database using ($n = 373$) images for training, the Table 3 shows how many of these 111,900 (373×300) vectors fall into *IC* number one, how many vectors fall into *IC* number two, and so on until *IC* number 10. This gives us somehow the probability that a given *IC* belongs to the 373 images.

In Tables 4–6 we can see how the sum of each line is equal to 300 through $\sum_{i=1}^{10} IC_i = 300$. Each line is voted for its corresponding 10 *IC*s. As example, in Table 4, line 1, the most voted *IC* was *IC* number 5, while the least voted were *IC*s 7, 8 and 10. This process is repeated for the remaining n images belonging to every dataset (OT database and FP (Caltech-101) database) respectively, as they are depicted in Tables 5 and 6.

As shown in Table 4–6, each register contains ten numbers and one pointer to an image. For a given image, each of the ten numbers is the number of times a given *IC* is found. We are aware that this is not the best computational structure (in terms of image searching) but as we will see later, it will allow us to test the functioning of our proposal.

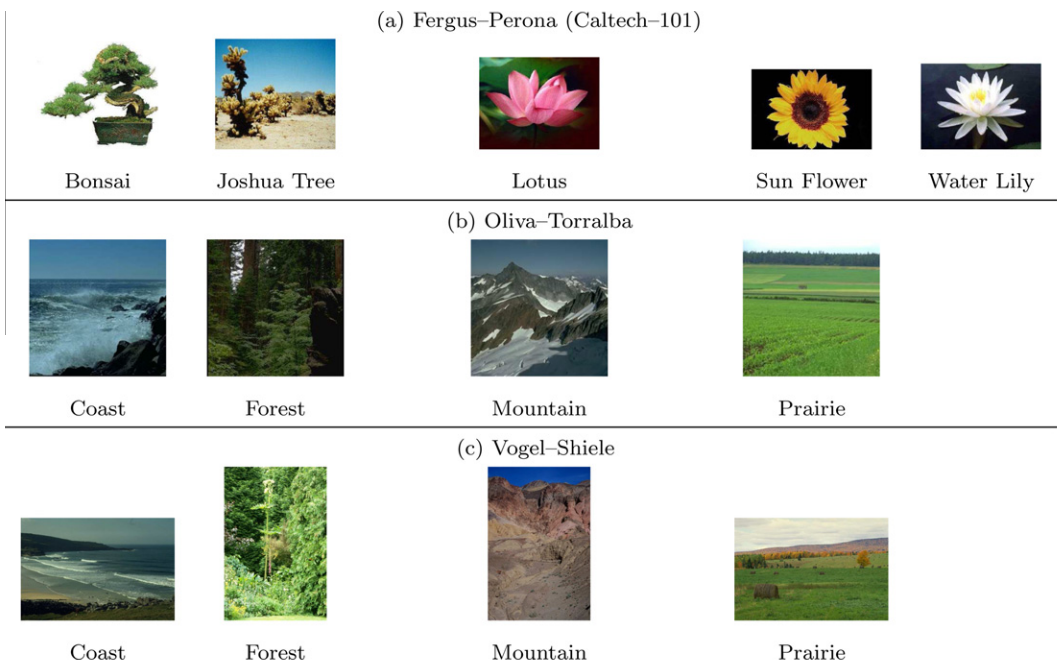


Fig. 6. Categories for image organization and retrieval for each of the three databases: (a) FP, (b) OT and (c) VS respectively.

Table 7
Efficiency of each classifier for each database.

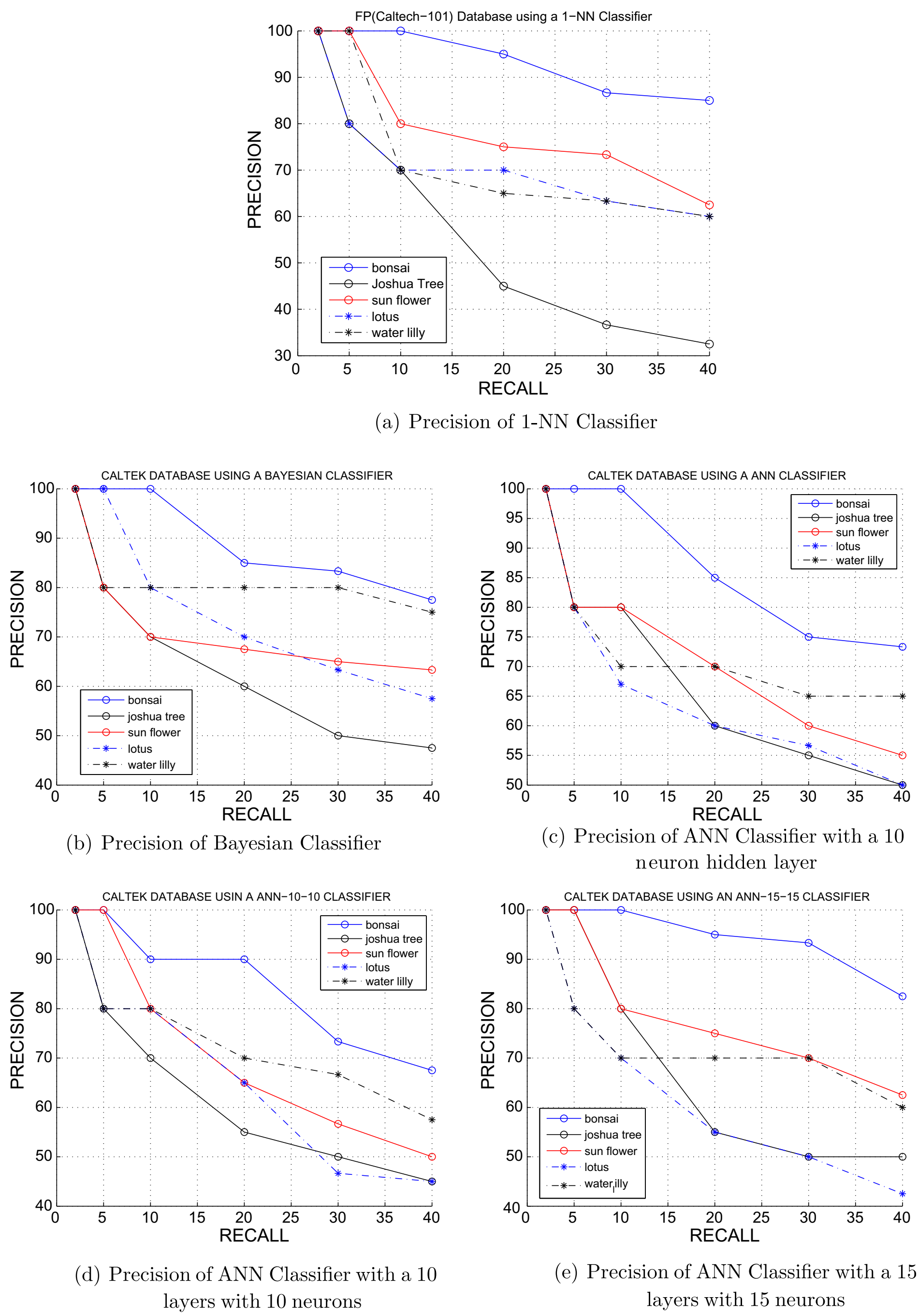
Database/classifier	1-NN	Bayesian	ANN with a 10 Neuron hiddean layer	ANN with a 10 layers With 10 neurons each	ANN with a 15 layers With 15 neurons each
FP (Caltech–101):373 images					
Bonsai (128)	94.44%	90.97%	88.88%	86.80%	95.13%
Joshua tree (60)	60.69%	67.91%	70.83%	66.67%	75.20%
Sun Flower (85)	81.80%	74.30%	74.16%	75.28%	81.25%
Lotus (64)	73.88%	78.47%	68.94%	69.44%	66.25%
Water Lilly (36)	76.38%	82.50%	75.00%	75.69%	75.00%
Mean	74.44%	78.83%	75.56%	74.77%	78.02%
OT:1472 images					
Forest (328)	81.12%	83.43%	83.50%	91.36%	86.62%
Coast (360)	68.56%	64.32%	63.61%	74.07%	67.51%
Mountains (374)	64.97%	67.33%	58.33%	72.13%	73.50%
Prairies (410)	39.64%	76.33%	59.20%	68.02%	61.91%
Mean	63.57%	72.86%	66.16%	76.39%	72.38%
VS: 668 images					
Forest (103)	74.93%	64.30%	75.56%	92.73%	82.51%
Coast (255)	60.70%	70.33%	81.41%	78.73%	74.83%
Mountains (179)	68.34%	76.69%	62.24%	69.05%	83.24%
Prairies (131)	51.47%	67.07%	66.32%	53.13%	63.46%
Mean	63.86%	69.60%	71.38%	70.91%	76.01%

3.2. Retrieval stage

This stage proceeds as depicted in Fig. 2. A query image is presented to the system, and the same six first processing steps used for learning are applied to it. As a result we get 300 describing vectors. These 300 vectors are presented to the input of the already trained classifier. As output from the classification process we just get one vector. This indexing vector contains the number of times (votes), that each one of the ten image components: $IC1, IC2, \ldots, IC10$ is supposed to be contained in the query image. To get from the database the most similar images, this vector is compared with the n vectors stocked in the indexed database. To reduce a bit the computing time, we took into account the four image components with the higher number of votes. The distance used in this case was the Euclidean distance.

4. Experimental results

In order to test the performance of our proposal three datasets were used as mentioned before. The first dataset contained 668 images of natural scenarios from the VS database. The second dataset contained 373 images from the FP (Caltek-101) database, and the third dataset contained 1472 images from the OT database. The classes shown in Fig. 6 were used in each database for testing purposes. So we used four for the VS database, 5 for the FP (Caltek-101) database and 4 for the OT database.



(a) Precision of 1-NN Classifier

(b) Precision of Bayesian Classifier

(c) Precision of ANN Classifier with a 10 neuron hidden layer

(d) Precision of ANN Classifier with a 10 layers with 10 neurons

(e) Precision of ANN Classifier with a 15 layers with 15 neurons

Fig. 7. The precision for the FP (Caltech-101) database.

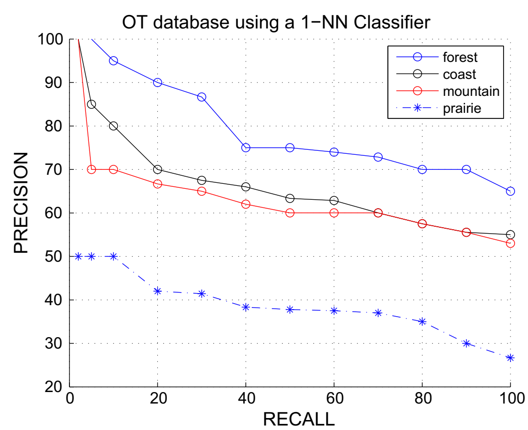
We compared the efficiency of our proposal with three classifiers: a non-parametric classifier (1-NN), a Bayesian classifier and an artificial neural network based classifier (ANNC). For the ANNC, different configurations have been tested. Table 7 resumes the efficiency of three classifiers (1-NN, Bayesian and ANNC). As we can see according to the Table 7, the ANNC presents the best efficiency. In some isolated cases, the 1-NN and Bayesian classifiers outperform the ANNC classifier.

To test the performance of each classifier for a given database, we proceeded as follows. For the FP (Caltech-101) database, we took a random selection of 5 images from each of the 5 classes of the same database (5 bonsai, 5 Joshua tree, 5 sun flower, 5 lotus and 5 Water Lilly images). For the VS database, we took a random selection of 12 images from each of the four classes of the OT database (12 forest, 12 coast, 12 mountain and 12 prairies images). Finally, for the OT database, we took a random selection of 12 images from each of the four classes from the VS database (12 forest, 12 coast, 12 mountain and 12 prairies images).

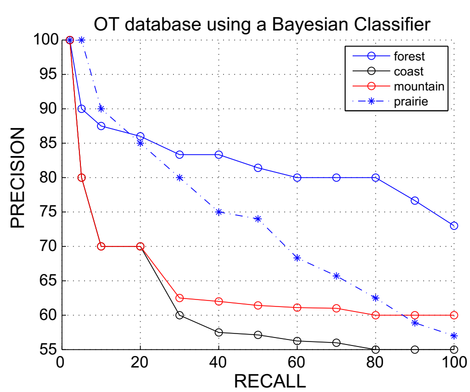
To measure the efficiency of our proposal we used the following two measures, P = Precision and R = Recall:

$$P = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \times 100\% \quad (1)$$

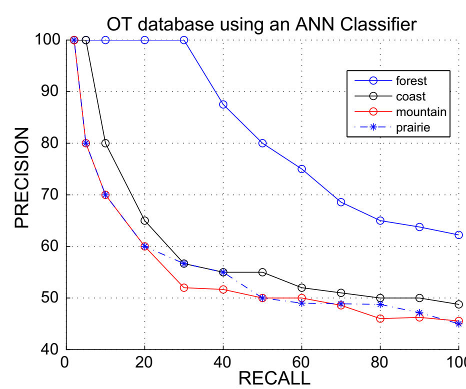
$$R = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images in the database}} \times 100\% \quad (2)$$



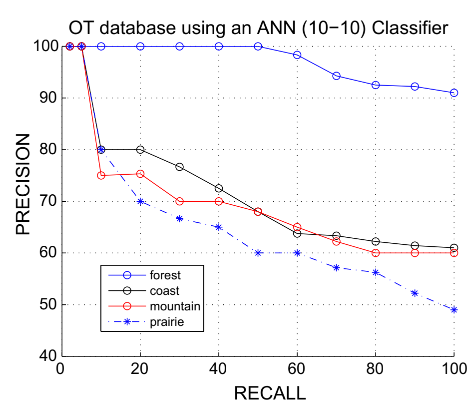
(a) Precision of 1-NN Classifier



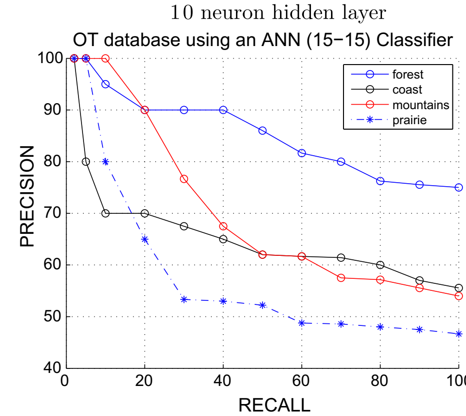
(b) Precision of Bayesian Classifier



(c) Precision of ANN Classifier with a 10 neuron hidden layer



(d) Precision of ANN Classifier with a 10 layers with 10 neurons



(e) Precision of ANN Classifier with a 15 layers with 15 neurons

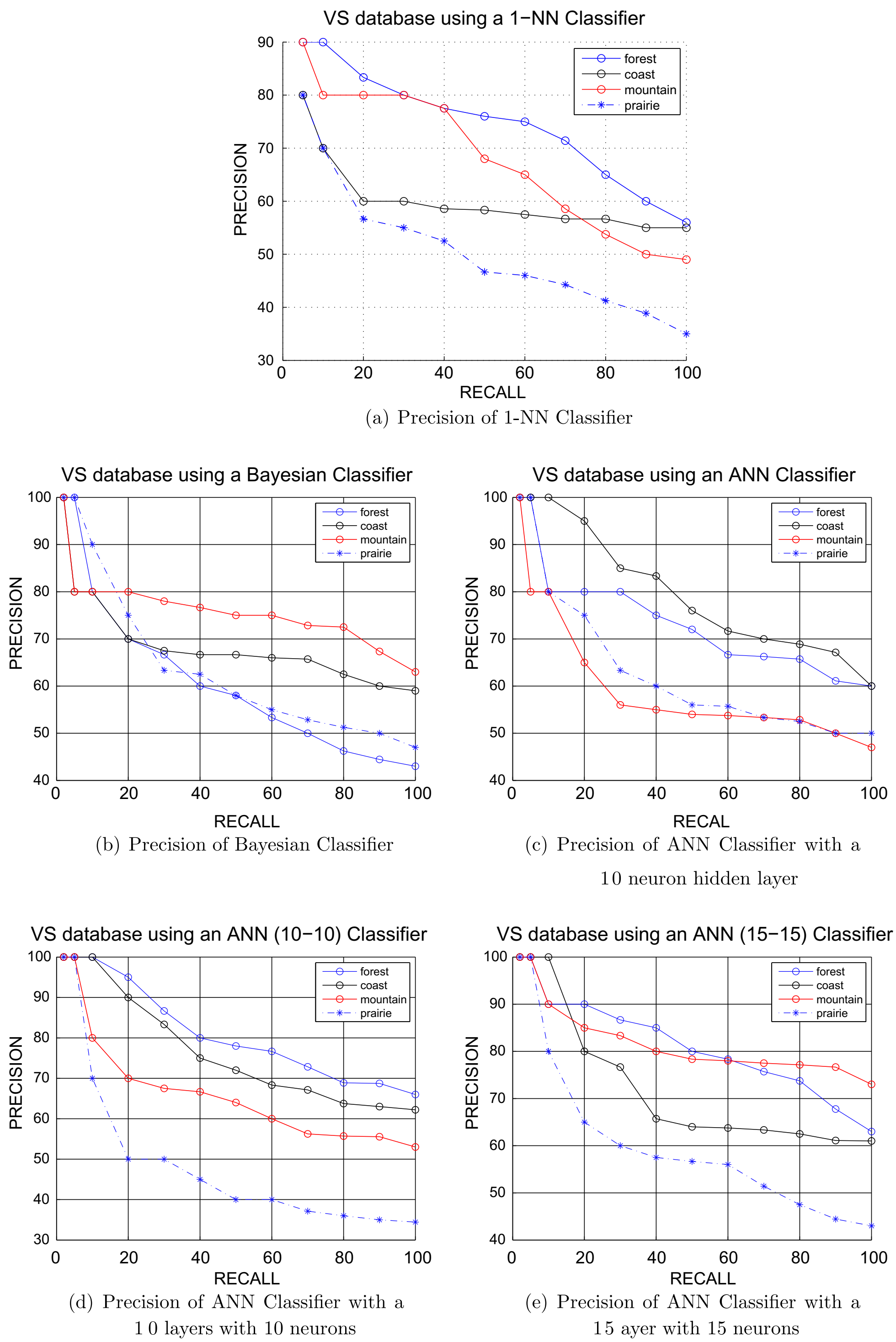
Fig. 8. The precision for the OT database.

The experimental results are plotted in the figure sets Figs. 7–9. In Fig. 7(a)–(e) show the precision for FP (Caltech-101) database respectively. Here we have the best precision vs recall for 1-NN classifier and the 15 ANN classifier, and the result is the same in mean values for the Figure set Fig. 8.

The Fig. 7(e) shows that the best image retrieval is reached in the bonzai image and in the Joshua tree image classes, in Fig. 7(a), the best image retrieval is reached by the Sun Flower image class, in Fig. 7(b), the best image retrieval is reached in the Lotus image and in the Water Lilly image classes respectively.

The graphics depicted in Fig. 8(a)–(e) show the precision vs recall for the OT database respectively. As we see in Fig. 8(d), the best image retrieval is reached for the forest image and the coast image classes, in Fig. 8(e) the best image retrieval is reached by the mountain image class, and in Fig. 8(b) the best image retrieval is reached for the prairie image class.

The efficiency of each classifier for each database is depicted in Table 7. For the first two image sets (FP and OT) the mean values show that the best classifier is the Bayessian, and the second one to get best retrieval is the ANN with 15 layers with 15 neurons



(a) Precision of 1-NN Classifier

(b) Precision of Bayesian Classifier

(c) Precision of ANN Classifier with a 10 neuron hidden layer

(d) Precision of ANN Classifier with a 1 0 layers with 10 neurons

(e) Precision of ANN Classifier with a 15 ayer with 15 neurons

Fig. 9. The precision for the VS database.

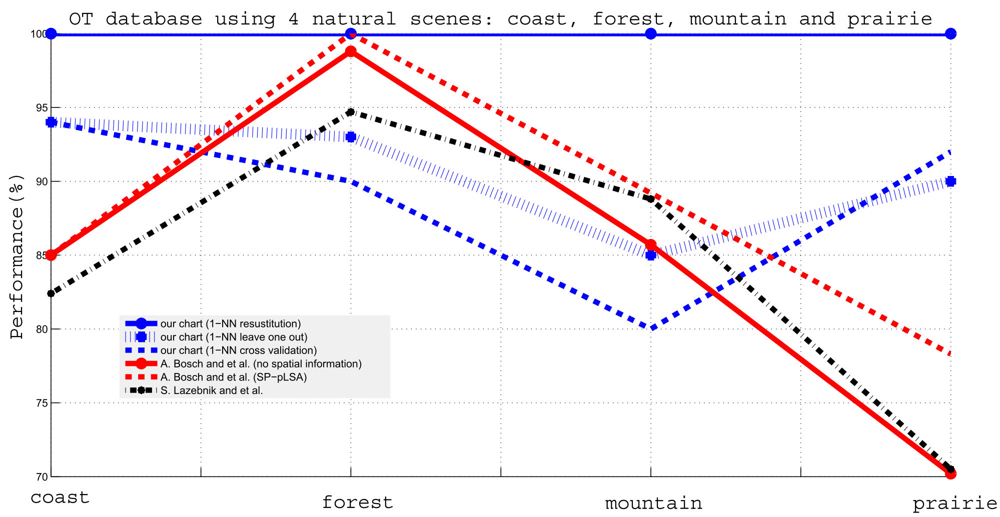


Fig. 10. Classification performance of our proposal against the algorithm of Bosch et al. (2008), and Lazebnik et al. (2006).

Table 8
Optimized values from the Oliva–Torralba database for each case.

Database	# of categ.	#train	#test	Authors
OT	4	1472	1472	Resustitution (our proposal = 100%)
OT	4	1471	1	Leave one out (our proposal = 90.5%)
OT	4	736	736	Cross validation (our proposal = 89%)
OT	4	1000	472	Bosch et al. (2008) in Fig. 13a = 84.92%
OT	4	1000	472	Bosch et al. (2008) in Fig. 15a = 88.12%
OT	4	400	1072	Lazebnik et al. (2006) in Fig. 3 = 84.1%

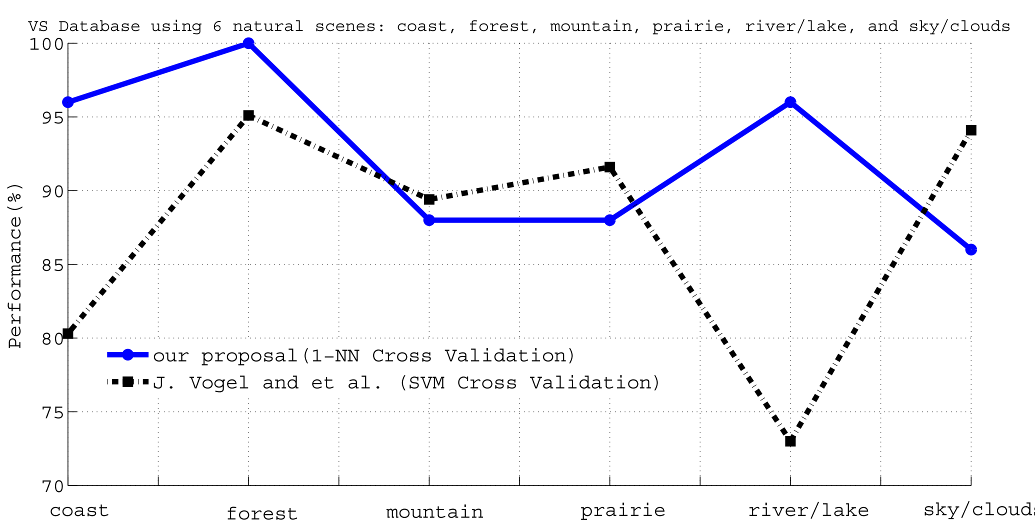


Fig. 11. Classification performance of our proposal against the algorithm of Vogel and Schiele (2007).

each. The third image set was best classified using the ANN 15 layers classifier, but the main aspect here is the different number of images per every set. The OT set has the greatest number of images, and here the Bayesian classifier shows the best performance, versus the ANN classifier for smaller image number set as the classic perspective use for the ANN purposes.

For these three cases our proposal reaches its higher performance using the Bayesian classifier, taking in mind the last observation in average result. The last image data set (VS), seems to

Table 9
Optimized values from the Vogel–Shiele database for each case.

Database	# of categ.	#train	#test	Authors
VS	4	600	100	Vogel and Schiele (2007) in Table 3 = 87.25%
VS	4	350	350	Cross validation (our proposal = 92.33%)

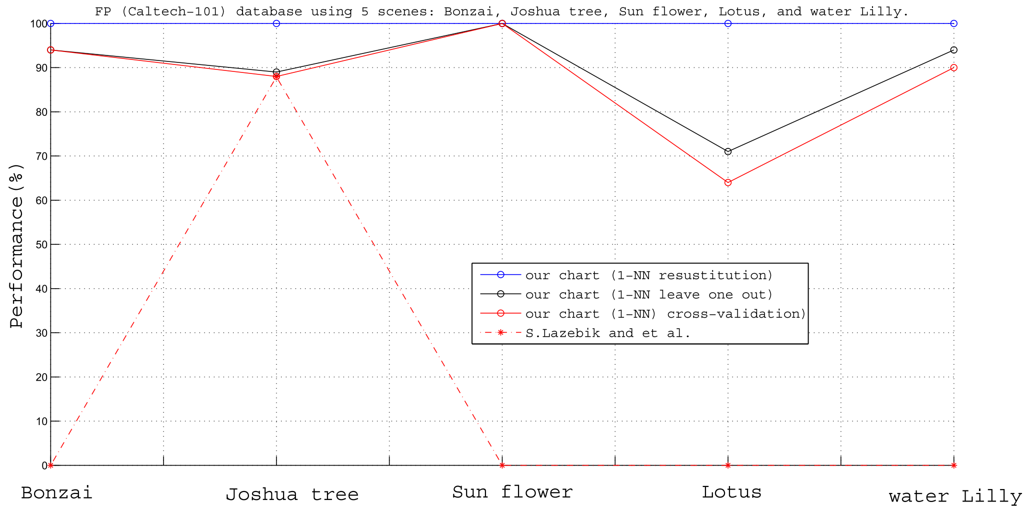


Fig. 12. Classification performance of our proposal against the algorithm of Lazebnik et al. (2006).

Table 10
Optimized values from the Fergus–Perona (Caltech–101) database for each case.

Database	# of categ.	#train	#test	Authors
FP (Caltech-101)	5	373	373	Resubstitution (our proposal = 100%)
FP (Caltech-101)	5	372	1	Leave one out (our proposal = 89.6%)
FP (Caltech-101)	5	187	186	Cross validation (our proposal = 87.20%)
FP (Caltech-101)	1	30	50	Lazebnik et al. (2006) in Fig. 5 (Joshua_tree) = 87.9%

reach the best results with ANN but the image number set is less than the second one image set.

5. Comparison to previous classification results: natural scene classification

We compare the performance of our classifier K-NN ($K = 1$) against the algorithm of Bosch, Zisserman, and Muoz (2008), and Lazebnik, Schmid, and Ponce (2006), see Fig. 10.

From Fig. 10, we can see that in order to evaluate and compare our classifier versus the ones presented in Bosch et al. (2008) and Lazebnik et al. (2006), we used the "Resustitution method". This method consist of using all the images in the database (1472 in this case), for the training and the testing stages. We also used the method "leave one out" (LOO), which consists of using one image to test and the remaining images to train the classifier; once the training phase ends, the image used for testing purposes is returned into the database. This procedure is repeated with all the images of the database. Finally, we use the "Cross-validation" method, it consists of random selection of half image set part from database of each class for training purposes, and the other half part for testing purposes. This procedure is carried out during 20 iterations and finally we obtain an average amount. In each of the three methods we form one confusion matrix for each class of the image database to evaluate the classifier performance on the main diagonal.

From Fig. 10 we obtain Table 8, which shows the performance for classification purposes of each chart method, using cross-validation in order to evaluate our classifier versus the other ones. Our proposal percentage numbers outperform better than the others (100% for resubstitution method, first row, right column).

We also compare the performance of our classifier K-NN ($K = 1$) against the algorithm of Vogel and Schiele (2007). The resulting approach is depicted in Fig. 11. From this data we obtain Table 9, where our best result (92.33% classification) which is better than the result cited in Vogel and Schiele (2007).

Finally, the FP (Caltech-101) classification result turns out to be our best one result of our proposal. It reaches out until 100% for the resubstitution method and 89.6% for the LOO method, which result better than the ones cited in Lazebnik et al. (2006). The corresponding plot of this approach is depicted in Fig. 12, and the numbers are shown in Table 10.

6. Conclusions and further research

In this paper we have described a new methodology that allows the retrieval of natural images automatically and it has been proofed for three different image databases. During the learning phase, the proposal takes as input one set of images divided into four classes: coasts, mountains, forests, and plains. It extracts from these images the describing features from sets of points randomly and automatically seeded. A K-means algorithm is used to form 10 different clusters from the describing features obtained from the chosen points. Several classifiers such as 1-NN, Bayesian and ANN backpropagation were used to build one indexed database from the combination of all the describing vectors which take image texture features. During the retrieval phase the same classifiers already trained were used to retrieve images from the indexed database, the retrieved images are the most similar images given by the query image. The experimental results, which were validated using the precision/recall measure, show that our proposal performs better than the results reported method in Vogel and Schiele (2007), Vogel. (2004).

We have also shown the case of natural scenarios images, here our proposal presents good retrieval performance, although different classifiers front were used.

One of the advantages of the proposal is that neither the image contents or the image areas are not manually labeled, as it is the case by which human vision proceeds for their retrieval, or for image query identification. Also our proposal improves the classification performance using 1-NN classifier against proposals presented in, Lazebnik et al. (2006), Vogel et al. (2006), Bosch et al. (2008).

To extensively validate the utility of our proposal some tasks will be done:

(a) Testing the proposal by changing the number of image categories. Until now we have only used 4 and 5 categories. One experiment to be done will consist on testing with less and more categories, for example, 2, 3, 4, 5, 6, and so on and verify if the effectiveness of the proposal is kept, improve or reduce.
(b) Testing the performance of the proposal with more classifiers, such as support vector machine bases classifiers, associative memory based classifiers, fuzzy based classifiers, and so on.

References

Gonzalez, A. C., Sossa, J. H., Felipe, E. M., & Pogrebnyak, O. (2006). Wavelet transforms and neural networks applied to image retrieval. In *Proceedings of the 18th International Conference on Pattern Recognition (ICPR 2006)* (Vol. 02, pp. 909–912). Washington, DC, USA: IEEE Computer Society.

Gonzalez-Garcia, Alain C. (2007). Image retrieval based on the contents. Ph. D. thesis. Center for Research in Computing (CIC)-IPN, Mexico DF.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation on the spatial envelope. *Computer Vision, 42*(3), 145–175.

Bimbo, A. D. (1998). A perspective view on visual information retrieval systems, content based access of image and video libraries. IEEE Workshop on Volume 21 IEEE 1998, 108–109.

Bosch, A., Muñoz, X., & Martí, R. (2007). Review: Which is the best way to organize/classify images by content? *Image and Vision Computing, 25*, 778–791.

Bosch, A., Zisserman, A., & Muoz, X. (2008). Scene classification using a hybrid generative/discriminative approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30*, 712–727.

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys, 40*, 5:1–5:60.

Diaz, M., & Sturm, P. (2009). Finding images with similar lighting conditions in large photo collections. In *Proceedings of the 14th Iberoamerican conference on pattern recognition: Progress in pattern recognition, image analysis, computer vision, and applications* (pp. 53–60). Berlin, Heidelberg: Springer-Verlag.

ElAlami, M. (2011). Unsupervised image retrieval framework based on rule base system. *Expert Systems with Applications, 38*, 3539–3549.

Fei-Fei, L., Fergus, R., & Perona., P. (2004). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. IEEE CVPR 2004, Workshop on Generative-Model Based Vision.

Feng, D., Siu, W. C., & Zhang, H. J. (2010). *Multimedia information retrieval and management: technological fundamentals and applications*. *Signals and communication technology* (1st ed.,). Springer Publishing Company, Incorporated.

Fukunaga, K. (1990). *Introduction to statistical pattern recognition* (2nd ed.). San Diego, CA, USA: Academic Press Professional, Inc..

Lazebnik, S., Schmid, C., & Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer vision and pattern recognition, 2006 IEEE computer society conference on* (pp. 2169–2178).

Li, J., & Wang, J. Z. (2006). Real-time computerized annotation of pictures. In *Proceedings of the 14th annual ACM international conference on multimedia* (pp. 911–920). New York, NY, USA: ACM.

Manjunath, B., & Ma, W. (1996). Texture features for browsing and retrieval of image data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 18*, 837–842.

Marceau, D. J., Howarth, P. J., & Gratton, D. J. (1990). Evaluation of grey level co-ocurrency matrix method for land classification using spot imagery. *IEEE Transactions on Geoscience and Remote Sensing, 28*(4), 513–519.

Hiremath, P. S., & Pujari, J. (2007). Content based image retrieval using color, texture and shape features. In *15th International conference on advanced computing and communications* (pp. 780–784).

Rui, Y., Huang, Th. S., & Chang, S. F. (1999). Image retrieval: Currente techniquies, promissing directions, and open issues. *Journal of Visual Communication and Image Representation, 10*, 39–62.

Schmid, C. (2004). Weakly supervised learning of visual models and its application to content-based image retrieval. *International Journal of Computer Vision, 56*(12), 7–16.

Serrano, J., Avilés, C., Sossa, H., Villegas, J., & Olague, G. (2010). Unsupervised image retrieval with similar lighting conditions. In *Pattern Recognition (ICPR), 2010 20th International Conference on* (pp. 4368–4371).

Serrano, J., Sossa, J., Avilés, C., Barrón, R., Olague, G., & Villegas, J. (2009). Scene retrieval of natural images. In E. Bayro-Corrochano & J. O. Eklundh (Eds.), *Progress in pattern recognition, image analysis, computer vision, and applications*. *Lecture Notes in Computer Science* (Vol. 5856, pp. 774–781). Berlin/Heidelberg: Springer.

Su, W., Chen, J., & Lien, J. (2010). Region-based image retrieval system with heuristic pre-clustering relevance feedback. *Expert Systems with Applications, 37*, 4984–4998.

Subrahmanyam, M., Maheshwari, R., & Balasubramanian, R. (2012). Expert system design using wavelet and color vocabulary trees for image retrieval. *Expert Systems with Applications, 39*, 5104–5114.

Sumana, I., Islam, M., Zhang, D., & Lu, G. (2008). Content based image retrieval using curvelet transform. In *Multimedia Signal Processing, 2008 IEEE 10th Workshop on* (pp. 11–16).

Vogel, J. (2004). Semantic scene modeling and retrieval. Ph. D. thesis. Swiss Federal Institute of technology Zurich, Zurich Germany.

Vogel, J., & Schiele, B. (2007). Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision, 72*(2), 133–157.

Vogel, J., & Schiele, B. (2006). Performance evaluation and optimization for content-based image retrieval. *Pattern Recognition, 39*(5), 897–909.

Vogel, J., Schwaninger, A., Wallraven, C., & Bulthoff, H. H. (2006). Categorization of natural scenes: Local vs. global information. In *Proceedings of the 3rd symposium on applied perception in graphics and visualization* (pp. 33–40). New York, NY, USA: ACM.

Yildizer, E., Balci, A. M., Hassan, M., & Alhajj, R. (2012). Efficient content-based image retrieval using multiple support vector machines ensemble. *Expert Systems with Applications, 39*, 2385–2396.

Liu, Y., Zhang, D., et al. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition, 40*, 262–282.

Zhu, X. D., & Huang, Z. Q. (2008). Conceptual modeling rules extracting for data streams. *Knowledge-Based Systems, 21*, 934–940.