Object and image indexing based on region connection calculus and oriented matroid theory

Ernesto Staffetti[a], Antoni Grau[b], Francesc Serratosa[c], Alberto Sanfeliu[d]

[a]*Department of Computer Science, University of North Carolina at Charlotte, 9201 University City Blvd., 28223 Charlotte NC, USA*
[b]*Department of Automatic Control, Technical University of Catalonia, Pau Gargallo 5, 08028 Barcelona, Spain*
[c]*Department of Computer Engineering and Mathematics, Rovira i Virgili University, Av. Països Catalans 26, 43007 Tarragona, Spain*
[d]*Institute of Industrial Robotics (CSIC-UPC), Llorens i Artigas 4-6, 08028 Barcelona, Spain*

Abstract

In this paper a novel method for indexing views of 3D objects is presented. The topological properties of the regions of views of an object or of a set of objects are used to define an index based on region connection calculus and oriented matroid theory. Both are formalisms for qualitative spatial representation and reasoning and are complementary in the sense that, whereas region connection calculus characterize connectivity of couples of connected regions of views, oriented matroids encode relative position of disjoint regions of views and give local and global topological information about their spatial distribution. This indexing technique has been applied to hypothesis generation from a single view to reduce the number of candidates in 3D object recognition processes.

Keywords: Image indexing; Object recognition from a single view; Oriented matroid theory; Region connection calculus

E-mail addresses: ernesto.staffetti@ieee.org (E. Staffetti), antoni.grau@upc.es (A. Grau), francesc.serratosa@etse.urv.es (F. Serratosa), asanfeliu@iri.upc.es (A. Sanfeliu).

1. Introduction

Given one or more images containing different views of the same object or of the same set of objects, a fundamental problem of computer vision is that of recognizing the objects represented in the images using models of objects known a priori.

The components of an object recognition system are represented in Fig. 1. It consists in general of a database of models of views of known objects, in which they are represented by those attributes that are relevant to characterize each of them in relation to the others. The information stored in the database of models depends on the method employed for recognition and vice versa. Another component of an object recognition system is a feature extractor which computes the relevant features of the images of the objects to recognize. These features are used by the hypothesis generator to assign likelihoods to the objects present in the image. Hypothesis generation reduces the search space for the recognition process. The hypothesis verifier then uses object models to refine the likelihood and finally the systems selects the objects with the highest likelihood as the objects present in the views. However, the relative importance assigned to these two components in different object recognition systems varies.

Object representation for recognition can be based either on attributes of the 3D geometry of the objects or on attributes of their 2D projections. These two approaches are referred to as object-centered and view-centered representations, respectively. View-centered representations characterize 3D objects by means of a set of 2D views, resulting in a significant reduction in the dimensionality of the problem by comparing 2D images rather than comparing 3D objects. In this case 3D objects are represented as image features and relationships among them. Each 3D object has infinitely many different views that correspond to the infinitely many possible points of view. However, the complete set of views of a 3D object is redundant and for efficiency reasons the set of views used to characterize an objects must be reduced to a minimal set. Therefore, the concept of aspect or characteristic view [21] has been developed. Indeed, the view space can be partitioned into a finite number of regions called characteristic view domains so that views from different characteristic view domains are topologically distinct whereas views belonging to the same domain are not. The different characteristic views of an object can be related using a data structure called aspect graphs [11]. An aspect graph enumerates all the possible appearances of an object and their adjacency relationships. In this context, the change in appearance at the boundary between different aspects is called a visual event. It is easy to see that, using characteristic views, both the number of stored view models and the number of comparisons that must be performed at run time for object recognition is minimized. Using a set of view the problem of object recognition can be reduced to the problem of 2D image recognition. We suppose to use a single view for 3D object recognition.

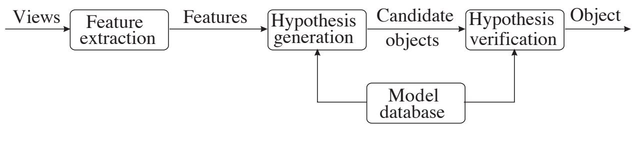

Fig. 1. Components of an object recognition system.

In object recognition systems based on a single view the objects can be characterized by global, local or relational features of their views. Global features are usually characteristics of interior points or of the boundary of regions of the views whereas local features make reference to small portions of the views. A better characterization of a view of an object can be done using relational features in which the relative position of global or local features is taken into account. The approach to pattern recognition based on relational features is also called structural pattern recognition and graphs are the most used relational data structures.

Classification is the standard object recognition method from single views but in some cases a classifier cannot be implemented, for instance because the a priori knowledge about feature probabilities and class probabilities is not available. In such cases direct matching of the models to the unknown object can be applied and the best matching model is selected. If the views are modeled using relational features, the matching will be symbolic, and, if graphs are used as relational data structures, the object recognition problem is then transformed into a graph matching problem [1,10,13,16,17].

Given a view of one or more 3D objects, the problem of object recognition using a single view becomes the problem of finding a subset of the set of regions of the view with a relational structure similar to that of a set of views stored in the database. Given the graph that represents a view of an object, the symbolic matching technique described above requires a sequential comparison between this graph and all the graphs that characterize all the views of the database. It is easy to see that this strategy is prohibitive with a large number of view models. To alleviate this problem, the concept of aspect graph introduced above can be used to reduce the number of graphs models in the database by representing each characteristic view using a single graph. Nevertheless, in the presence of a large number of objects, it is essential to reduce the search space using a hypothesis generator. Feature indexing can be used for this purpose. The idea behind features indexing is that when a certain feature is detected in the view of the object to recognize, it can be used to reduce the search space. Indeed, the search for the best matching model can be restricted to the subset of the set of view models of the database that contain that feature. However, to be of practical interest for object recognition, hypothesis generation should be a relatively fast although imprecise procedure in which a reduced number of possible candidates for matching are generated. In this way the verification can be carried out using a more complex, and therefore, slower procedure but over a reduced number of candidates [17].

In this paper an hypothesis generation strategy based on a indexing technique that combines region connection calculus and oriented matroid theory is presented [19]. More precisely, the type of connectivity between connected regions of the views is described by means of the formalism of region connection calculus [7], whereas the topological properties of the disconnected regions of the views are encoded into a data structure called set of cocircuits [4]. The set of cocircuits, that are one of the several combinatorial data structure referred to as oriented matroids, encode incidence relations and relative position of the regions of an image and give local and global topological information about their spatial distribution. Oriented matroids provide a straightforward mathematical interpretation of the concept of characteristic view and intrinsically contains information about their adjacency relationship, i.e, about the aspect graph of the object. Reasoning with region connection calculus is based on rules, while oriented matroids permit algebraic techniques to be used. These two relational data structures are used together to create an index of the database of views. This indexing method has been applied to hypothesis generation for 3D object recognition from a single view that can be regarded as a qualitative counterpart of the geometric hashing technique [12]. For other approach to shape representation and indexing based on combinatorial geometry see [6,2,3].

Region connection calculus and oriented matroid theory are introduced in Section 2 whereas Section 3 describes the proposed indexing method. In Section 4, some experimental results are reported and, finally, Section 5 contains the conclusions.

2. Qualitative spatial representation

Qualitative reasoning is based on comparative knowledge rather than on metric information. Many methods for shape representation and analysis are based on extracting points and edges that are used to define projectively invariant descriptors. In this paper, instead of points, regions of the images are taken into account. The motivation behind this choice is that the regions of an image can be more reliably extracted than vertices, edges or contours. Segmented views of 3D objects are considered, i.e., images partitioned into simply connected regions having perceptually homogeneous characteristics. Since our approach to image representation and indexing is combinatorial, it is important that the number of regions is low. Therefore, we implicitly suppose that we can control the maximum number of regions resulting from image segmentation, as in the method described in [8], in such a way that a too fine segmentation that would produce a prohibitive complexity of the resulting representation can be avoided. In the following sections two formalisms for qualitative representation and reasoning are described: the first one is based on region connection calculus and the second one is derived from oriented matroid theory.

2.1. Region connection calculus

For each simply connected region of an image we can qualitatively distinguish the interior, the boundary, and the exterior of the region, without taking into account its concrete shape or size. A set-theoretical analysis of the possible relations between objects based on the above partition is provided by Egenhofer and Franzosa [9]. The relation between objects that they examine is the intersection between their boundaries and interiors. This setting is based on the distinction of the values empty and non-empty for the intersection. If the exterior is regarded as a part of the object itself, there are three parts for each object that must be compared with three parts of another object. This gives rise to many set-theoretically distinguishable relations, but it is shown in [9] that some of them cannot occur. Actually only nine of them have a meaningful interpretation in physical space and are referred to as "disjoint", "meet", "equal", "inside", "covered by", "contains", "covers", and "overlap" (both with disjoint or intersecting boundaries). Consider for instance a region A with a hole B. Using this formalism, the relationship between A and B is described as "A contains B" or "B inside A".

Some variants of this theory were developed by Cohn and his coworkers in a series of papers (see for example [7]). In this work the distinction between interior and the boundary of an object is abandoned, and eight topological relations derived from the single binary

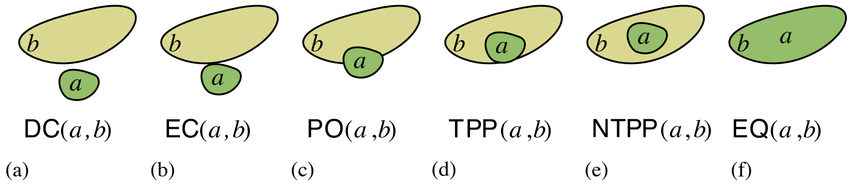

Fig. 2. Some of the 8 possible relative positions of two regions and the corresponding descriptions using the formalism of region connection calculus. The other two can be obtained from (d) and (e) interchanging a with b. In the situation (a) a is disconnected from b, in (b) a is externally connected to b, in the situation (c) a is partially overlapped to b, in (d) a is tangential proper part of b, in (e) a is non-tangential proper part of b and, finally, in the situation (f) a and b coincide.

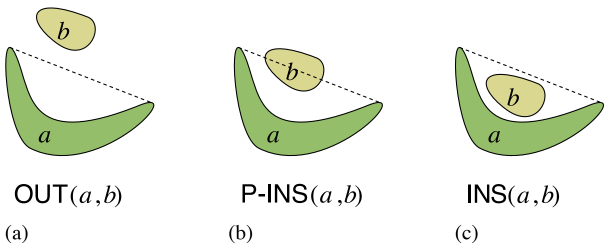

Fig. 3. Some of the possible positions of a convex region with respect to the convex hull of a non-convex one.

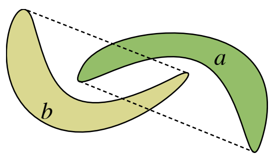

Fig. 4. With the formalism of region connection calculus the relation between these two disconnected non-convex regions, where a is partially inside the convex hull of b and vice versa, is denoted by $\mathsf{P\text{-}INS}^{\cdot}\mathsf{P\text{-}INSi}^{\cdot}\mathsf{DC}(a, b)$.

relation "connected to" are taken into account. Some of them are represented in Fig. 2. Two of these relations, namely those of Fig. 2(d) and (e), are not symmetrical and, following the notation used in [7], their inverses are denoted $\mathsf{TPPi}(a, b)$ and $\mathsf{NTTPi}(a, b)$, respectively. Furthermore in [7] the theory is extended to handle concave objects by distinguishing the regions inside and outside of the convex hull of the objects. A convex object can be inside, partially inside or outside the convex hull of a non-convex one (Fig. 3). If both regions are non-convex 23 relations between them can be defined. These relations permit qualitative description of rather complex relations, such as that represented in Fig. 4. Moreover, by means of this formalism called *region connection calculus* it is possible, for instance, to infer the type of connection between two regions knowing the type of connection they have with respect to a third one. Reasoning with region connection calculus is essentially based on rules. For other comprehensive descriptions of techniques for reasoning about qualitative spatial relationships see [18,22].

2.2. Oriented matroids

Oriented matroid theory [4,5,14] is a broad setting in which some combinatorial properties of geometrical configurations relevant for shape representation and indexing, such as, incidence and relative position, can be described and analyzed. It provides a common generalization of a large number of different mathematical objects usually treated at the level of usual coordinates. In this section oriented matroids will be introduced over sets of points called *arrangements of points* using two combinatorial data structures called chirotope and set of cocircuits, which represent the main tools to translate geometric problems into this formalism. In the abstraction process from the concrete arrangement of points to the oriented matroid, metric information is lost but the structural properties of the arrangement of points are represented at a purely combinatorial level. Then a technique to compute the oriented matroid representation for the set of regions that compose an image, called *arrangement of regions* is presented.

2.2.1. Oriented matroids of arrangements of points

Given a *arrangement of points* in $\mathbb{R}^{d-1}$ whose coordinates are the columns of the matrix $P=(p_1, p_2, \ldots, p_n)$ of $(\mathbb{R}^{d-1})^n$, the associated *arrangement of vectors* is a finite sequence of linearly independent vectors $\{x_1, x_2, \ldots, x_n\}$ in $\mathbb{R}^d$ represented as columns of the matrix $X = (x_1, x_2, \ldots, x_n)$ of $(\mathbb{R}^d)^n$ where each point p_i is represented in homogeneous coordinates as $x_i = \binom{p_i}{1}$.

To encode the combinatorial properties of the arrangement of points we can use a data structure called *chirotope* [14], which can be computed by means of the associated arrangement of vectors X. The chirotope of X is the map

$$\begin{aligned} \chi_X : \{1, 2, \ldots, n\}^d &\to \{+, 0, -\} \\ (\lambda_1, \lambda_2, \ldots, \lambda_d) &\mapsto \text{sign}\,([x_{\lambda_1}, x_{\lambda_2}, \ldots, x_{\lambda_d}]), \end{aligned} \tag{1}$$

in which the brackets stand for determinant, that assigns to each d-tuple of vectors of the finite arrangement X a sign $+$ or $-$ depending on whether it forms a basis of $\mathbb{R}^d$ having positive or negative orientation, respectively. Function (1) assigns the value 0 to those d-tuples that do not constitute a basis of $\mathbb{R}^d$. The chirotope describes the incidence structure between the points of X and the hyperplanes spanned by the same points and, at the same time, encodes the relative location of the points of the arrangement with respect to these hyperplanes themselves. This representation says, for example, which points of X lie on the positive side of a given hyperplane, which on the negative side, and which on the hyperplane itself.

Consider the arrangement of points P represented in Fig. 5 whose associated arrangement of vectors X is given in Table 1.

The chirotope χ_X of this arrangement of vectors is given by the orientations listed in Table 2. The element $\chi(1, 2, 3) = +$ indicates that in the triangle formed by p_1, p_2, and p_3 these points are counterclockwise ordered (Fig 6).

These orientations can be rearranged in an equivalent data structure called *set of cocircuits* of X shown in Table 3. The set of cocircuits of X is the set of all partitions generated by the lines passing through two points of the arrangement. For example, $(0, 0, +, +, +, +)$

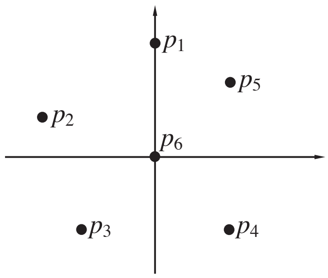

Fig. 5. A planar point configuration.

Table 1
Vectors that corresponds to the points of the planar arrangement represented in Fig. 5

$x_1 = (0, 3, 1)^T$	$x_2 = (-3, 1, 1)^T$	$x_3 = (-2, -2, 1)^T$
$x_4 = (2, -2, 1)^T$	$x_5 = (2, 2, 1)^T$	$x_6 = (0, 0, 1)^T$

Table 2
Chirotope of the planar arrangement of points represented in Fig. 5

$\chi(1, 2, 3) = +$	$\chi(2, 3, 4) = +$	$\chi(3, 4, 5) = +$	$\chi(4, 5, 6) = +$
$\chi(1, 2, 4) = +$	$\chi(2, 3, 5) = +$	$\chi(3, 4, 6) = +$	
$\chi(1, 2, 5) = +$	$\chi(2, 3, 6) = +$	$\chi(3, 5, 6) = 0$	
$\chi(1, 2, 6) = +$	$\chi(2, 4, 5) = +$		
$\chi(1, 3, 4) = +$	$\chi(2, 4, 6) = +$		
$\chi(1, 3, 5) = +$	$\chi(2, 5, 6) = -$		
$\chi(1, 3, 6) = +$			
$\chi(1, 4, 5) = +$			
$\chi(1, 4, 6) = -$			
$\chi(1, 5, 6) = -$			

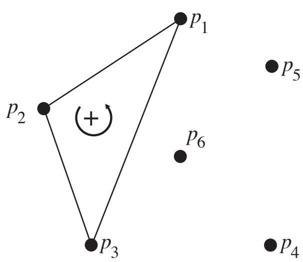

Fig. 6. In the triangle formed by p_1, p_2, and p_3 these points are counterclockwise ordered. This is recorded in the chirotope by $\chi(1, 2, 3) = +$.

Table 3
Set of cocircuits of the planar arrangement of points represented in Fig. 5

Ordered couples of points	Cocircuits					
	p_1	p_2	p_3	p_4	p_5	p_6
(p_1, p_2)	0	0	+	+	+	+
(p_1, p_3)	0	−	0	+	+	+
(p_1, p_4)	0	−	−	0	+	−
(p_1, p_5)	0	−	−	−	0	−
(p_1, p_6)	0	−	−	+	+	0
(p_2, p_3)	+	0	0	+	+	+
(p_2, p_4)	+	0	−	0	+	+
(p_2, p_5)	+	0	−	−	0	−
(p_2, p_6)	+	0	−	−	+	0
(p_3, p_4)	+	+	0	0	+	+
(p_3, p_5)	+	+	0	−	0	0
(p_3, p_6)	+	+	0	−	0	0
(p_4, p_5)	+	+	+	0	0	+
(p_4, p_6)	−	+	+	0	−	0
(p_5, p_6)	−	−	0	+	0	0

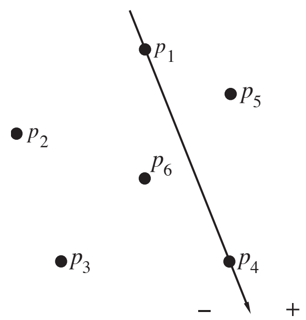

Fig. 7. Partition of the plane generated by the oriented ray from the point p_1 to p_4.

corresponding to the oriented couple (p_1, p_2) means that the points p_3, p_4, p_5, and p_6 lie on the half plane determined by the oriented ray from p_1 to p_2. Reversing all the signs of the set of cocircuits we obtain an equivalent description of the planar arrangement of points (Fig 7).

Besides chirotopes and cocircuits there are several data structures capable of encoding the topological properties of an arrangement of points. In [14] their definitions can be found and it is shown that all of them are equivalent and are referred to as *oriented matroids*.

2.2.2. Oriented matroid of arrangements of planar regions

Consider a segmented view of a 3D object. Computing the oriented matroid of the arrangement of regions of a view is not straightforward since the regions that form the

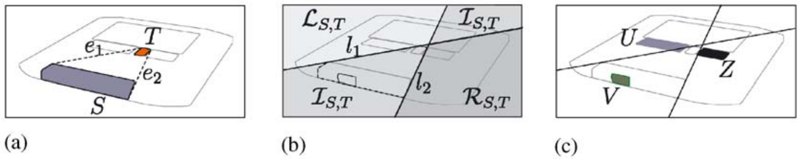

Fig. 8. Steps of encoding of the combinatorial properties of a view of an object into a chirotope.

image cannot be reduced to points, taking for instance their centroids, without losing essential topological information for object recognition. Therefore, the convex hull [15] of each region is employed to represent the region itself. Then, couples of the resulting convex hulls are considered and the oriented matroid is computed based on the location of the other convex regions of the image with respect to the two lines arising while merging them. Note that, since we decompose an image into a set of simply connected regions, holes are considered as separate regions. In this way we do not need to compute the convex hull of regions with holes that would eliminate the hole thus generating a change in the topology of the regions of the image.

Consider, for instance, the couple of convex hull of regions (S, T) of Fig. 8(a). It is easy to see that the convex hull of these two planar convex disconnected polygonal regions is a polygon whose set of vertices is included in the union of the set of vertices of S and T. On the contrary, the set of edges of the convex hull of S and T is not included in the union of their set of edges. Indeed, two new "bridging edges," e_1 and e_2, appear as illustrated in Fig. 8(a). Actually, efficient algorithms for merging convex hulls are based on finding these two edges [20].

Consider the two lines l_1 and l_2 that support e_1 and e_2. These two lines divide the image into three or four zones depending on the location of their intersection point with respect to the image. Let $\mathscr{R}_{S,T}$, $\mathscr{L}_{S,T}$ of Fig. 8(b) be, respectively, the rightmost and leftmost zones with respect to l_1 and l_2 and $\mathscr{I}_{S,T}$ the zone of the image comprised between them. Since, $\mathscr{R}_{S,T}$, $\mathscr{L}_{S,T}$ and $\mathscr{I}_{S,T}$ can be univocally determined from the ordered couple of regions S and T, the location of a region U with respect to the regions (S, T) of the image is encoded into a chirotope using the following rule

$$\chi(S, T, U) = \begin{cases} + & \text{if } U \in \mathscr{L}_{S,T}, \\ 0 & \text{if } U \in \mathscr{I}_{S,T}, \\ - & \text{if } U \in \mathscr{R}_{S,T}. \end{cases} \tag{2}$$

It has been implicitly assumed in (2) that U is completely contained into either $\mathscr{R}_{S,T}$ $\mathscr{L}_{S,T}$ or $\mathscr{I}_{S,T}$ but, in general, it belongs to more that one of them. In this case, since the ratio of areas is an affine invariant, introducing an approximation, we can choose the sign based on which region contains the largest portion of the area of U. If this region is split into 2 equally sized parts between $\mathscr{R}_{S,T}$ and $\mathscr{I}_{S,T}$, or between $\mathscr{L}_{S,T}$ and $\mathscr{I}_{S,T}$, the sign that corresponds to $\mathscr{R}_{S,T}$ and $\mathscr{L}_{S,T}$ will be assigned, respectively. On the contrary, if a region is split into three equally sized parts, among $\mathscr{R}_{S,T}$, $\mathscr{I}_{S,T}$ and $\mathscr{L}_{S,T}$, the 0 sign will be assigned. For instance, if regions U, V and Z are located as in Fig. 8(c) we have that $\chi(S, T, U) = +$, $\chi(S, T, V) = 0$ and $\chi(S, T, Z) = -$.

2.3. Invariance of the representation

Consider a 3D arrangement of points and one of its views. The chirotope of the 3D arrangement of points and that of its 2D perspective projection are related in the following way: if x_0 represents in homogeneous coordinates the center of the camera, p_0, we have that

$$\mathrm{sign}[\bar{x}_i, \bar{x}_j, \bar{x}_k] = \mathrm{sign}[x_i, x_j, x_k, x_0], \tag{3}$$

where x_i, x_j and x_k are the homogeneous coordinates of the 3D points p_i, p_j and p_k, and $\bar{x}_i$, $\bar{x}_j$ and $\bar{x}_k$ are those of the corresponding points in the view, $\bar{p}_i$, $\bar{p}_j$ and $\bar{p}_k$. Eq. (3) can be regarded as a projection equation for chirotopes.

It is easy to see that, whereas the *matrix* that represents in homogeneous coordinates the vertices of a projected set of points is *coordinate-dependent*, the oriented *matroid* of the projected set of points is a *coordinate-free* representation. Furthermore, it is a *topological invariant*, that is, an invariant under homeomorphisms. Roughly speaking, this means that the oriented matroid that represents the arrangement of points of a view of an object does not change when the points undergo a continuous transformation that does not change any orientation, that is, any sign of the chirotope. Due to this property, this representation is robust to discretization errors of the image as well as to small changes of the point of view that does not change any orientation of the chirotope. This does not occur when the viewing direction is nearly parallel to the plane spanned by the three observed points. It is easy to see that in this case a small displacement of the point of view may flip the plane and hence the sign of the determinant formed by these three points and the point of view.

The concept of chirotope of the set of views of a 3D object is related to the concepts of characteristic view and aspect graph of a 3D object. An equivalency relationship can be defined in the set of views of a 3D object so that all the views from the same characteristic view domain belongs to one equivalency class, the so-called *characteristic view class*. Each characteristic view class can be represented by any of its view called *aspect* or *characteristic view*. The equivalence relationship of characteristic view classes is defined as follows. Two views are equivalent if and only if they have isomorphic graph representation and can be related by a 3D geometric transformation that can be a Euclidean or a projective transformation. It is easy to see that the concept of characteristic view class has a natural mathematical interpretation by means of the concept of chirotope. Indeed, as explained in [4, Chapter 1] the oriented matroid of a view of a 3D object contains information about the underlying graph representation of the view and isomorphic graphs have the same oriented matroid representation. Furthermore, since projective transformations can be regarded as special homeomorphisms, we can assert that the representation of the projected set of points based on oriented matroids is *projectively invariant*. However, since affine and Euclidean transformations are special projective transformations, the oriented matroid of the projected set of points of a view of an object does not change under rotations, translations, and affine transformations of the planar arrangement of points themselves. Moreover, the oriented matroid representation of a set of views of a 3D object intrinsically contains information about their adjacency relationships, i.e., about the aspect graph of the object.

These considerations can be extended to the case in which oriented matroids represent an arrangement of planar regions. Since the ratio of areas is not invariant under projective

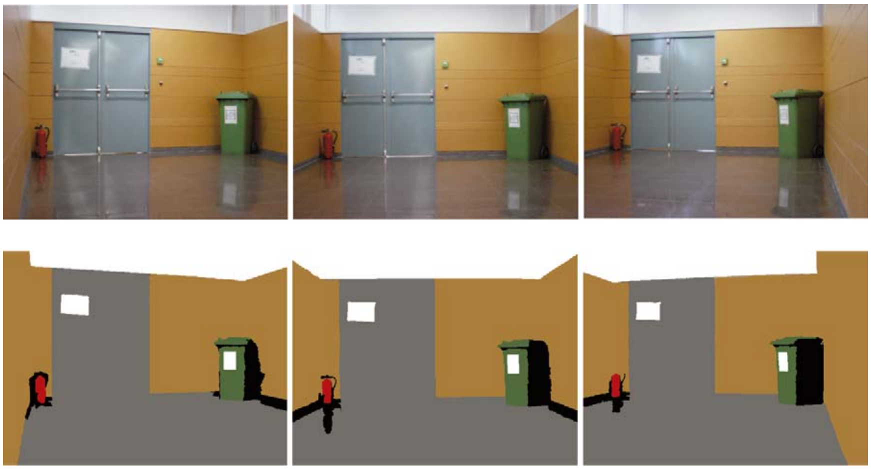

Fig. 9. Three different views of the same scene which have the same representation using oriented matroids.

transformations this representation will be invariant to affine and Euclidean transformations of the views. In Fig. 9 three different views of the same scene having the same oriented matroids are represented.

3. Indexing views of 3D objects

The process of indexing a database of views of a set of objects starts with some preliminary choices, namely the features used to characterize the regions of the segmented views of the set of 3D objects. Suppose that hue and area are used to characterize each region. Another parameter to choose is the number of levels in which the hue is quantized and the number of regions having the same hue that will be taken into account. These choices, of course, depend on the properties of the views of the database. Then, the views are segmented according to these choices and the convex hull of each region is computed. As a consequence, the resulting images are compositions of convex polygonal regions called *relevant regions* which can be disconnected or partially or completely overlapped. In Fig. 10 are represented two views of two objects in which a hue quantization with 6 levels W, R, Y, G, B and N has been applied and only the two regions having largest area with the same hue value are taken into account. Let (W, R, Y, G, B, N) be the ordered tuple of hue levels considered. For example, labels G_1 and G_2 in Fig. 10 denote, respectively, the first and the second regions of the views with the largest area having the same hue value G.

The type of connection between the existing regions is described using the formalism of region connection calculus. For each couple of disconnected regions the set of cocircuits is computed. This is done for each view of the database and this information is combined into a unique index table whose entries are relational features and whose records contain a list of the views in which each feature is present. The order of the rows is not relevant: it is a consequence of the ordering of the hue levels in the tuple (W, R, Y, G, B, N). The size of

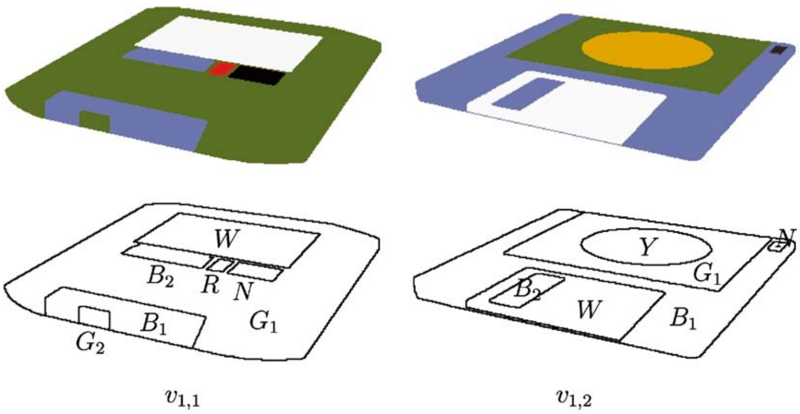

Fig. 10. Two views of two objects whose topological properties are indexed in Table 4.

Table 4
Index of the topological properties of the two views $v_{1,1}$ and $v_{1,2}$ of the two objects represented in Fig. 10

Ordered couples of regions	Type of connection	Cocircuits								Views
		W	R	Y	G_1	G_2	B_1	B_2	N	
(W, R)	DC	0	0	$*$	0	0	0	−	+	$v_{1,1}$
(W, Y)	DC	0	$*$	0	0	$*$	0	0	−	$v_{1,2}$
(W, G_1)	NTPP									$v_{1,1}$
(W, G_1)	DC	0	$*$	0	0	$*$	0	0	0	$v_{1,2}$
(W, G_2)	DC	0	0	$*$	0	0	+	0	0	$v_{1,1}$
(W, B_1)	DC	0	0	$*$	0	0	0	0	0	$v_{1,1}$
(W, B_1)	NTPP									$v_{1,2}$
(W, B_2)	DC	0	0	$*$	+	+	+	0	+	$v_{1,1}$
(W, B_2)	NTPPi									$v_{1,2}$
(W, N)	DC	0	0	$*$	−	−	−	−	0	$v_{1,1}$
(W, N)	DC	0	$*$	+	+	$*$	0	0	0	$v_{1,2}$
(R, Y)	Ø									
(R, G_1)	NTPP									$v_{1,1}$
...	...	...	...	...	...	...	...	...	...	...
(B_2, N)	DC	+	0	$*$	−	−	−	0	0	$v_{1,1}$
(B_2, N)	DC	−	$*$	+	+	$*$	+	0	0	$v_{1,2}$

the index table depends on both the number of hue levels and the number of regions having the same hue level that are taken into account.

In Table 4 the index of the topological properties of the two views $v_{1,1}$ and $v_{1,2}$ of the objects represented in Fig. 10 is reported. In the first column the relation between ordered couples of regions is described in terms of region connection calculus. The symbol "Ø" for a certain couple (S, T) indicates that no view contains two regions having features S and T. This is the case of the ordered couple of regions (R, Y). When S and T are disconnected, the corresponding cocircuit is present in the index. The symbol "$*$" in correspondence with a certain feature indicates that no region with that feature is present in the views listed in the record. For example, the cocircuit corresponding to the couple (W, R) contains a $*$ in the column Y because no region with the Y feature is present in $v_{1,1}$. If (S, T) is a couple of connected regions, the corresponding row of the index is empty because the cocircuit cannot be computed.

3.1. Hypothesis generation for object recognition

Consider a database of views of a set of known 3D objects and a view v_i of one of them that we want to recognize. As explained in the introduction, hypothesis generation for object recognition from a single view entails retrieving from the database the m views most similar to v_i to reduce the search space in object recognition from a single view. Roughly speaking an index indicates which views contain a certain relational feature. In this paper a strategy for hypothesis generation based on indexing of relational features will be presented. The relational features used to create the index are type of connection and relative location of the relevant regions of the views.

In the hypothesis generation process, the same type relational features are calculated for the view v_i and used to access the index table. Then, the m views that best match v_i are selected based on counting the number of correspondences they have with v_i in terms of type of connection and set cocircuits which together can be thought of as a feature matrix in which each relational feature is represented by a row and associated to a couple of relevant regions of v_i. To find the m views that best match v_i we follow an approach based on voting [12,6]. For each relational feature of v_i, i.e., for each couple of relevant regions of v_i, say (U, V), we access the database to retrieve the list of views in which the relational feature corresponding to the couple (U, V) is present. At each iteration each of the retrieved views receives one vote. The votes obtained by each view at each iteration are added together and the m best matches of v_i will be those views of the database that received the m highest scores. It is easy to see that this method for hypothesis generation, which can be regarded as a qualitative version of the geometric hashing technique [12], is also robust to partial occlusions of the objects. Indeed, if a region of an image is occluded, the set of cocircuits can still be computed and therefore, the number of correspondences with the views of the database can still be calculated. In this case, obviously, its selectivity decreases.

3.2. Computational complexity

Suppose that the segmented views are characterized by h hue values, and that the index is built taking into account the k largest regions for each value of hue. Then the maximum number of relevant regions of a view used for indexing is $n = kh$.

Since the index is based on a combinatorial characterization of the views in which all the possible ordered couples or relevant regions are considered, the number of entries of the index table, i.e., the number of different relational features that will be considered once the value n has been chosen, is given by the product of two factors. The first factor is the number of different relational features that can exist considering all the combinations of couples of regions. This corresponds to the number of combinations of the relevant

n regions taken 2 by 2, that is $n(n-1)/2$. Consider a couple of relevant regions of the image. The second factor represents the number of different relational features that can exist in correspondence to the same couple of relevant regions. It is given by the number of combinations of the other $(n-2)$ relevant regions taken 2 by 2, that is, $(n-2)(n-3)/2$.

Each relational feature has $(n+1)$ elements, namely 1 symbol that indicates the type of connection and n symbols that represent the cocircuit. Thus, the memory space needed to store the index table is $O(n^5)$. This value is independent of the number of views on the database. Additionally each entry contains a list of views whose size depends on the number of views of the database.

Feature extraction from each image returns $n(n-1)/2$ relational feature each composed by $(n+1)$ elements. In the hypothesis generation process the index table has to be accessed to understand which views contain a certain feature vector. This entails matching $(n+1)$-dimensional feature vectors that represent relational features. Each relational feature of v_i corresponding to a certain ordered couple of relevant regions has to be compared only with the entries of the index table corresponding to the same ordered couples of relevant regions. This amounts to $n(n-1)/2 \times (n-2)(n-3)/2$ comparisons. The resulting computational cost is $O(Cn^4 + D)$, where C is the cost of comparing two $(n+1)$-tuples of symbols and D is the cost of finding the view of the database having the maximum number of correspondences with the given view. The value Cn^4 is independent of the number of view of the database.

It is important to bear in mind that in the hypothesis generation method described in this paper n is usually a small number. If we use, for instance, 16 hue levels and consider the 2 largest regions having the same view $n = 32$. These are the values used in the experiments described in the next section.

4. Experimental results

The method described in this paper has been fully implemented and several experiments have been carried out to assess its effectiveness in hypothesis generation for object recognition and to compare it to other object recognition techniques based on relational features. In particular, it has been compared to two other structural methods based on graphs. The first experiment has been based on the adjacency graph of the regions of the views whereas in the second one a structure called *function-described graph* is used to characterize each view. In both cases a distance between graphs [16] has been computed to select the view most similar to a given view. Fig. 11 schematically shows the learning and recognition processes of the three methods (see [17] for more details).

Sixteen views of each object with angular separation of 22.5° have been used for the experiments. These images have been segmented using the method described in [8]. For illustrative purposes, Fig. 12 shows one view of each object and the corresponding segmented image, with the extracted adjacency graph in which the attributes of the nodes represent the average hue of the region. For each object, the reference set was composed by the views taken from the angles 0, 45, 90, 135, 180, 225, 270 and 315 and used to synthesize the structure of the database. The test set was composed by the eight views not used in the reference set, that is, the views taken at angles: 22.5°, 67.5°, 115.5°, 157.5°, 202.5°, 247.5°, 292.5°and 337.5°.

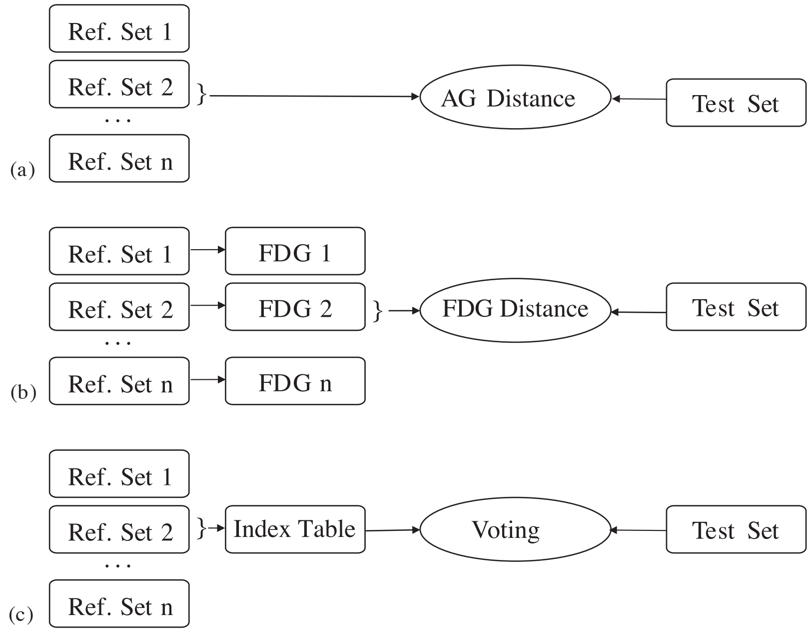

Fig. 11. The three structural object recognition methods from a single view considered in the experiments. (a) Method based on adjacency graphs. (b) Method based on function described graphs. (c) Method presented in this paper. Whereas in (a) and (b) a distance measure is used, in (c) the selection of the *m* views of the test set most similar to a given view is based on counting the number of correspondences based on the type of connection between regions and on the set of cocircuits.

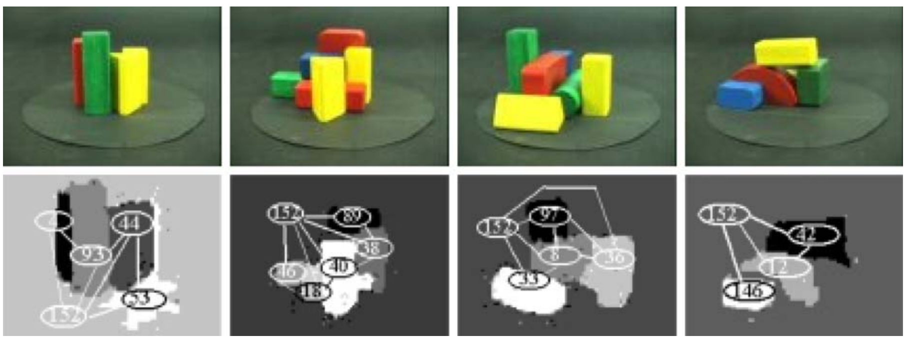

Fig. 12. Some of the views used for the experiments of object recognition, their segmented views and the adjacency graphs of the regions of the segmented views.

Table 5 shows the correctness of the adjacency graph (AG), function-described graphs (FDG) and the method presented in this paper based on indexing the topological properties

Table 5
Correctness of the three object recognition methods from a single view considered in the experiments

Method	Correctness
AG	59%
FDG	78%
RCC & OMT	87%

of the regions of the views of a 3D object using region connection calculus and oriented matroid theory (RCC & OMT)

5. Conclusions

In this paper a new method for indexing a database of views of 3D objects has been presented. It is based on the combination of two qualitative representations derived from region connection calculus and oriented matroid theory. It has been shown that this combination of qualitative representations characterizes the local and global topology of the regions of an image, is invariant under affine and Euclidean transformation of the views, intrinsically robust to discretization errors of the image and insensitive to small displacements of the point of view. This indexing method has been applied to hypothesis generation for 3D object recognition from a single view. The experimental results are encouraging and currently we are refining the method introducing a similarity measure between sets of cocircuits.

References

[1] S. Berretti, A. Del Bimbo, E. Vicario, Efficient matching and indexing of graph models in content-based retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 23 (10) (2001) 1089–1105.

[2] S. Berretti, A. Del Bimbo, E. Vicario, Spatial arrangement of colors in retrieval by visual similarity, Pattern Recognition 35 (8) (2002) 1661–1674.

[3] S. Berretti, A. Del Bimbo, E. Vicario, Weighted walkthrough between entities for retrieval by spatial arrangement, IEEE Trans. Multimedia 5 (1) (2003) 52–70.

[4] A. Björner, M. Las Vergnas, B. Sturmfels, N. White, G.M. Ziegler, Oriented Matroids, Encyclopedia of Mathematics and its Applications, vol. 43, Cambridge University Press, Cambridge, 1993.

[5] J. Bokowski, B. Sturmfels, Computational Synthetic Geometry, Lecture Notes in Mathematics, vol. 1355, Springer, Berlin, 1989.

[6] S. Carlsson, Combinatorial geometry for shape representation and indexing, in: J. Ponce, A. Zisserman (Eds.), Object Representation in Computer Vision II, Lecture Notes in Computer Science, vol. 1144, Springer, Berlin, 1996, pp. 53–78.

[7] A. Cohn, B. Bennett, J. Gooday, N.M. Gotts, Qualitative spatial representation and reasoning with the region connection calculus, GeoInformatica 1 (3) (1997) 275–316.

[8] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 603–619.

[9] M.J. Egenhofer, R.D. Franzosa, Point set topological relations, Internat. J. Geogr. Inform. Systems 5 (2) (1991) 161–174.

[10] S. Gunter, H. Bunke, Self-organizing map for clustering in the graph domain, Pattern Recognition Lett. 23 (4) (2002) 405–417.

[11] J. Koenderink, A. Van Doorn, The internal representation of solid shape with respect to vision, Biol. Cybern. 32 (1979) 211–216.

[12] Y. Lamdan, J.T. Schwartz, H.J. Wolfson, Affine invariant model-based object recognition, IEEE Trans. Robot. Automat. 6 (5) (1990) 578–589.

[13] B. Luo, E.R. Hancock, Structural graph matching using the EM algorithm and singular value decomposition, IEEE Trans. Pattern Anal. Mach. Intell. 23 (10) (2001) 1120–1136.

[14] J. Richter-Gebert, G.M. Ziegler, Oriented matroids, in: J.E. Goodman, J. O'Rourke (Eds.), Handbook of Discrete and Computational Geometry, CRC Press, Boca Raton, FL, 1997.

[15] J. O' Rourke, Computational Geometry in C, Cambridge University Press, Cambridge, 1999.

[16] A. Sanfeliu, K.-S. Fu, A distance measure between attributed relational graphs for pattern recognition, IEEE Trans. Systems, Man Cybern. 13 (1983) 353–362.

[17] F. Serratosa, R. Alquézar, A. Sanfeliu, Function-described graphs for modeling objects represented by attributed graphs, Pattern Recognition 36 (3) (2003) 781–798.

[18] A.P. Sistla, C. Yu, Reasoning about qualitative spatial relationships, J. Automated Reason. 25 (4) (2000) 291–328.

[19] E. Staffetti, A. Grau, F. Serratosa, A. Sanfeliu, Shape representation and indexing based on region connection calculus and oriented matroid theory, in: I. Nyström, G. Sanniti di Baja, S. Svensson (Eds.), Discrete Geometry for Computer Imagery, Lecture Notes in Computer Science, vol. 2886, Springer, Berlin, 2003, pp. 267–276.

[20] G.T. Toussaint, Solving geometric problems with the rotating calipers, in: Proceedings of IEEE MELECON'83, Athens, Greece, 1983.

[21] R. Wang, H. Freeman, Object recognition based on characteristic view classes, in: Proceedings of the International Conference on Pattern Recognition, 1990, pp. 8–12.

[22] S. Winter, Uncertain topological relations between imprecise regions, Internat. J. Geogr. Inform. Sci. 14 (5) (2000) 411–430.