

Research Article

Topology: A Theory of a Pseudometric-Based Clustering Model and Its Application in Content-Based Image Retrieval

I. Osuna-Galán ¹, Y. Pérez-Pimentel ¹,
Carlos Avilés-Cruz ², and Juan Villegas-Cortez ²

¹Polytechnic University of Chiapas. Carretera Tuxtla Gutierrez., Portillo Zaragoza Km. 21 + 500, Suchiapa, Chiapas, Mexico

²Universidad Autónoma Metropolitana, Unidad Azcapotzalco. Departamento de Electrónica. Av. San Pablo 180 Col. Reynosa Tamaulipas, C.P. 02200, Ciudad de México, Mexico

Correspondence should be addressed to Carlos Avilés-Cruz; cavilesc@gmail.com

Received 24 October 2018; Revised 10 December 2018; Accepted 16 December 2018; Published 21 January 2019

Academic Editor: Benjamin Ivorra

Copyright © 2019 I. Osuna-Galán et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The clustering problem has been extensively studied over the last 50 years; however, it still has the attention of researchers. This paper presents a topological basis of a pseudometric-based clustering model which takes into account the local and global topological properties of the data to be clustered, as per the definition of homogeneity measurement. The proposed approach takes into account the homogeneity effect produced when a new particle is added to a group. The additional element can be accumulated in the group if its local homogeneity is not altered and, therefore, it is not necessary to carry out tests in another group. A new group needs to be generated if the threshold of the local homogeneity of the group exceeds. Theoretical results, their implementation, and their application to the problem of Content Based Image Retrieval (CBIR) are presented. The tests were performed using three image databases widely used in the literature, which are “Vogel and Shiele,” “Oliva and Torralba,” and “L. Fei- Fei, R. Fergus and P. Perona.” The results are presented and compared with the most competitive methods available in the literature.

1. Introduction

Nowadays, the large amount of data on the Internet requires grouping or clustering to obtain the relevant information from them. The typical clustering algorithms are extensively used in the areas of data sciences, data mining, and pattern recognition for grouping the information having common characteristics or defining the optimal number of groups.

Literature reports many clustering paradigms among which the most important can be categorized into Partitional Clustering [1–5] and its variants [6, 7], Hierarchical Clustering [1, 3, 8–12], Density-based Clustering [1, 3, 13–16], Grid-based Clustering [1, 3, 17–20], Spectral Clustering [1, 3, 21–24], and Gravitational Clustering [25–29]. The literature review on clustering is given in the in the next section.

In this research paper, a new clustering theory inspired of the thermodynamics principle of energy and based on the topological paradigm is presented. The proposal works in accordance with the homogeneity effect that occurs when

a group receives a new particle. A test in another group can be avoided if the new element does not alter the local homogeneity of the group in which it is added. On the contrary, when the threshold value of the group is surpassed, a new group must be generated. The process continues until all the elements are assigned groups. Subsequently, the proposal is applied to the Content-Based Image Retrieval (CBIR) technique in databases of natural scenery images.

The CBIR technique provides a support-oriented tool for image understanding where the aim is to associate and recognize an image only by its content. Research and development in CBIR technique enabled to make its place in the market in the form of co-marketing products such as QBIC-IBM (<http://www.qbic.almaden.ibm.com/>), VisualSEEL (<http://www.ee.columbia.edu/ln/dvmm/researchProjects/Multimedia-Indexing/VisualSEEK/VisualSEEK.htm>), and MARS (<https://www.ideals.illinois.edu/handle/2142/25947>). The increasing attention given and search for efficient methods have made it an active research area [30–33].

Regardless of the purpose of the users, each type of image has its own specific problems in terms of recognition and classification. Therefore, CBIR based techniques are designed to focus on one type of image with the natural scenery being one of the most complicated due to the mixture of irregular shapes and colors.

Our novel approach was tested and validated using three databases widely used in CBIR literature, which are described below:

- (i) Vogel and Shiele (VS) [34]: 700 images classified as 144 coast, 103 forest, 179 mountain, 131 prairie, 111 river/lake, and 32 sky/cloud.
- (ii) Oliva and Torralba (OT) [35]: 1472 images classified as 360 coast, 328 forest, 374 mountain, and 410 prairie.
- (iii) L. Fei-Fei, R. Fergus, and P. Perona (FP) [36]: 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, and FP (Caltech-101).

Our results are compared with those of [34, 37–39]. We would like to point out that our proposal achieves a recognition percentage of 100%.

The system is implemented using the LabVIEW programming language (Laboratory Virtual Instrument Engineering Workbench) [40]. Due to the graphical ease of programming and the facility to interface input-output devices, this aspect is considered important because commercial computer-based devices can use this software. Visualizing in the future to implement an autonomous system of recognition of natural scenes mounted on a car, which will be managed by LabView as autonomous system [41–43]. The fact of recognizing natural scenarios in the navigation of an autonomous car or possibly a drone, with a 100% certainty, this proposed system will be an important element in the safety of autonomous vehicles.

The rest of the paper is organized as follows. In Section 2, related works are discussed. The proposed approach is presented in Section 3. A detailed description of the topology-based theory is given in Section 4. Section 5 presents the system methodology. Experimental results are given in Section 6. Finally, conclusions and future works are given in Section 7.

2. State-of-the-Art

Clustering algorithms are suitable for identifying and separating large databases into classes and they are among the major pattern recognition techniques. A clustering algorithm is expected to divide the set of features into the subsets which optimizes the intra-subset similarity and inter-subset dissimilarity, where a similarity measure is defined beforehand.

An extensive use of clustering algorithms is made in areas of data science and data mining, where the objective is to group information that has common characteristics, as well as, to define the optimal number of groups.

There are survey works giving a landscape on clustering techniques [44–50], where, most of the works divide clustering algorithms on the basis of their paradigms. Thus, the most

important clustering paradigms reported in the literature are: Partitional Clustering, Hierarchical Clustering, Density-based Clustering, Spectral Clustering, and Gravitational Clustering, which are summarized in Table 1. Despite the large number of clustering algorithms, one does not work 100% for all types of clustering problems. The algorithms reported in the literature have been developed and adopted for different data to be clustered. The fundamental problem in the previously proposed algorithms reported in the literature is the overlooking of the topology of the data, both locally and globally, which limits them to work only for the data they analyze.

This proposal proposes a homogeneity measure that takes into account the local and global topological properties of the data to be clustered (see Sections 3 and 4).

Our proposal is an attempt to overcome the main limitations of typical clustering algorithms. With the topological model based on pseudometric, unlike partitional and gravitational paradigms, we no longer have the need to establish prior knowledge of the data, thus, it is not necessary to define any density-based function and the use of working only with the hyper-spherical separation functions is avoided. Therefore, based on Algebra of sets, in addition to not occupying excessive memory as the density based, spectral, and gravitational paradigms do, the calculations and results are faster. This research papers aims to compare the clustering and classification to databases of natural sceneries through CBIR methodology of different paradigms in terms of performance. The advantages and disadvantages of different clustering algorithms are beyond the scope of this research paper.

Concerning the most competitive CBIR works leading the same natural databases and scenarios are

- (i) Serrano [39]: used *k-means* algorithm (partitional paradigm) as clustering technique and *k-nearest neighbor K-NN* classifier.
- (ii) Bosch [37]: used *Probabilistic Latent Semantic Analysis (PLSA)* which is based on a density function (density-based and partitional paradigms). Authors used *K-NN* and support vector machine *SVM* classifiers.
- (iii) Lazebnik [38]: used multiresolution grouping methodology based on histograms (partitional paradigm) and a *SVM* classifier.
- (iv) Vogel [34]: used *grid segmentation* (partitional paradigm), *K-NN*, and *SVM* classifiers.

Authors take into account eleven classes: *coast, forest, mountain, prairie, schooner, bonsai, lotus flower, aquatic flower, sky, cloud, and Joshua's tree*. The classification performance of the authors does not reach 100% (about 90%). Throughout our clustering method, 100% is reached.

3. Proposal

The traditional models categorize the images of different natural sceneries (forest, desert, mountain, river, etc.) by recognizing the necessary and sufficient characteristics of that

TABLE 1: Comparing existing clustering paradigms.

Clustering paradigm	Related works Works	Limitations
<i>Partitional</i>	[1–7]	(i) Only suitable for hyper-spherical separation. (ii) Optimization problem (non-guaranteed minimum value).
<i>Hierarchical</i>	[1, 3, 8–12]	(i) Suitable for small databases. (ii) Problem to find optimal number of groups.
<i>Density-Based</i>	[1, 3, 13–16]	(i) Main disadvantage is defining the priori density function. (ii) Suitable for databases following a given density function. (iii) Computationally heavy.
<i>Spectral</i>	[1, 3, 21–24]	(i) Frequency domain transformation. (ii) Preserving time-frequency correspondence. (iii) Optimization problem (non-guaranteed minimum value). (iv) Computationally heavy.
<i>Gravitational</i>	[25–29]	(i) Suitable for small databases. (ii) Only suitable for hyper-spherical separation. (iii) Optimization problem (non-guaranteed minimum value). (iv) Computationally heavy.

particular image or images. However, there is little consensus on what these characteristics should be.

The clustering process reflects the way visual information is used, associating common elements between different images. Therefore, the hypothesis is the way to associate and create the categories of natural sceneries using mathematical functions which measure association and the decision is made on the basis of that assessment. The approach developed in this research paper compares the categorization patterns and decides whether or not the images belong to the same category.

Therefore, the following model of human thought is considered. The way in which different objects are grouped by a person is by evaluating the attributes and properties of the object in a given class. This is done by generating association values which are based on one function; the affinity function which measures how close one image is to another. The individual generates values and decides the convenience of associating an object to a group and thus form partitions.

The principle of homogeneity states that, for a closed system, internal homogeneity maximum is reduced to a minimum value in equilibrium. Following this line of thought, if we associate the ensembles with a homogeneity function, which is obtained in principle with the affinity function, a group of images will be formed by common elements if they are kept below their homogeneity level.

A subset of a group may be designated, which we will call *representatives*. A good way of associating an object to a group is when there is affinity between that new object and the group of *representatives*, that is, when the object added to the group keeps its homogeneity level low, by maintaining the qualities of that group and therefore the association of new objects will be clearer. In such case, it is sought that even with the least affinity that is on the object, it remains below the level of homogeneity. If it turns out that the object has no relation to this grouping, then this leads to a rejection, and a new group is created in which the elements are labelled as ‘other’ because they have no similar properties to each other or to the created groups.

4. Topological-Based Clustering Theory

Definition 1. Let X be a nonempty set. The power set 2^X is the set of all subsets of X . The Cartesian product $X \times Y$ of sets X and Y is the set of all ordered pairs (x, y) with $x \in X$ and $y \in Y$ [51].

A partition $P(X)$ of a set X is a set of nonempty subsets of X such that every element $x \in X$ is in exactly one of these subsets. Equivalently, a family of sets $P(X) = \{P_i \mid i \in I\}$ is a partition of X if and only if the following conditions are satisfied:

- (i) For each $i \in I, P_i \subset X$.
- (ii) For $i, j \in I$ and $i \neq j$ implies $P_i \cap P_j = \emptyset$.
- (iii) $\bigcup_{i \in I} P_i = X$.

Definition 2. A pseudometric of a set X is a function $\rho : X \times X \rightarrow [0, \infty)$ satisfying the conditions defined by (1), for all x, y , and $z \in X$.

$$\begin{aligned} \rho(x, y) &= \rho(y, x) \\ \rho(x, y) &\leq \rho(x, z) + \rho(z, y) \\ \rho(x, x) &= 0 \end{aligned} \quad (1)$$

The pair (X, ρ) is called a pseudometric space.

Besides, for all $x \in X$ and $\lambda \geq 0$, an open-ball function of radius λ and around x is defined as $B_\rho(x, \lambda) = \{x \in X : \rho(x, k) < \lambda\}$.

Definition 3. Let (X, ρ) be a pseudometric space and A be a subset of X . If an element $r \in A$ satisfies the condition (2), then r is called representative of A , denoted by $R(A)$.

$$\sum_{y \in A \setminus \{r\}} \rho(r, y) \leq \sum_{y \in A \setminus \{x\}} \rho(x, y) \quad (2)$$

For all $x \neq y, x, y \in X$.

Definition 4. Let (X, ρ) be a pseudometric space and A be a subset of X with $R(A) = \{x \in A : x \text{ is representative of } A\}$; then the distance between A and x is defined as

$$\rho(A, x) = \min \{\rho(r, x) : r \in X\} \quad (3)$$

Also, a homogeneity function $H(A)$, with $H : 2^X \rightarrow R$, is defined as

$$H(A) = \frac{1}{|A| - 1} \sum_{x \neq r} \rho(x, r) \quad (4)$$

where $r \in R(A)$.

It is noted that the calculation of the homogeneity of a set A is independent of the choice of the representative of A . By definition, if r_1 and r_2 are representative of A , then, the following condition must be satisfied:

$$\sum_{y \in A \setminus \{r_1\}} \rho(r_1, y) = \sum_{y \in A \setminus \{r_2\}} \rho(r_2, y) \quad (5)$$

A point p is related to A if $H(A \cup \{p\}) \leq H(A)$.

Definition 5. Let (X, ρ) be a pseudometric space and $\delta \geq 0$. A partition $P(X) = \{P_i : i \in I\}$ of X is called a δ -partition if $H(P_i \leq \delta)$ for each $i \in I$.

Definition 6. Let (X, ρ) be a pseudometric space, A be a subset of X , and $R(A)$ be the set of representatives of A . The set (6) given below is called a λ -cover of A .

$$\text{star}(A, \lambda) = \bigcup_{r \in R(A)} B_\rho(r, \lambda) \quad (6)$$

Lemma 7. Let (X, ρ) be a pseudometric space. For each finite set $A \in 2^X$, the set of representatives $R(A)$ is a nonempty set.

Proof. For the set $A = \{a_i : i = 1, \dots, n\}$, considers

$$D_i = \sum_{y \in A \setminus \{a_i\}} \rho(a_i, y) \quad (7)$$

take $D^* = \min\{D(a_i) : i = 1, \dots, n\}$. There is an i^* such that $D^* = D_{i^*}$. For definition of a_{i^*} it follows that

$$\sum_{y \in A \setminus \{a_{i^*}\}} \rho(a_{i^*}, y) \leq \sum_{y \in A \setminus \{x\}} \rho(x, y) \quad (8)$$

for each $x \in A$. This proves that a_{i^*} is a representative of A . \square

Example 8. Consider the sets A, B , and C in R^2 with the Euclidean metric (see Figure 1). The elements of the set A marked with a dot are distributed in two concentric circles with radii 1, and 2. The representative of the set is the element in the center of the circle marked with a star. Consider the subsets B and C as removing a circle from A (see Figure 1). Also, it can be noted that the homogeneity of set A, B , and C is 1.5, 2, and 1, respectively.

This example shows that the homogeneity is not invariant under subsets, for $B \subseteq A$ does not imply $H(B) \leq H(A)$.

The following result shows one way to preserve the homogeneity.

Proposition 9. Let (X, ρ) be a pseudometric space, $A \in 2^X$, and $R(A)$ and $H(A)$ be the set of representatives and homogeneity of A , respectively. The λ -star cover $st(A, \lambda)$ with $\lambda = H(A)$ satisfies $H(st(A, \lambda)) \leq H(A)$.

Proof. Consider $r \in R(A)$ and $p \in B_\rho(r, H(A)) \setminus A$; then $\rho(r, p) \leq H(A)$. Let $A_1 = A \cup \{p\}$.

$$\begin{aligned} H(A_1) &= \frac{1}{|A_1| - 1} \sum_{y \in A_1 \setminus \{r\}} \rho(y, r) \\ &= \frac{1}{|A_1| - 1} \sum_{y \in A \setminus \{r\}} \rho(y, r) + \frac{1}{|A_1| - 1} \rho(p, r) \\ &\leq \frac{1}{|A|} \sum_{y \in A \setminus \{r\}} \rho(y, r) + \frac{1}{|A|} H(A) \end{aligned}$$

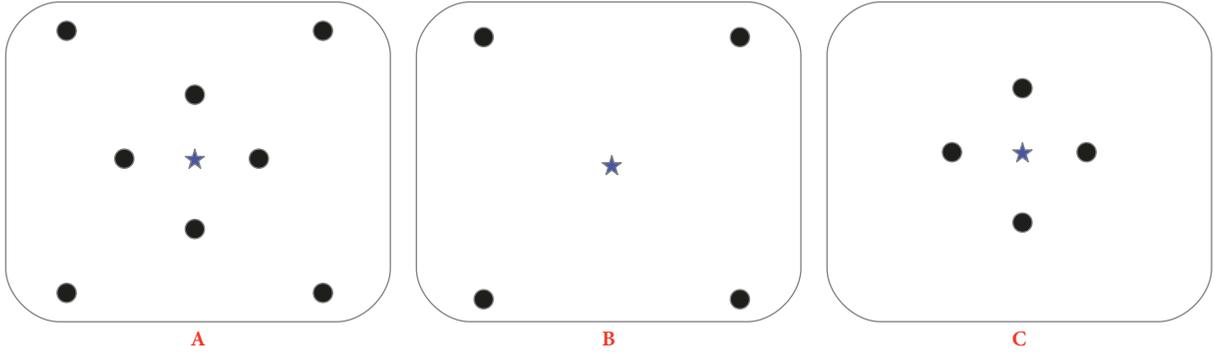


FIGURE 1: Example of homogeneity and subsets.

$$\begin{aligned}
 &= \frac{|A| - 1}{|A|(|A| - 1)} \sum_{y \in A \setminus \{r\}} \rho(y, r) + \frac{1}{|A|} H(A) \\
 &= \frac{|A| - 1}{|A|} H(A) + \frac{1}{|A|} H(A) = H(A)
 \end{aligned} \tag{9}$$

It can construct the succession $\{A_j : j = 1, \dots, n\}$ which satisfied $H(A_m) \leq H(A_n)$ for $n \leq m$ and that $A_j \rightarrow st(A, \lambda)$. Therefore $H(st(A, \lambda)) \leq H(A)$. \square

Theorem 10. Let (X, ρ) be a pseudometric space and $\delta \geq 0$. There exists a partition $P(X) = Q \cup \{P_i : i \in I\}$ of X , where $\{P_i : i \in I\}$ is δ -partition and the element Q are not related to P_j for each $j \in I$.

Proof. It will construct a family of sets $\{F_i\}_{i=1}^k = \{Q_k \cup \{P_1, P_2, \dots, P_k\}\}_{i=1}^k$ which satisfies the following properties for each $i, j = 1, \dots, n$:

$$\begin{aligned}
 (i) \quad &P_i \cap P_j = \emptyset \quad \forall i \neq j \\
 (ii) \quad &Q_k \cap P_j = \emptyset \quad \forall j = 1, \dots, k \\
 (iii) \quad &H(P_i) \leq \delta
 \end{aligned} \tag{10}$$

Take $x_1, x_2 \in X$. If $\rho(x_1, x_2) < \delta$ then $P_1 = \{x_1, x_2\}$. Otherwise, define $Q_1 = \{x_1, x_2\}$. It can assume without loss of generality that there exists n such that condition $\rho(x_m, x_n) < \delta$ is satisfied for $m \leq n$. Let $P^1 = \{x_m, x_n\}$ and $Q_1 = \{x_1, x_2, \dots, x_{m-1}, x_{m+1}, \dots, x_{n-1}\}$; then take $P_1 = st(P^1, \delta)$ and consider $F_1 = Q_1 \cup \{P_1\}$. By Proposition 9 the following inequalities hold:

$$H(P_1) = H(st(P^1, \delta)) \leq H(P^1) < \delta \tag{11}$$

Therefore condition $P_1 = st(P^1, \delta)$ is satisfied. The properties (i) and (ii) are fully satisfied.

Suppose that, for $k \in N$, a succession of families of sets $F_k = Q_k \cup \{P_1, P_2, \dots, P_k\}$ has been constructed satisfying the conditions (i)-(iii).

Let $Y = X \setminus (\bigcup_{k=1}^K F_k)$ and take $y \in Y$.

Case I (y is not related to any P_i). Define $Q^k = B_\rho(y, \delta) \cap Q_k$.

If $Q^k = \emptyset$ then $Q_{k+1} = Q_k \cup \{y\}$, and $F_{k+1} = Q_{k+1} \cup \{P_1, P_2, \dots, P_k\}$. By definition of Q^k the property $Q_{k+1} \cap P_j = \emptyset$, therefore property (ii) is satisfied. The property (i) and (ii) are preserved due to the family of sets $\{P_1, P_2, \dots, P_k\}$ is the same in F_{k+1} and F_k .

If $Q^k \neq \emptyset$ then $P_{k+1} = Q^k$, $Q_{k+1} = Q^k \setminus Q^k$, and $F_{k+1} = Q_k \cup \{P_1, P_2, \dots, P_k\}$.

Note $Q_{k+1} \subseteq Q^k$ and $P_{k+1} \subseteq X \setminus (\bigcup_{j=1}^k P_j)$. Properties (i) and (ii) are easily verified. y is a P_{k+1} the condition (iii) is satisfied.

Case II (there exists $i \leq k$ such that y is related to P_i). Consider the set of indexes $j = \{j \in N : y \text{ is related to } P_j\}$ and let $i^* = \min\{\rho(P_i, y) : i \in I\}$. Defines $F_{k+1} = Q_k \cup \{P_1, P_2, \dots, P_{i^*} \cup \{y\}, \dots, P_k\}$. Since y is related to P_{i^*} the inequality $H(P_{i^*} \cup \{y\}) \leq H(P_{i^*}) \leq \delta$ implies property (iii). Note that $y \in X \setminus (F_k)$. Therefore properties (i) and (ii) are satisfied.

By induction, there exists a decomposition of the set X of the form $P(X) = Q \cup \{P_i : i \in I\}$, where the family of sets $\{P_i : i \in I\}$ satisfies with $H(P_i) \leq \delta$ for each $j \in I$. The theorem is proved. \square

Observe that the theorem indicates an algorithm which allows to create partitions of a set from a pseudometric and a parameter δ . This partition consists of a family of sets $\{P_i : i \in I\}$ whose elements are clusters. Elements of the set Q do not belong to any class.

5. Methodology

In this section, we describe each stage of the proposed methodology (applying our topological clustering theory based on a pseudometric) for image retrieval of natural scenes from a database. As shown in Figure 2, the flow diagram for the whole proposal, there is a training phase and a testing phase explained as follows.

5.1. Training Stage. This stage is divided into four main phases: *Image database*, *Histogram estimation*, *Calculate the distance*, and *Add the object into the suitable cluster*, see Figure 2 upper line, which are described below.

5.1.1. Image Database. To develop a CBIR system, it is necessary to use databases of images for the training of the

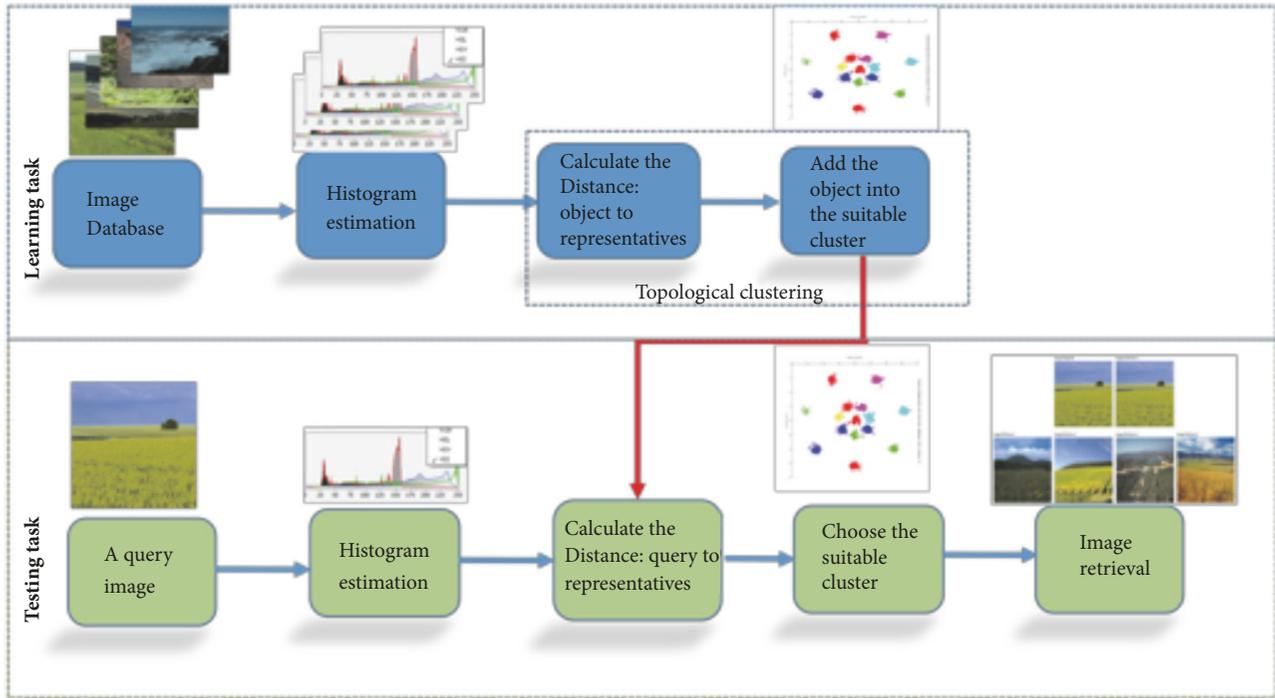


FIGURE 2: Flow diagram for the training and testing stage.

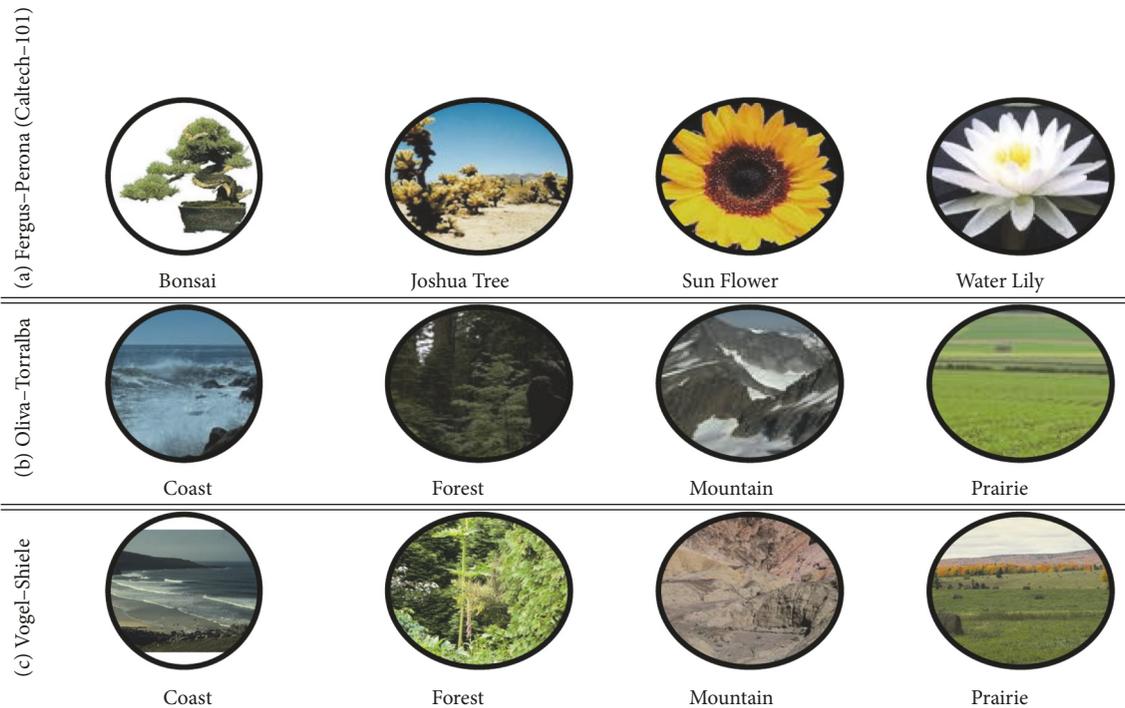


FIGURE 3: Categories for image organization and retrieval for each of the three databases: (a) FP, (b) OT, and (c) VS respectively.

system. In this phase, standard and free-use databases such as Vogel and Shiele (VS) [34], Oliva and Torralba (OT) [35], and L. Fei-Fei, R. Fergus, and P. Perona (FP) [36] can be used.

The classes used in both the training and the experimentation phases are shown in Figure 3.

5.1.2. Histogram Estimation. The histogram provides a summary of the distribution of pixel values in an image. The color histogram of an image is relatively invariant with respect to translation and rotation on the axes. The comparison of the colors contained in two images can be made

using histograms. The color histograms are suitable for this investigation [52–54]. A histogram represents the number of occurrences of the values of a data set (see the following):

$$H_i(A) = \sum_{j=1}^n C_{ij} \quad \text{where } c_{ij} = \begin{cases} 1 & \text{if } a_i = x_j \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

A good practice is to perform normalization. It is usually represented as a density probability function. This preprocessing task is carried out to bring the values to an interval regardless of the number of elements they have.

$$H^*(A) = \frac{H(A)}{\sum_{j=1}^n H_j(A)} \quad (13)$$

In this part, the values of the image are still expressed in RGB color space. We also use HSI (Hue, Saturation, and Intensity) values. The first step for the conversion is to normalize the RGB values, as

$$\begin{aligned} r &= \frac{R}{R+G+B}; \\ g &= \frac{G}{R+G+B}; \\ b &= \frac{B}{R+G+B} \end{aligned} \quad (14)$$

After normalization, the HSI components are obtained by

$$\begin{aligned} h &= \cos^{-1} \left(\frac{0.5 [(r-g)(r-b)]}{[(r-g)^2 + (r-b)(g-b)]^{(1/2)}} \right) \\ &\quad \forall h \in [0, \pi], \quad b \leq g \\ h &= 2\pi - \cos^{-1} \left(\frac{0.5 [(r-g)(r-b)]}{[(r-g)^2 + (r-b)(g-b)]^{(1/2)}} \right) \\ &\quad \forall h \in [\pi, 2\pi], \quad b > g \end{aligned} \quad (15)$$

$$s = 1 - 3 \times \min(r, g, b) \quad \forall s \in [0, 1]$$

$$i = \frac{R+G+B}{3 \times 255} \quad \forall i \in [0, 1]$$

For convenience, for neutral color values such as black, white, and gray, where values for R, G, and B are equal, H = 0. Similarly, the H, S, and I values are converted into the following ranges:

$$\begin{aligned} H &= \frac{h \times 180}{\pi} \\ S &= s \times 100 \\ I &= i \times 255 \end{aligned} \quad (16)$$

5.1.3. Calculating the Distance. Consider the set Φ of images. There is an association of an image P to the corresponding

histogram $H(P)$. For two images P_1 and P_2 , the following relationship is defined:

$$\begin{aligned} \rho &: \Phi^2 \longrightarrow R \\ \rho(P_1, P_2) &= d(H(P_1), H(P_2)) \end{aligned} \quad (17)$$

In this case the function $\rho : \Phi^2 \longrightarrow R$ is a pseudometric.

Take P_1, P_2 , and P_3 images in Φ and $H(P_1), H(P_2)$, and $H(P_3)$ with their corresponding histograms. The properties of Definition 2 are fulfilled:

$$\begin{aligned} \text{(i)} \quad \rho(P_1, P_1) &= d(H(P_1), H(P_1)) = 0 \\ \text{(ii)} \quad \rho(P_1, P_2) &= d(H(P_1), H(P_2)) \\ &= d(H(P_2), H(P_1)) = \rho(P_2, P_1) \\ \text{(iii)} \quad \rho(P_1, P_2) &= d(H(P_1), H(P_2)) \\ &= d(H(P_2), H(P_1)) \\ &\leq d(H(P_1), H(P_3)) \\ &\quad + d(H(P_3), H(P_2)) \\ &= \rho(P_1, P_3) + \rho(P_3, P_2) \end{aligned} \quad (18)$$

A pseudometric has been built in the space of the images. Theorem 10 will allow creating clusters and perform image recovery. For this study, we consider using the following distance functions [55].

Intersection:

$$I(x, y) = \sum_{i=1}^n \min(x_i, y_i) \quad (19)$$

Chi-Square:

$$\chi^2(x, y) = \frac{1}{2} \sum_{i=1}^n \frac{(x_i - y_i)^2}{(x_i + y_i)} \quad (20)$$

Correlation:

$$\text{Corr}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (21)$$

The calculation of the distances must be modified in order to eliminate the indeterminate forms. If the direct substitution produces an indeterminate form then these values are discarded.

5.1.4. Add the Object into the Suitable Cluster. Having the *representatives* of each cluster, the homogeneity of the element to be added to the *representatives* is calculated. Through *Theorem 10* and *Proposition 9* of Section 4, homogeneities are calculated for each representative. Then, the subject image is assigned to the cluster where homogeneity is not altered.

TABLE 2: Eleven classes for learning and testing tasks.

Class	Number of images
<i>coast</i>	504
<i>forest</i>	431
<i>mountain</i>	553
<i>prairie</i>	541
<i>schooner</i>	350
<i>bonsai</i>	128
<i>lotus flower</i>	64
<i>aquatic flower</i>	36
<i>sky</i>	32
<i>cloud</i>	32
<i>Joshua's tree</i>	60
<i>Total</i>	2,731

5.2. *Testing Stage.* This stage is divided into five main tasks and proceeds as depicted in Figure 2 bottom line. A query image is presented to the system, and the same two first processing steps used for learning are applied to it: *histogram estimation* and *Calculate the Distance: query to representatives*. At this point, theorem one is applied measuring through a defined pseudometric. By *Proposition 9* of Section 4, homogeneities are calculated for each cluster defined in learning phase. Then, the query image is assigned to the cluster where homogeneity is not altered. Therefore, the n -images of the cluster are recovered.

6. Experimental Results

In order to test the whole performance of this proposal, 3 databases were concatenated and then divided by classes. The division was done in eleven classes (See Table 2) because natural sceneries are one of the most complicated due to the textures, the mixture of irregular shapes, and color variety.

All tests were carried out under “Cross-validation” method, which consists of random selection of half images set apart from database of each class for training task, and the other half part for testing task. This procedure is carried out during 20 iterations and, finally, an average value is obtained of retrieval percentages.

The system was implemented using the LabVIEW programming language, which is a computer systems engineering software that facilitates testing, measurement, and provides control with quick access to hardware and data information. The programs developed with LabVIEW are called Virtual Instruments, or VIs, and their origin came from instrument control, although today it has expanded widely not only to control all types of electronics but also to the embedded programming, communications, mathematics, etc. Its main characteristic is the ease of use for professional programmers. Relatively complex programs can be made and the facility to interface input-output devices. Visualizing in the future is to implement an autonomous system of recognition of natural scenes mounted on a car or a drone, which will be managed by LabView as autonomous system [41–43]. The fact of recognizing natural scenarios in

the navigation of an autonomous car or possibly a drone, with a 100% certainty, contributes to the safety of autonomous vehicles.

The developed LabVIEW application screenshot is shown in Figure 4 where the query image, the main feature of the query image which is the histogram, and the retrieved image can be seen at the top left corner, at the lower left corner, and at the center right, respectively. The most related image file can be seen at the right bottom of the figure. Also, the histogram of the query image and the histogram of the most similar images can be seen.

Several subroutines were made to integrate two main modules: training and testing tasks. Only the latter has the image display for performance and processing speed issues. For each image of the selected database, the histogram is computed and sent to a *.tdms file.

The value of parameter δ is experimentally estimated as an intermediate value between the *representatives*. For each class an average of the attributes was taken and it was defined as the *representative*. In order to determine the parameter δ , only the *representatives* were taken. Therefore, the parameter $\delta = 0.5\mu$ was defined. From now, this parameter is used for the whole clustering algorithm. It was observed that a value of parameter δ smaller restricts very much the classes of objects and created many classes that apparently were different however the classes contain common elements. A higher value of parameter δ significantly increased the size of the set, adding elements that do not have common characteristics.

As there are eleven possible classes of natural scenarios, only 3 examples of query and retrieval images are presented: Bonzai (See Figure 5(a)), Prairie (See Figure 5(b)), and Forest (See Figure 5(c)). The query image (red label in each sub figure) is presented and the system was configured to present the 5 most similar images. Note how, in the three cases of query images, the results of the 5 images are very similar to the query one. As the query image is included in the training database, consequently, in the search result, the first image found is equal to the searched one.

In order to evaluate the performance of the system created, precision and recall formulas are used [56]:

$$\begin{aligned}
 \text{Precision} &= \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \\
 \text{Recall} &= \frac{\text{Number of relevant images retrieved}}{\text{Number of relevant images in the database}}
 \end{aligned} \tag{22}$$

In the experiment, the precision vs. recovery graph is used for a query image. Graphs are made with the three distances in RGB and HSI color formats; averages were considered for each measure used [57].

The precision-recall analysis for the mountain, bonsai, and forest scenarios is presented in Figures 6, 7, and 8. It can be seen that the best results are obtained with the distance function “Intersection” for the RGB color model (dark blue line). Besides, the worst result was obtained for the “correlation” distance function in the HSI color model (green line).

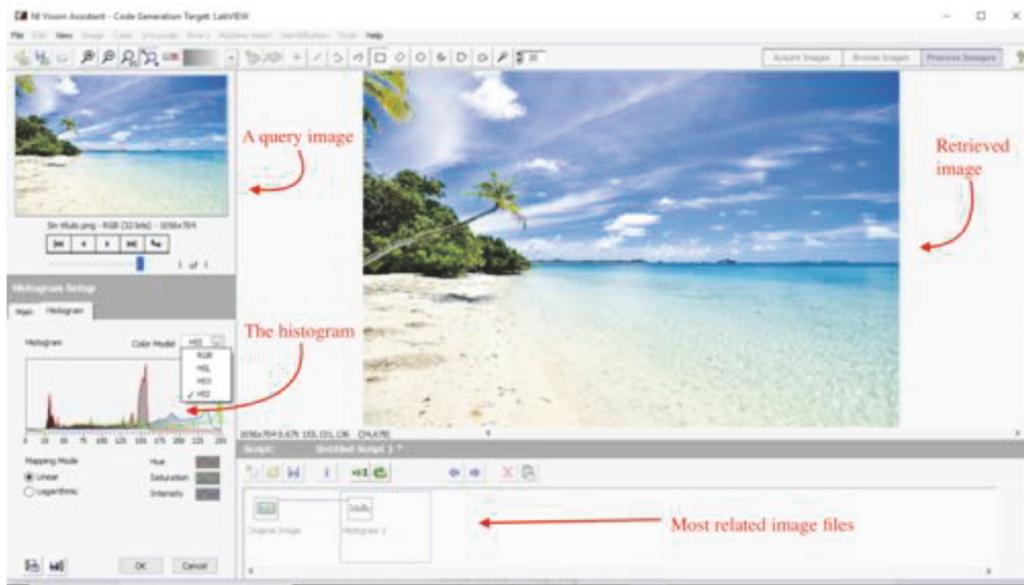


FIGURE 4: LabVIEW Vision Assistant.

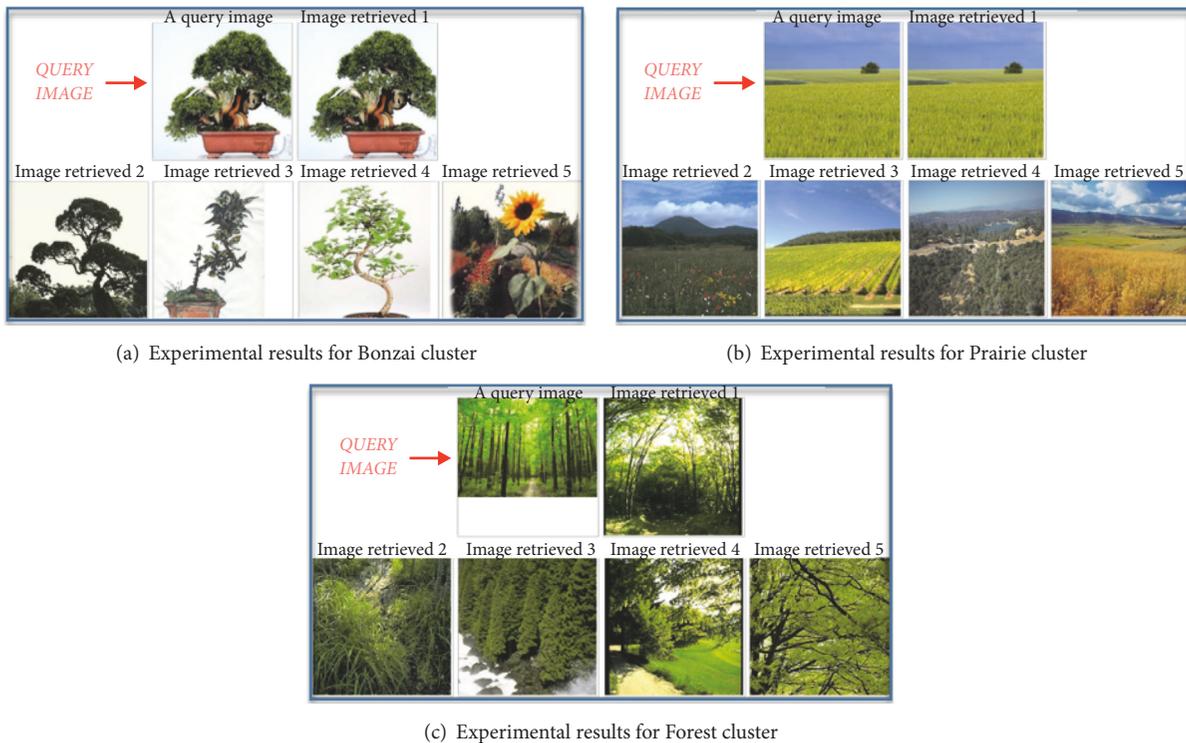


FIGURE 5: Experimental results for classes clustering.

Applying our proposal to the CBIR problem, a query image is always retrieved in the correct scenario. If the query image belongs to the training database, then it is found first. Otherwise, it is in the right scenario within the first 4 retrieved images (as an example, see Figure 5 for bonsai, prairie, and forest classes).

6.1. Comparison to Previous Classification Results: Natural Scene Classification. The comparison between the proposed method in this research paper and other competitive methods reported in the literature is focused on natural scenarios. Therefore, in the comparison, eleven classes are taken into account, coast, forest, mountain, prairie, schooner, bonsai,

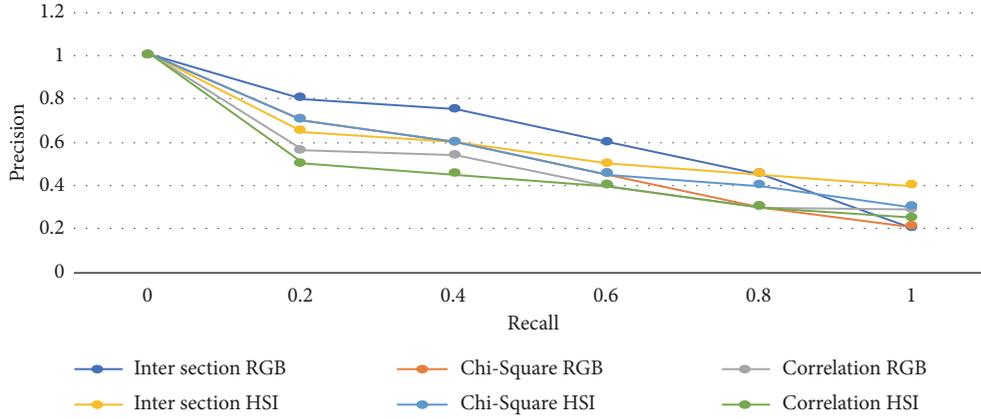


FIGURE 6: CBIR Effectiveness (Mountains, average).

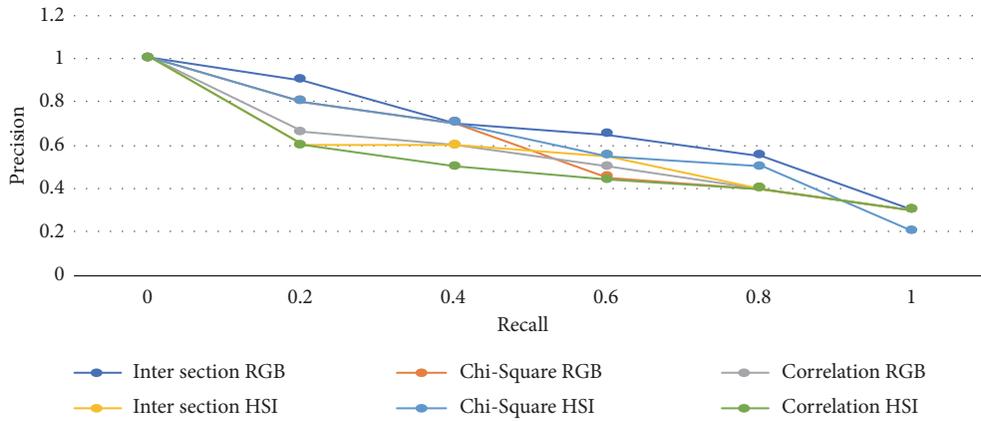


FIGURE 7: CBIR Effectiveness (Bonzai, average).

TABLE 3: Comparison values against the most competitive methods.

	Our proposal	Serrano	A. Bosh	Lazebnik	Vogel
<i>Coast</i>	100	94	88	82	80
<i>Forest</i>	100	90	100	95	95
<i>Mountain</i>	100	80	89	88	90
<i>Prairie</i>	100	92	78	70	92
<i>Schooner</i>	100	0	0	79	0
<i>Bonsai</i>	100	92	0	92	0
<i>Lotus flower</i>	100	64	0	70	0
<i>Aquatic flower</i>	100	90	0	92	0
<i>Skies</i>	100	86	0	0	94
<i>Clouds</i>	100	86	0	0	94
<i>Joshua's tree</i>	100	90	0	88	0

lotus flower, aquatic flower, sky, cloud, and Joshua's tree. The most competitive results leading to the same natural databases and scenarios are Serrano [39], Bosch [37], Lazebnik [38], and Vogel [34].

As shown in Table 3 and Figure 9 where authors used the same database, the proposed method improves the results in

all cases and for the all classes. Clarifying that in Table 3 and Figure 9, the zero values of the other authors indicate that they did not work with the respective image class.

Remarking that Bosch [37] has the same performance for the *forest* class as our model, but Bosch has worked only with four classes. Besides that, Vogel [34] has percentages above 90% in four classes: *forest*, *prairie*, *skies*, and *clouds* classes; the proposed method improves the result about 6%, i.e., from 94% (mean value for the four best classification) to 100%. Finally, compared with the two authors (Lazebnik [38] and Serrano [39]), our results are far better.

Finally, comparing with articles [30, 58–60] that use other image data bases (Corel A, Corel B, and Caltech) which include some natural scenarios: costs, flowers, and mountains. It can be seen in Table 4 that this proposal is superior.

We consider that we have obtained better results than the previous published results since the proposed clustering algorithm, based on topology theory, takes into account both the local and the global topology of the data to be clustered.

7. Conclusions

The proposed theory of a pseudometric-based clustering model and its application in Content-Based Image Retrieval

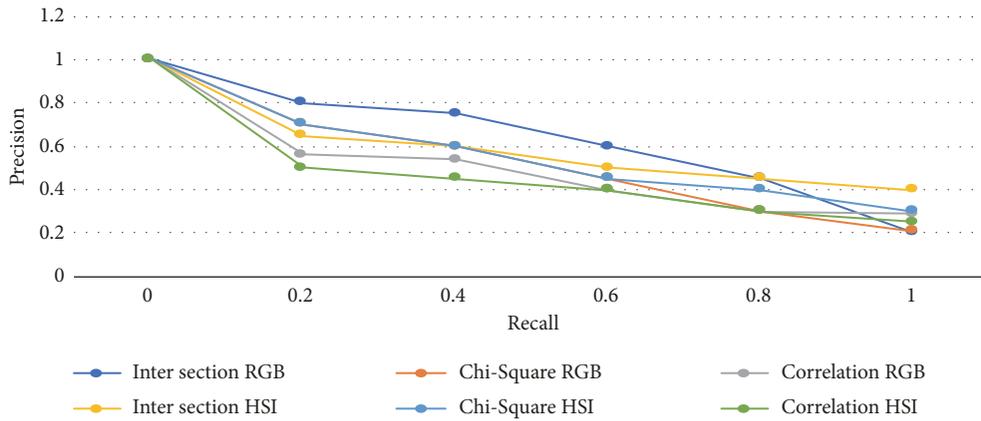


FIGURE 8: CBIR Effectiveness (Forest, average).

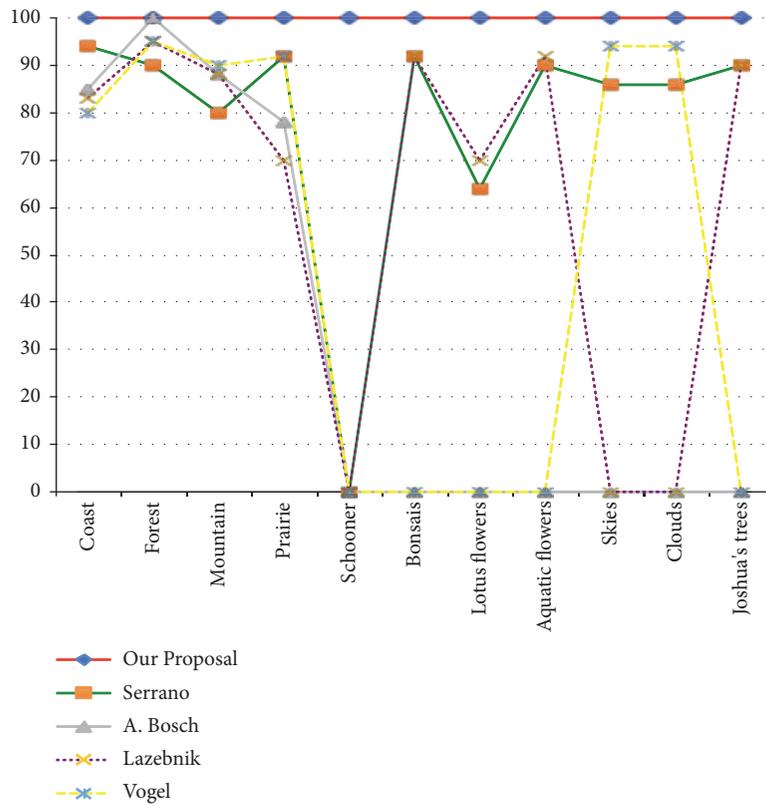


FIGURE 9: Classification performance of our proposal against the most competitive methods focused on natural scenarios.

worked as it was expected. The developed clustering method based on topology theory, has successfully operated for the clustering of natural scenery. The proposal based on a topological pseudometric, provides a paradigm within the literature of clustering algorithms.

This proposal takes into account the local and global topological properties of the data to be clustered, in a definition of homogeneity measurement.

The proposal tries to overcome the main limitations of typical clustering algorithms: (1) no longer have the need to establish an a prior knowledge of the data, thus, it is not

necessary to define any density-based function, (2) there is no need to work only with hyper-spherical separation functions (partitional and gravitational paradigms). Finally, (3) as our proposal is based on Algebra of sets, calculations and results are faster and without excessive memory consumption (as density-based, spectral and gravitational paradigms do).

Using the same image databases, the comparison with the most competitive works, the proposed method improves the results in all cases and for all the classes.

A query image is always retrieved in the correct scenario. If the query image belongs to the training database, then it is

TABLE 4: Comparison values against the most competitive methods using others DB.

	Our proposal	Mehmood-2018	Nouman	Mehmood-2016	Yousuf
Coast	100	74.58	60.39	68.23	75
Lotus flower	100	85.13	88.37	84.97	98.3
Mountain	100	82.43	61.88	87.43	83.8

found first. Otherwise, it is in the right scenario within the first 4 retrieved images. As a consequence, the whole system has an efficiency of 100% for eleven natural sceneries.

The application was deployed using the LabVIEW vision assistant. This aspect is considered important because commercial computer-based devices can use this software. The observed processing time was linear with respect to the number of elements in the database. There was a considerable reduction in the training and testing times.

Theorem 10 is not limited to images; besides in any space that defines a pseudometric, this algorithm can be used. The results obtained are very promising and future work includes the application of these functions in other sets and applications.

As future work, the authors propose developing a CBIR system in FPGA hardware or mobile device, such as a camera device to acquire the images to process in real time using this methodology. Finally, the authors have shown a real-world experiment, in which this function obtained very good results.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

I. Osuna-Galán, Y. Pérez-Pimentel, Carlos Aviles-Cruz, and Juan Villegas-Cortez contributed equally to this work.

References

- [1] R. Xu and D. Wunsch, *Clustering*, vol. 1, John Wiley and Sons, 2008.
- [2] H. Xiong, J. Wu, and J. Chen, "K-means clustering versus validation measures: a data-distribution perspective," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 2, pp. 318–331, 2009.
- [3] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, vol. 1, Academic Press, 2008.
- [4] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.
- [5] S. B. Belhaouari, S. Ahmed, and S. Mansour, "Optimized K-means algorithm," *Mathematical Problems in Engineering*, vol. 2014, Article ID 506480, 14 pages, 2014.
- [6] D. Arthur and S. Vassilvitskii, "K-means: the advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, SODA 07*, pp. 1027–1035, Philadelphia, USA, 2007.
- [7] Y. Xu, W. Qu, Z. Li, G. Min, K. Li, and Z. Liu, "Efficient k-means++ approximation with MapReduce," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 12, pp. 3135–3144, 2014.
- [8] Z. Nazari, D. Kang, M. R. Asharif, Y. Sung, and S. Ogawa, "A new hierarchical clustering algorithm," in *Proceedings of the International Conference on Intelligent Informatics and Biomedical Sciences, ICIBMS 2015*, pp. 148–152, Japan, November 2015.
- [9] F. Liu, Y. Wei, M. Ren, X. Hou, and Y. Liu, "An agglomerative hierarchical clustering algorithm based on global distance measurement," in *Proceedings of the 7th International Conference on Information Technology in Medicine and Education (ITME)*, pp. 363–367, 2015.
- [10] O. Arslan, D. Guralnik, and D. Koditschek, "Coordinated robot navigation via hierarchical clustering," *IEEE Transactions on Robotics*, vol. 32, pp. 352–371, 2016.
- [11] E. Rashedi, A. Mirzaei, and M. Rahmati, "Optimized aggregation function in hierarchical clustering combination," *Intelligent Data Analysis*, vol. 20, no. 2, pp. 281–291, 2016.
- [12] W. Yao, C. O. Dumitru, O. Loffeld, and M. Datcu, "Semi-supervised Hierarchical Clustering for Semantic SAR Image Annotation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 1993–2008, 2016.
- [13] F. Ros and S. Guillaume, "DENDIS: A new density-based sampling for clustering algorithm," *Expert Systems with Applications*, vol. 56, pp. 349–359, 2016.
- [14] S. Jahirabadkar and P. Kulkarni, "Algorithm to determine ϵ -distance parameter in density based clustering," *Expert Systems with Applications*, vol. 41, no. 6, pp. 2939–2946, 2014.
- [15] S. Yokoyama, A. Bogardi-Meszoly, and H. Ishikawa, "Ebscan, An entanglement-based algorithm for discovering dense regions in large geo-social data streams with noise," 2015.
- [16] J. E. Chacon, "A population background for nonparametric density-based clustering," *Statistical Science. A Review Journal of the Institute of Mathematical Statistics*, vol. 30, no. 4, pp. 518–532, 2015.
- [17] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan, "Automatic subspace clustering of high dimensional data for data mining applications," *SIGMOD Record*, vol. 27, no. 2, pp. 94–105, 1998.
- [18] Q. Zhao, Y. Shi, Q. Liu, and P. Fránti, "A grid-growing clustering algorithm for geo-spatial data," *Pattern Recognition Letters*, vol. 53, pp. 77–84, 2015.
- [19] M. A. Rashad, H. El-Deeb, and M. W. Fakh, "Document classification using enhanced grid based clustering algorithm," *Lecture Notes in Electrical Engineering*, vol. 312, pp. 207–215, 2015.
- [20] T. Sajana, C. M. Sheela Rani, and K. V. Narayana, "A survey on clustering techniques for big data mining," *Indian Journal of Science and Technology*, vol. 9, no. 3, pp. 1–12, 2016.

- [21] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and algorithm," in *In Advances in Neural Information Processing Systems*. MIT Press, pp. 849–856, 2001.
- [22] Y. Yang, Z. Ma, Y. Yang, F. Nie, and H. T. Shen, "Multitask spectral clustering by exploring intertask correlation," *IEEE Transactions on Cybernetics*, vol. 45, no. 5, pp. 1069–1080, 2014.
- [23] T. Inkaya, "A parameter-free similarity graph for spectral clustering," *Expert Systems with Applications*, vol. 42, no. 24, pp. 9489–9498, 2015.
- [24] R. Shang, Z. Zhang, L. Jiao, W. Wang, and S. Yang, "Global discriminative-based nonnegative spectral clustering," *Pattern Recognition*, vol. 55, pp. 172–182, 2016.
- [25] W. E. Wright, "Gravitational clustering," *Pattern Recognition*, vol. 9, no. 3, pp. 151–166, 1977.
- [26] E. Rashedi and H. Nezamabadi-Pour, "A stochastic gravitational approach to feature based color image segmentation," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 4, pp. 1322–1332, 2013.
- [27] M. B. Dowlatshahi and H. Nezamabadi-Pour, "GGSA: a grouping gravitational search algorithm for data clustering," *Engineering Applications of Artificial Intelligence*, vol. 36, pp. 114–121, 2014.
- [28] V. Kumar, J. K. Chhabra, and D. Kumar, "Automatic cluster evolution using gravitational search algorithm and its application on image segmentation," *Engineering Applications of Artificial Intelligence*, vol. 29, pp. 93–103, 2014.
- [29] H. Nikbakht and H. Mirvaziri, "A new algorithm for data clustering based on gravitational search algorithm and genetic operators," in *Proceedings of the 2015 International Symposium on Artificial Intelligence and Signal Processing, AISP 2015*, pp. 222–227, Iran, March 2015.
- [30] Z. Mehmood, S. M. Anwar, N. Ali, H. A. Habib, and M. Rashid, "A Novel image retrieval based on a combination of local and global histograms of visual words," *Mathematical Problems in Engineering*, vol. 2016, Article ID 8217250, 12 pages, 2016.
- [31] W. Zhou, H. Li, and Q. Tian, "Recent advance in content-based image retrieval: A literature survey," *CoRR*, 2017.
- [32] M. Yousuf, Z. Mehmood, H. A. Habib, and et al, "A novel technique based on visual words fusion analysis of sparse features for effective content-based image retrieval," *Mathematical Problems in Engineering*, vol. 2018, Article ID 2134395, 13 pages, 2018.
- [33] A. Varma and D. K. Kaur, "Survey on content based image retrieval," *International Journal of Engineering and Technology*, vol. 7, pp. 471–476, 2018.
- [34] J. Vogel and B. Schiele, "Semantic modeling of natural scenes for content-based image retrieval," *International Journal of Computer Vision*, vol. 72, no. 2, pp. 133–157, 2007.
- [35] A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation on the spatial envelope," *Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.
- [36] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '04)*, pp. 59–70, July 2004.
- [37] A. Bosch, A. Zisserman, and X. Muñoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 712–727, 2008.
- [38] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the Society Conference on Computer Vision and Pattern Recognition (CVPR06)*, vol. 1, p. 12, 2006.
- [39] J. F. Serrano-Talamantes, C. Avilés-Cruz, J. Villegas-Cortez, and J. H. Sossa-Azuela, "Self organizing natural scene image retrieval," *Expert Systems with Applications*, vol. 40, no. 7, pp. 2398–2409, 2013.
- [40] "National-Instruments," 2018 <http://www.ni.com/en-us/shop/labview.html>.
- [41] R. Begum and S. V. Halse, "The smart car parking system using gsm and labview," *Journal of Computer and Mathematical Sciences*, vol. 9, pp. 135–142, 2018.
- [42] S. Blaifi, S. Moulahoum, I. Colak, and W. Merrouche, "Monitoring and enhanced dynamic modeling of battery by genetic algorithm using labview applied in photovoltaic system," *Electrical Engineering*, vol. 100, pp. 1021–1038, 2018.
- [43] A. Alam and Z. A. Jaffery, "A vision-based system for traffic light detection," *Applications of Artificial Intelligence Techniques in Engineering*, vol. 1, pp. 333–343, 2018.
- [44] K. Sim, V. Gopalkrishnan, A. Zimek, and G. Cong, "A survey on enhanced subspace clustering," *Data Mining and Knowledge Discovery*, vol. 26, no. 2, pp. 332–397, 2013.
- [45] D. Xu and Y. Tian, "A comprehensive survey of clustering algorithms," *Annals of Data Science*, vol. 2, pp. 165–193, 2015.
- [46] D. Dinler and M. K. Tural, *A Survey of Constrained Clustering*, Springer International Publishing, Cham, Switzerland, 2016.
- [47] A. S. Khandare and A. Anandand, "Clustering algorithms: Experiment and improvements," in *Computing and Network Sustainability*, H. Vishwakarma and S. Akashe, Eds., pp. 263–271, Springer, Singapore, 2017.
- [48] R. S. M. L. Patibandla and N. Veeranjaneyulu, "Survey on clustering algorithms for unstructured data," in *Intelligent Engineering Informatics*, V. Bhateja, C. A. Coello Coello, S. C. Satapathy, and P. K. Pattnaik, Eds., Springer, Singapore, 2018.
- [49] M. M. A. Osman, S. K. Syed-Yusof, N. N. N. Abd Malik, and S. Zubair, "A survey of clustering algorithms for cognitive radio ad hoc networks," *Wireless Networks*, vol. 24, pp. 1451–1475, 2018.
- [50] K. Bindra, A. Mishra, and Suryakant., "Effective data clustering algorithms," in *Soft Computing: Theories and Applications*, K. Ray, T. K. Sharma, S. Rawat, R. K. Saini, and A. Bandyopadhyay, Eds., Springer, Singapore, 2019.
- [51] R. Engelking, *General Topology*, Springer International Publishing, 1989.
- [52] A. J. Afifi and W. M. Ashour, "Image retrieval based on content using color feature," *ISRN Computer Graphics*, vol. 2012, Article ID 248285, 11 pages, 2012.
- [53] G.-H. Liu and J.-Y. Yang, "Content-based image retrieval using color difference histogram," *Pattern Recognition*, vol. 46, pp. 188–198, 2013.
- [54] M. A. Stricker and M. Orengo, "Similarity of color images," W. Niblack and R. C. Jain, Eds., vol. 2420 of *Proceedings of SPIE*.
- [55] Q. Zhang and R. L. Canosa, "A comparison of histogram distance metrics for content-based image retrieval," in *Imaging and Multimedia Analytics in a Web and Mobile World*, S. P. Ed., vol. 9027, 9 edition, 2014.
- [56] J. Makhoul, F. Kubala, R. Schwartz, and R. Weischedel, "Performance measures for information extraction," in *Proceedings of the DARPA Broadcast News Workshop*, pp. 249–252, 1999.
- [57] M. Mosbah and B. Boucheham, "Matching measures in the context of cbir: A comparative study in terms of effectiveness

- and efficiency,” in *Proceedings of the Recent Advances in Information Systems and Technologies*, Á. Rocha, A. M. Correia, H. Adeli, L. P. Reis, and S. Costanzo, Eds., pp. 245–258, Springer International Publishing, Cham, Switzerland, 2017.
- [58] Z. Mehmood, M. Rashid, A. Rehman, T. Saba, H. Dawood, and H. Dawood, “Effect of complementary visual words versus complementary features on clustering for effective content-based image search,” *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, vol. 35, no. 5, pp. 5421–5434, 2018.
- [59] N. Ali, K. B. Bajwa, R. Sablatnig et al., “A Novel Image Retrieval Based on Visual Words Integration of SIFT and SURE” *PLoS ONE*, vol. 11, no. 6, pp. 1–20, 2016.
- [60] M. Yousuf, Z. Mehmood, and H. A. Habib, “A novel technique based on visual words fusion analysis of sparse features for effective content-based image retrieval,” *Mathematical Problems in Engineering*, vol. 2018, Article ID 2134395, 13 pages, 2018.



Hindawi

Submit your manuscripts at
www.hindawi.com

