

Introducing the Φ -Descriptor

A Most Versatile Relative Position Descriptor

P. Matsakis, M. Naeem and F. Rahbarnia

School of Computer Science, University of Guelph, Guelph, Canada

Keywords: Image Descriptors, Relative Position Descriptors, Spatial Relationships, *F*-Histograms, Affine Invariance.

Abstract: Spatial prepositions, like *above*, *inside*, *near*, denote spatial relationships. A relative position descriptor is a basis from which quantitative models of spatial relationships can be derived. It is an image descriptor, like colour, texture, and shape descriptors. Various relative position descriptors can be found in the literature. In this paper, we introduce a new relative position descriptor—the Φ -descriptor—that has about all the strengths of each and every one of its competitors, and none of the weaknesses. Our approach is based on the concept of the *F*-histogram and on an original categorization of pairs of consecutive boundary points on a line.

1 INTRODUCTION

The position of an object relative to another is an important feature people rely on to understand and communicate about space. In daily conversation, relative positions are described through the use of spatial prepositions, e.g., the apple *in* the bowl, the bowl *near* the vase, the vase *in front of* the window. These prepositions denote spatial relationships, which can be categorized into topological (e.g., *in*), distance (e.g., *near*) and directional (e.g., *in front of*) relationships. From a mathematical perspective, an object is a subset of the 2D or 3D space, and topological relationships include set relationships. For example, the condition $A \cap B = \emptyset$ defines the set (and hence topological) relationship *disjoint*, while $A \cap B \neq \emptyset$ and $\text{int}(A) \cap \text{int}(B) = \emptyset$ (disjoint interiors) define the topological (but non-set) relationship *touch*.

Models of spatial relationships have been investigated in many disciplines, including cognitive science, linguistics, geography, and artificial intelligence. In the qualitative approach (and contrary to the quantitative approach), the set of relationships is discrete (not continuous); a relationship either holds or does not hold (it cannot hold to some degree); spatial relationship information is decoupled from the individual features of the objects (like shape and size). For example, in the qualitative approach, one might consider the set {*east*, *northeast*, *north*, *northwest*, *west*, *southwest*, *south*, *southeast*} of

directional relationships; say that the playground is northeast of the building; argue that the exact shape of the playground is of no importance. In the quantitative approach, one may want to specify that the playground is 37° east of north of the building; allow partial truth when considering whether the playground is northeast of the building; argue that the shape of the playground might have an impact on the degree to which this relationship holds. The qualitative approach has been used extensively for spatial reasoning, and qualitative models are by far the most common models. However, many practical image processing and computer vision tasks call for quantitative models. Moreover, qualitative measures can easily be derived from quantitative measures, while the converse does not hold.

A relative position descriptor is an image descriptor, and it is a basis from which quantitative models of spatial relationships can be derived. As such, it provides a link between low-level spatial data features and high-level concepts. Moreover, it is a natural complement to colour, texture, and shape descriptors. Applications include human-robot interaction (Skubic et al., 2004), semantic metadata generation for image digital libraries (Wang, Makedon, Ford et al., 2004), suspected minefield risk estimation (Chan et al., 2005), melanocytic image analysis and recognition (Kwasnicka and Paradowski, 2005), geospatial information retrieval and indexing (Shyu et al., 2007), scene matching (Sjahputera and Keller, 2007), land cover classification (Vaduva et al., 2010), graphical symbol retrieval

(Santosh et al., 2010), shape matching (Wang et al., 2012), spatiotemporal reasoning (Salamat and Zahzah, 2012a), and map-to-image conflation (Buck et al., 2013).

In the light of these and other publications, here are what seem to be the most important properties that may be expected from a relative position descriptor. P1 — The descriptor can handle raster objects, whatever their topology (e.g., connected or disconnected, without or with holes), and whether they are disjoint or not. P2 — The descriptor can handle these objects efficiently (e.g., in linear time with respect to the number of pixels in the image). P3 — The descriptor can handle vector objects. P4 — The descriptor can handle distance relationships, i.e., meaningful distance relationship information can be extracted in no time from the descriptor. P5 — The descriptor can handle set relationships; at the very least, it can be used to determine whether two objects intersect, and whether one object includes the other. P6 — The descriptor can handle topological, non-set relationships; at the very least, it can be used to determine whether the boundaries of two objects intersect, and whether the interiors intersect. P7 — The descriptor can handle directional relationships; at the very least, it can be used to assess relationships like *to the right of*, *to the left of*, *above* and *below*. P8 — The descriptor can handle the relationship *surround*. P9 — Relative positions (as defined by the descriptor) can be somehow compared, and similar positions detected, regardless of which relationships hold. P10 — Given two objects A and B , the position of B relative to A can be derived from the position of A relative to B . P11 — Given an affine transformation t and two objects A and B , the position of $t(A)$ relative to $t(B)$ can be derived from t and the position of A relative to B . P12 — Given an affine transformation t and two objects A and B , the transformation t can be derived from the position of A relative to B and the position of $t(A)$ relative to $t(B)$. P13 — Consider four objects A, B, A' and B' ; whether there exists an affine transformation t such that $A'=t(A)$ and $B'=t(B)$ can be derived from the position of A relative to B and the position of A' relative to B' .

Various relative position descriptors can be found in the literature. Most of them are **histogram-based** descriptors. Each one meets a few of the above properties. As far as we know, however, none of them meets P1 to P13, or even P4 to P8, or P12 (Naeem and Matsakis, 2015). For example, the **histogram of forces** (Matsakis and Wendling, 1999), which is probably the relative position descriptor backed up with the most theoretical and applied

results (Matsakis et al., 2010), does not satisfy P4-6, P8, P12-13; the R*-histogram (Wang et al., 2004) does not satisfy P3, P6, P8, P11-13; the spread histogram (Kwasnicka and Paradowski, 2005) does not satisfy P2-4, P6-7, P10-13; the radial line model (Santosh et al., 2010) does not satisfy P4-6, P8, P11-13; the Allen histograms (Malki et al., 2002) (Matsakis and Nikitenko, 2005) (Salamat and Zahzah, 2012b) do not satisfy P2, P4, P8, P12-13; the ratio histogram (Wang et al., 2012) does not satisfy P4-8, P12.

In this paper, we introduce a histogram-based relative position descriptor—the Φ -descriptor—that meets each and every one of the 13 properties. Necessary background information is provided in Section 2. A detailed definition of the Φ -descriptor is presented in Section 3. In Section 4, we briefly explain why each property holds. Conclusions and future work are discussed in Section 5.

2 BACKGROUND

The Φ -descriptor is based on the concept of the ***F*-histogram**, which is briefly reviewed in Section 2.1. The relative position descriptor it is the closest to may be the one defined by the Allen histograms. These histograms are reviewed in Section 2.2 with the intent to help the reader understand the rationale behind our approach (Section 3).

2.1 The ***F*-Histogram**

Notation and terminology are as follows. See also Fig. 1. The symbol S denotes the Euclidean space. The *origin* ω is an arbitrary point of S . A *direction* θ is a unit vector. $\theta(p)$ is the line in direction θ that passes through the point p , and $\theta^\perp(p)$ is the subspace orthogonal to θ that passes through p . Note that $\theta^\perp(p)$ is a line if S is of dimension 2 and is a plane if S is of dimension 3. Now, consider a nonempty bounded subset A of S . The intersection $A \cap \theta(p)$ is a *core* of A . If any core of A is a closed set with a finite number of connected components, i.e., if it is the union of a finite number of pairwise disjoint segments, then A is an *object*. Consider a real function F that takes inputs of the form (θ, S_1, S_2) , where θ is a direction and S_1 and S_2 are two subsets of S . The *F-histogram* associated with the pair (A, B) of objects is the function F^{AB} defined by:

$$F^{AB}(\theta) = \int_{p \in \theta^\perp(\omega)} F(\theta, A \cap \theta(p), B \cap \theta(p)) dp \quad (1)$$

Its intended purpose is to represent, in some way, the position of A (the *argument*) with respect to B (the *referent*).

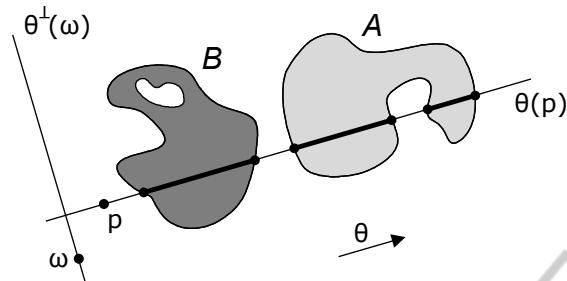


Figure 1: Notation: a direction θ , the origin ω , a point p , the line $\theta(p)$, the orthogonal line (or plane) $\theta^\perp(\omega)$, and two objects A and B .

The idea and assumption behind the concept of the F -histogram (Matsakis and Wendling, 1999) (Matsakis et al., 2010) are that acceptable representations of relative positions can be obtained by reducing the handling of multidimensional objects to the handling of 1D entities. The force histogram, the R^* -histogram, the ratio histogram and the Allen histograms mentioned in Section 1 are based on this concept.

2.2 The 13 Allen Histograms

The F -histogram naturally conveys directional information. It is therefore tempting to use a fuzzy approach and choose the function F such that the real number $F(\theta, A \cap \theta(p), B \cap \theta(p))$ measures the extent to which a given topological relationship holds between $A \cap \theta(p)$ and $B \cap \theta(p)$. If there are n possible topological relationships between two such cores, n histograms (one per relationship) should convey comprehensive quantitative information on the directional and topological relationships between the two objects

A and B . Unfortunately, there are infinitely many binary relations definable in the algebra generated by unions of segments on a directed line (Ladkin, 1986). In the simple case of a segment and the union of two disjoint segments, there are already over 40 topological relationships (Egenhofer, 2007). It seems wise to avoid a combinatorial explosion and rely on the very well known 13 Allen relations between two segments (Allen, 1983). See Fig. 2. For every segment (i.e., connected component) I of $A \cap \theta(p)$ and for every segment J of $B \cap \theta(p)$, the value $F(\theta, I, J)$ then measures the extent to which a given (fuzzified) Allen relation holds between I and J ; as for $F(\theta, A \cap \theta(p), B \cap \theta(p))$, it is some aggregate of all the $F(\theta, I, J)$ values.

There are several issues with this approach, which has been explored in various publications (Malki et al., 2002) (Matsakis and Nikitenko, 2005) (Salamat and Zahzah, 2012b). For example, it is hard to extract meaningful 2D topological relationship information from the 13 histograms of 1D fuzzy Allen relations, especially when the objects are not convex. This is apparent in (Matsakis et al., 2010) (Salamat and Zahzah, 2012c). Moreover, it is often impossible to extract crisp topological relationship information. To illustrate this, let F_P and F_M be the functions F attached to the Allen relations P (precedes) and M (meets). Assume $F_P^{AB}(\theta) \neq 0$ and $F_M^{AB}(\theta) \neq 0$. It may be because the statements " $A \cap \theta(p_1)$ precedes $B \cap \theta(p_1)$ " and " $A \cap \theta(p_2)$ meets $B \cap \theta(p_2)$ " are both totally true (and $A \cap B \neq \emptyset$). However, since P and M are conceptual neighbours and have been fuzzified, it may also be because the statements " $A \cap \theta(p)$ precedes $B \cap \theta(p)$ " and " $A \cap \theta(p)$ meets $B \cap \theta(p)$ " are both partially true (and $A \cap B = \emptyset$). As a result, one cannot answer with 'yes' or 'no' the question: "Are these objects disjoint?"

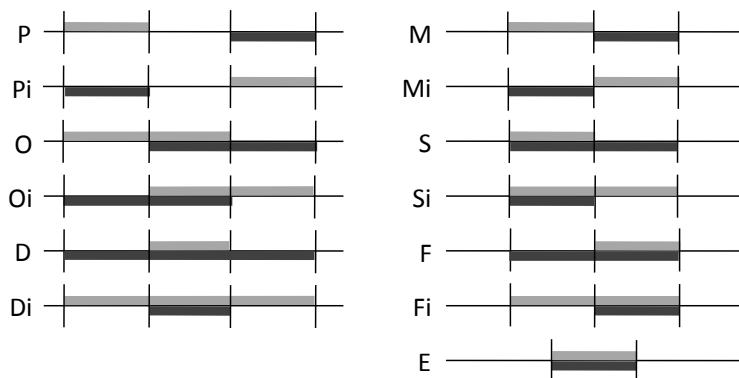


Figure 2: The 13 relations between two aligned segments (Allen, 1983). In each case, the argument is the light gray segment and the referent is the dark gray segment. P =Precedes, M =Meets, O =Overlaps, S =Starts, D =During, F =Finishes, E =Equals, Pi = P -inverse=Preceded by, Mi =Met by, Oi =Overlapped by, Si =Started by, Di =Contains, Fi =Finished by.

3 DEFINITION

The Φ -descriptor is built upon 13 F -histograms. Each histogram value corresponds to the area (in dimension 2) or volume (in dimension 3) of a region delimited by parallel lines and the boundaries of the objects in hand (Fig. 3). The outline of this section is as follows: a directed straight line intersects the boundaries of two objects in several points; these points fall into 12 categories, and the pairs of consecutive points fall into 36 categories (Section 3.1) divided into 10 groups (Section 3.2); a function is attached to each group and maps each pair to a real number; the 10 functions and 3 others are used to define the 13 F -histograms (Section 3.3) that are the basis of the Φ -descriptor (Section 3.4).

3.1 Boundary Points and Categories

Let A and B be two objects and let θ be a direction. Consider a line L in direction θ . Its intersection with A has a finite number of connected components, and each component is a line segment. Let p and q be the endpoints of one of these segments. If $p \neq q$ and $pq/|pq| = \theta$, where pq denotes the vector from p to q and $|pq|$ denotes its length, then p is an A -entry (on L , in direction θ) and q is an A -exit. Now, consider the set $\{p_1, p_2, \dots, p_n\}$ of all A -entries, A -exits, B -entries and B -exits on L and in direction θ . Assume that for any i we have $p_i p_{i+1}/|p_i p_{i+1}| = \theta$. The point p_{i+1} is then the *successor* of p_i . See Fig. 4. Consider two elements p and q of $\{p_1, p_2, \dots, p_n\}$ such that q is the successor of p . The point p falls into one of 12 categories, which can be named and represented as in Fig. 5. The same applies to q . As a result, the pair (p, q) falls into one of 36 categories. These categories, numbered from 1 to 36, are shown in Fig. 6. They may remind the reader of the Allen's relations. The two concepts are, in a sense, the reverse of each other: an Allen relation involves 2 segments, and up to 4 distinct points are the endpoints of these segments (Fig. 2); on the other hand, a point pair category involves 2 distinct points, and up to 4 segments have these points as endpoints.

3.2 Grouping the Point Pair Categories

In Fig. 6, the 36 point pair categories are divided into 9 groups (A - A , B - B , A - B , etc.). The division was convenient when trying to list all the categories. In this section, however, other groups are considered. Seven are labeled with a verb (third person singular form in the simple present tense): *trails*, *overlaps*, *covers*, *uncovers*, *follows*, *leads*, or *starts*. See Fig. 7.

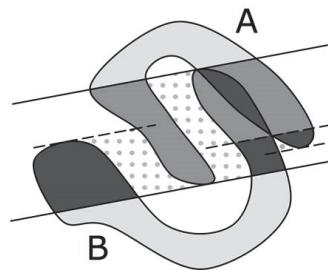


Figure 3: A glimpse into the area F -histograms. Each histogram value corresponds to the area of a region. Ten such regions are represented here (medium gray, dark gray, and dotted regions).

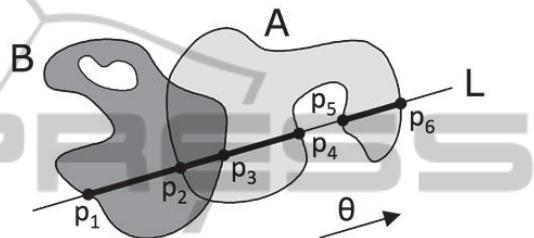


Figure 4: Entry and exit points. On L , in direction θ , the point p_1 is a B -entry not in A , the point p_2 is an A -entry in B , etc. (see Fig. 5); the pair (p_1, p_2) falls into the category 13, the pair (p_2, p_3) into the category 11, etc. (see Fig. 6).

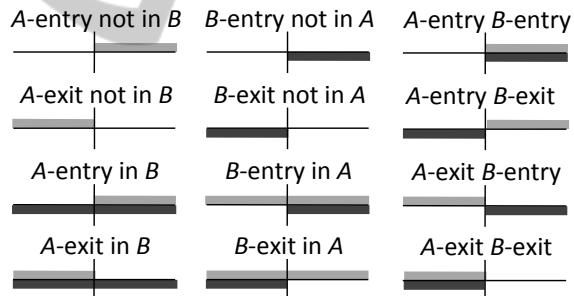


Figure 5: The 12 point categories. In each case, the argument is the light gray segment and the referent is the dark segment.

Each verb indicates a particular relationship between a segment of the argument A and a segment of the referent B . For example, in categories 10, 18, 32 and 36 a segment of A (left) is far behind (i.e., *trails*) a segment of B (right), while in categories 26, 30 and 34 a segment of A (left) is right behind (i.e., *follows*) a segment of B (center). Note that the terms *overlaps*, *follows* and *starts* are commonly used to denote the Allen relations O, F and S (Fig. 2). The reader should assume their meaning here is unrelated. Three groups of categories are labeled with a noun: *void*, *argument*, or *referent*. See Fig. 8. In these categories, there is no useful relationship between a segment of

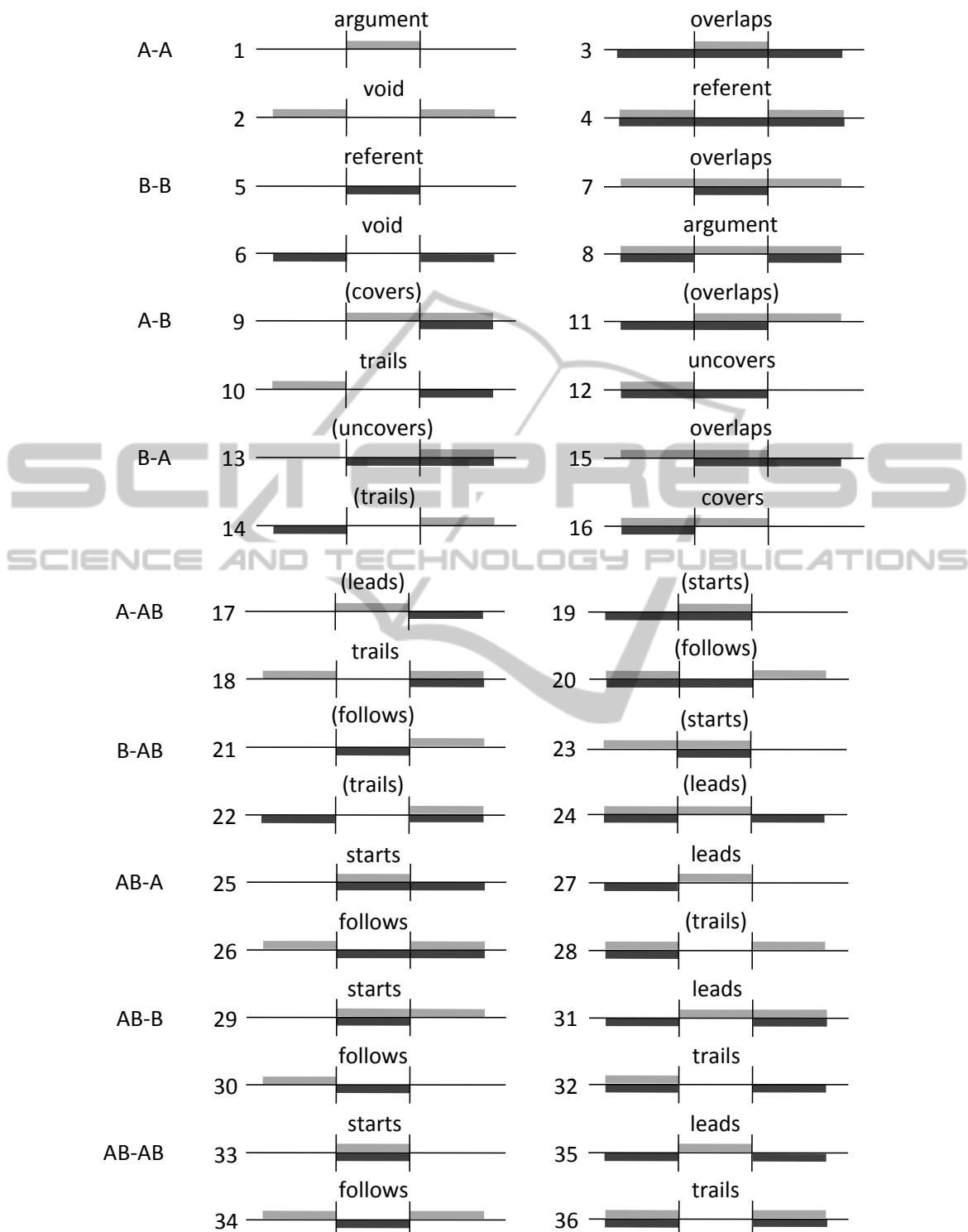


Figure 6: The 36 point pair categories. They are here divided into 9 groups of 4 categories. For example, AB-B means that the first point p is an A-entry or A-exit and a B-entry or B-exit, while its successor q is a B-entry or B-exit only. In each case, the argument is the light gray segment and the referent is the dark gray segment.

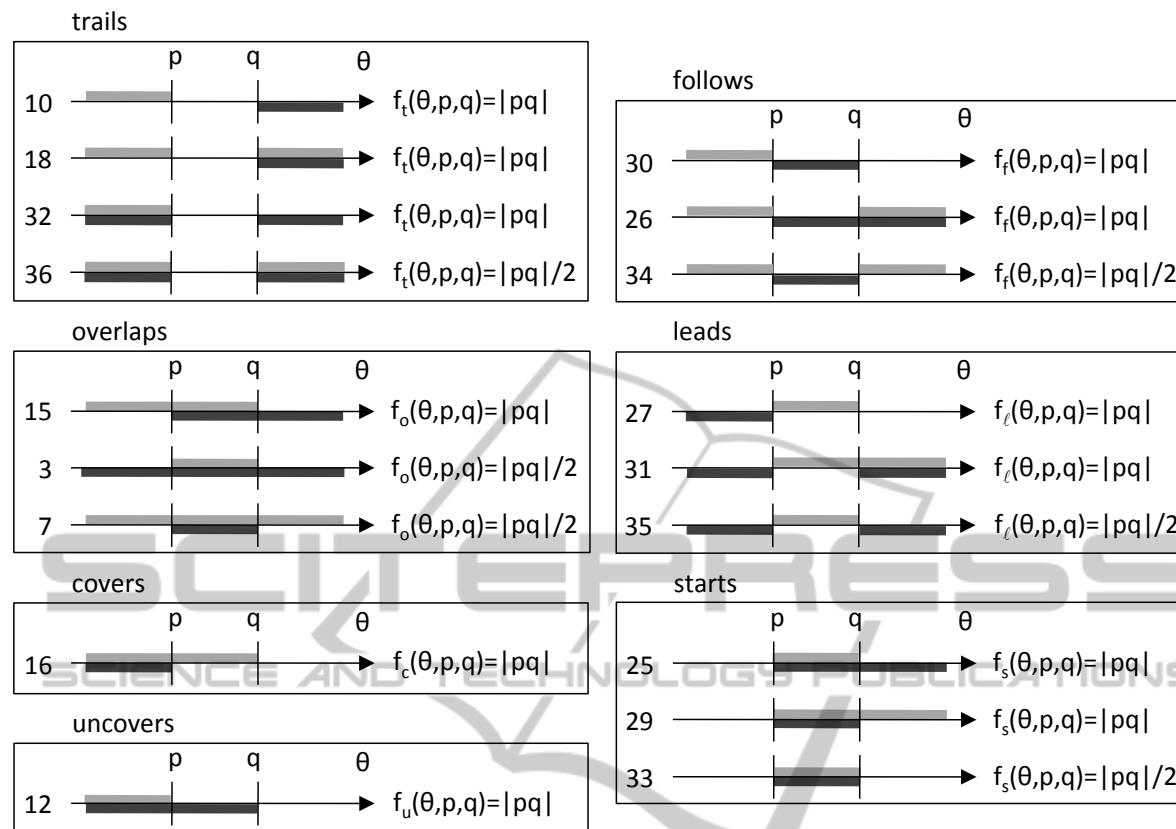


Figure 7: Nonzero values for the functions f_t , f_o , f_c , f_u , f_f , f and f_s .

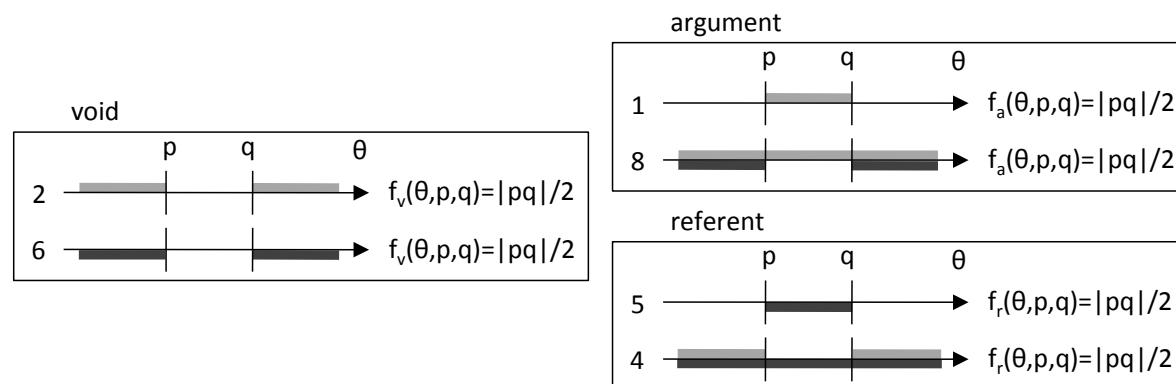


Figure 8: Nonzero values for the functions f_v , f_a and f_r .

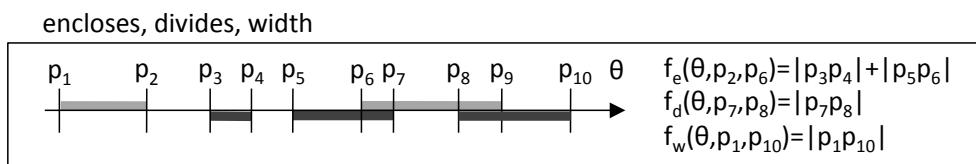


Figure 9: Nonzero values for the functions f_e , f_d and f_w . Examples.

A and a segment of *B*. Each noun refers to the object (if any) that occupies the space between *p* and *q*. Finally, note that the 10 groups shown in Figs. 7 and 8 include 24 categories only (instead of 36). Twelve are ignored to avoid redundancy and keep information directional; in Fig. 6, these categories are labeled with names in brackets. For example, category 9 is ignored because it is the *directional inverse* of category 16: if *p* is a *B*-exit and *q* an *A*-exit on line *L* and in direction θ (category 16), then *q* is an *A*-entry and *p* a *B*-entry on line *L* and in direction $-\theta$ (category 9).

3.3 Area and Volume *F*-Histograms

Consider two objects *A* and *B*. First, we define 10 real functions $f_t, f_o, f_c, f_u, f_f, f_\ell, f_s, f_v, f_a$ and f_r (the subscript *t* stands for *trails*, *o* for *overlaps*, *c* for *covers*, etc.). These functions take inputs of the form (θ, p, q) , where θ is a direction and *p* and *q* are two points. Let *L* be the line in direction θ that passes through *p*. Assume $\{p_1, p_2, \dots, p_n\}$ is the set of all *A*-entries, *A*-exits, *B*-entries and *B*-exits on *L* and in direction θ , as in Section 3.1. Each function *f* maps (θ, p, q) to 0 unless (p, q) is a pair (p_i, p_{i+1}) that falls into a category *f* is attached to. In that case, we have $f(\theta, p, q) = |pq|$ (and $f(-\theta, q, p) = 0$) if the category is not its own directional inverse; we have $f(\theta, p, q) = |pq|/2$ (and $f(-\theta, q, p) = |pq|/2$) otherwise. See Figs. 7 and 8. In a nutshell, the greater the distance between *p* and *q*, the more a segment of *A* *trails*, or *overlaps*, *covers*, etc., a segment of *B*.

We now define 3 more functions: f_e, f_d and f_w . The reason for this will be clarified later. The value $f_e(\theta, p, q)$ is 0 unless (p, q) is a pair (p_i, p_j) with $j > i$, the point *p_i* is an *A*-exit, *p_j* is an *A*-entry, and for any *k* in the integer interval $i+1..j-1$ the point *p_k* is neither an *A*-exit nor an *A*-entry; in that case, $f_e(\theta, p, q)$ is the total length of $B \cap [p_i, p_j]$. The value $f_d(\theta, p, q)$ is 0 unless (p, q) is a pair (p_i, p_j) with $j > i$, the point *p_i* is a *B*-exit, *p_j* is a *B*-entry, and for any *k* in $i+1..j-1$ the point *p_k* is neither a *B*-exit nor a *B*-entry; in that case, $f_d(\theta, p, q)$ is the total length of $A \cap [p_i, p_j]$. Finally, the value $f_w(\theta, p, q)$ is 0 unless (p, q) is the pair (p_1, p_n) ; in that case, $f_w(\theta, p, q) = |p_1 p_n|$. See Fig. 9.

The next step is to define 13 other functions $F_t, F_o, F_c, F_u, F_f, F_\ell, F_s, F_v, F_a, F_r, F_e, F_d$ and F_w . Each one maps $(\theta, A \cap L, B \cap L)$ to 0 if the set of all *A*-entries, *A*-exits, *B*-entries and *B*-exits on *L* and in direction θ is empty. Otherwise:

$$F_t(\theta, A \cap L, B \cap L) = \sum_{i=1}^{n-1} f_t(\theta, p_i, p_{i+1}) \quad (2)$$

$$F_o(\theta, A \cap L, B \cap L) = \sum_{i=1}^{n-1} f_o(\theta, p_i, p_{i+1}) \quad (3)$$

$$\dots \quad (4-10)$$

$$F_r(\theta, A \cap L, B \cap L) = \sum_{i=1}^{n-1} f_r(\theta, p_i, p_{i+1}) \quad (11)$$

$$F_e(\theta, A \cap L, B \cap L) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n f_e(\theta, p_i, p_j) \quad (12)$$

$$F_d(\theta, A \cap L, B \cap L) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n f_d(\theta, p_i, p_j) \quad (13)$$

$$F_w(\theta, A \cap L, B \cap L) = f_w(\theta, p_1, p_n) \quad (14)$$

These functions *F* and (1) allow us to define 13 *F*-histograms:

$$F_t^{AB}, F_o^{AB}, F_c^{AB}, F_u^{AB}, F_f^{AB}, F_\ell^{AB}, F_s^{AB}, F_v^{AB}, F_a^{AB}, F_r^{AB}, F_e^{AB}, F_d^{AB}, F_w^{AB}.$$

Each histogram value corresponds to an area (in dimension 2) or volume (in dimension 3). See Fig. 10.

Note that $F_w^{AB}(\theta)$ is the area or volume of the *region of interaction* in direction θ .

3.4 Length Histograms and the *Φ*-Descriptor

Let δ be the real function of a real variable defined by $\delta(0)=0$ and $\delta(x)=1$ if $x \neq 0$. The *length histogram* \bar{F}_w^{AB} is the real function defined by Equation (15).

$$\bar{F}_w^{AB}(\theta) = \frac{F_w^{AB}(\theta)}{\int_{p \in \theta^\perp(\omega)} \delta(F_w(\theta, A \cap \theta(p), B \cap \theta(p))) dp} \quad (15)$$

$\bar{F}_w^{AB}(\theta)$ is the average width of the region of interaction in direction θ . We may also say that it is the average nonzero f_w value in direction θ . Note that \bar{F}_w^{AB} is undefined at θ if $F_w^{AB}(\theta) = 0$. Likewise, we can define the length histograms $\bar{F}_t^{AB}, \bar{F}_o^{AB}, \dots$. For example, $\bar{F}_t^{AB}(\theta)$ is the average nonzero f_t value in direction θ .

We can now introduce the *Φ*-descriptor associated with the pair (A, B) of objects. It is a tuple Φ^{AB} of area (dimension 2) or volume (dimension 3) *F*-histograms and of length histograms. Its intended purpose is to represent the position of *A* relative to *B*. One possible definition is given by Equation (16), although more length histograms may be considered. $\text{measure}(A)$ denotes the area (dimension 2) or the volume (dimension 3) of *A*. Note that F_w^{AB} can actually be derived from $F_t^{AB}, F_o^{AB}, \dots, F_r^{AB}$.

$$\Phi^{AB} = (F_t^{AB}, F_o^{AB}, F_c^{AB}, F_u^{AB}, F_f^{AB}, F_\ell^{AB}, F_s^{AB}, F_v^{AB}, F_a^{AB}, F_r^{AB}, F_e^{AB}, F_d^{AB}, F_w^{AB}, F_w^{AB}, \text{measure}(A), \text{measure}(B)) \quad (16)$$

4 PROPERTIES

The Φ -descriptor satisfies each and every one of the properties P1 to P13 (Section 1). In this section, we briefly explain why. Complete proofs are given in separate papers, including (Matsakis and Naeem, submitted). Although the definition of the Φ -descriptor holds in any Euclidean space, it is assumed here that S is of dimension 2.

4.1 P1 to P3, P9 and P10 (Basics)

P1 and P2 — In the case of raster objects, the Φ -descriptor, which has obviously been designed with arbitrary objects in mind, can be computed in a very efficient way. For every direction θ , the image is partitioned into parallel raster lines. The pixels in a line are examined one by one and all the $F^{AB}(\theta)$ and $F^{AB}(-\theta)$ values are updated on the fly; basically, it is just a matter of counting the number of pixels between every two consecutive entry or exit points (i.e., boundary pixels). In the end, the Φ -descriptor is computed in $\mathcal{O}(KN)$ time, where N is the number of pixels in the image and K is the number of directions θ considered. Note that the higher K , the more complete the collected histogram data, but the longer the processing time. Practically, there does not seem to be any interest in considering more than a few hundred directions when computing F -histograms, and K is chosen between 4 and 360 (Matsakis et al. 2010).

P3 — In the case of vector objects, updating the $F^{AB}(\theta)$ and $F^{AB}(-\theta)$ values comes down to calculating the areas of polygons delimited by the boundaries of the objects and lines in direction θ . The Φ -descriptor is computed in $\mathcal{O}(K \eta^3)$ time, where η is the total number of object vertices. However, this worst-case performance falls to $\mathcal{O}(K \eta^2)$ when the objects intersect in η points or less, which is typical in practice.

P9 — A simple way to compare two relative positions Φ^{AB} and $\Phi^{A'B'}$ is to compare their corresponding elements, i.e., F_t^{AB} with $F_t^{A'B'}$, F_o^{AB} with $F_o^{A'B'}$, etc. For example, the similarity between two histograms h_1 and h_2 can be calculated using a

measure introduced by Pappis and Karacapilidis (1993):

$$\frac{\sum_{\theta} \min\{h_1(\theta), h_2(\theta)\}}{\sum_{\theta} \max\{h_1(\theta), h_2(\theta)\}} \quad (17)$$

Likewise, the similarity between two object areas a_1 and a_2 can be evaluated as:

$$\frac{\min\{a_1, a_2\}}{\max\{a_1, a_2\}} \quad (18)$$

The similarity, sim , between Φ^{AB} and $\Phi^{A'B'}$ can then be defined as the minimum similarity between corresponding histograms and object areas.

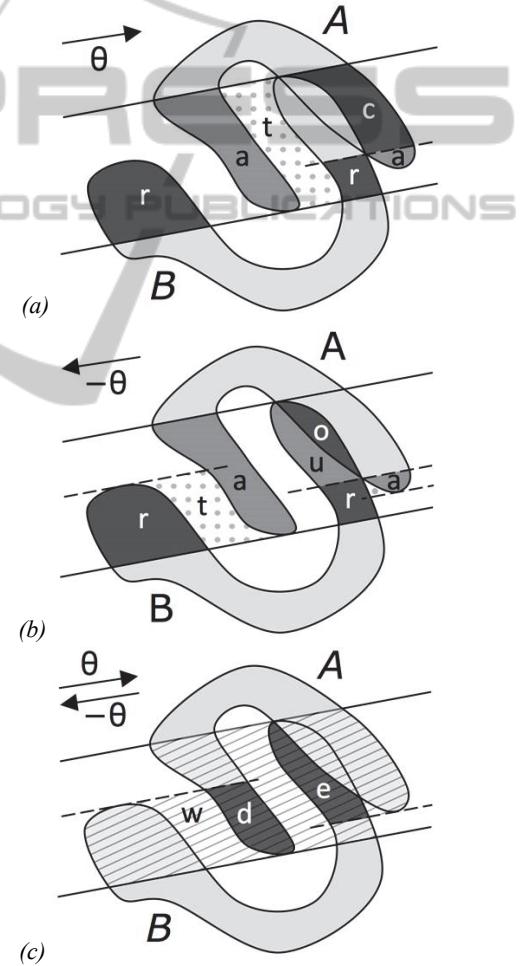


Figure 10: Area F -histogram values. For example, $F_r^{AB}(\theta)$ is half the total area of the two regions in (a) labeled r , and $F_r^{AB}(-\theta) = F_r^{AB}(\theta)$; the value $F_u^{AB}(\theta)$ is 0, but $F_u^{AB}(-\theta)$ is the area of the region in (b) labeled u ; the value $F_w^{AB}(\theta)$ is the area of the region in (c) filled with diagonal lines, and $F_w^{AB}(-\theta) = F_w^{AB}(\theta)$.

P10 — Each F^{BA} or $\overline{F^{BA}}$ histogram can be derived from an F^{AB} or $\overline{F^{AB}}$ histogram. For example, for any direction θ , we have $F_t^{BA}(\theta) = F_t^{AB}(-\theta)$ and $\overline{F_c^{BA}}(\theta) = \overline{F_u^{AB}}(\theta)$. As a result, the position of B relative to A can be derived from the position of A relative to B .

4.2 P4 to P8 (Spatial Relationships)

A great amount of spatial (including topological, directional and distance) relationship information can be extracted from the Φ -descriptor. For example, consider Fig. 11. Figure 11a is a world view of a robot in an environment with corridors and doorways. The

robot is a mobile Nomad 200 with 16 sonar sensors evenly distributed along its circumference. The sensor readings were used to build an approximate polygonal representation of the surrounding obstacles. The experiment was done using the Nomadic simulator. Figure 11b shows an egocentric robot view of the scene. The position of A (the robot) relative to B (the perceived environment) is described using the Φ -descriptor. See Figs. 11cd. Since $F_o^{AB} = F_s^{AB} = 0$ (everywhere zero histograms), the interiors of A and B do not intersect (this is good news for the robot). Moreover, since $F_f^{AB} = F_\ell^{AB} = 0$, the objects A and B do not even touch; they are disjoint (the robot is not leaning against the wall). The average distance, in direction θ , between A and B is

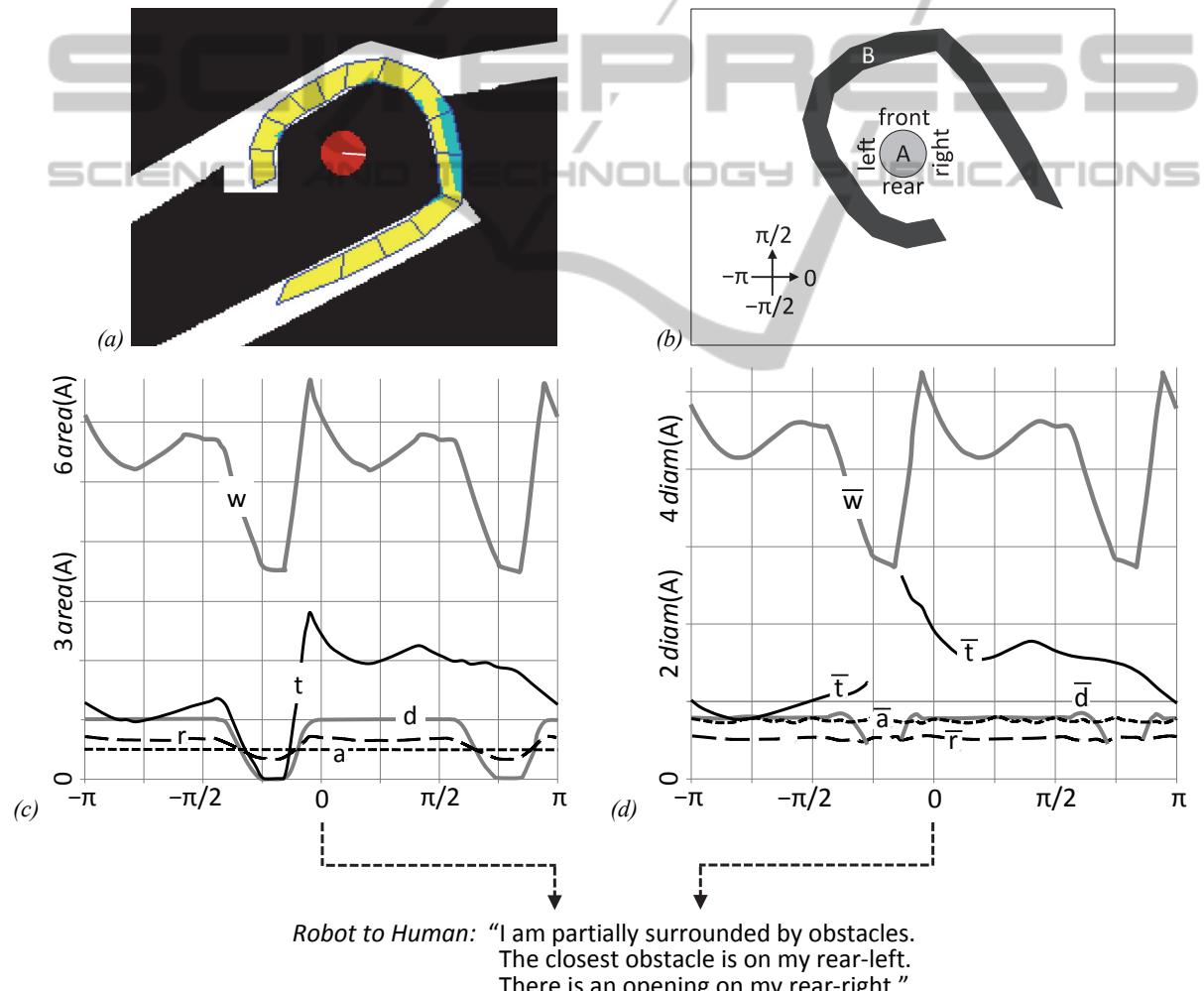


Figure 11: (a) A robot with sonar sensors, its environment, and its perception of the environment (Skubic et al., 2003). (b) An egocentric robot view of the scene. (c) The corresponding area histograms (w for F_w^{AB} , t for F_t^{AB} , etc.), and (d) the corresponding length histograms. On the vertical axes, $\text{area}(A)$ denotes the area of A while $\text{diam}(A)$ denotes its diameter.

$F_t^{AB}(\theta)$; it is minimum when $\theta = -3\pi/4$ (on the robot's rear-left). The object A would be totally surrounded by B if we had $F_d^{AB}(\theta) = \text{area}(A)$ for all θ , but this is not the case. However, A is partially surrounded by B , since $F_d^{AB}(\theta) = \text{area}(A)$ for all θ in the interval $[0, \pi/2]$ and $F_d^{AB}(\theta) \neq 0$ for all θ in $[-\pi/8, 5\pi/8]$. There is a $\pi/8$ -wide opening in direction $-\pi/4$ (on the robot's rear-right), since F_d^{AB} and F_t^{AB} are 0 on $[-\pi/4, -\pi/8]$.

Note that Fig. 11b may be seen as an illustration of one of the RCC23 spatial relationships (Cohn et al. 1997): the objects A and B do not intersect ($A \cap B = \emptyset$), the convex hull of A does not intersect B ($\text{conv}(A) \cap B = \emptyset$), and the convex hull of B includes A ($\text{conv}(B) \cap A = A$). The Φ -descriptor is able to identify every single one of the RCC23 relationships (and many, many more). In other words, it is able to provide crisp information and indicate whether yes or no a given relationship holds. With all the numerical histogram values available, it is also able, of course, to provide fuzzy information and indicate to what extent one may say that the relationship holds.

4.3 P11 to P13 (Affine Transformations)

Consider Fig. 12. Figures 12ab show two RGB pictures taken with a commercial digital camera, while Figs. 12cd show the pictures after segmentation. Segmentation was achieved by choosing the color channel with the best contrast (red channel), running an optimum thresholding algorithm (like Otsu's) on the corresponding gray-level histogram, and performing 7×7 median filtering on the thresholded image. Are the RGB pictures two pictures of the same scene? If so, which can (Figs. 12ac) is which (Figs. 12bd)? Color, texture, and shape descriptors would clearly not be very helpful in answering these questions. Later in this section, we focus on the two cans A_1 and A_2 , and we use this matching problem to illustrate the affine transformation properties of the Φ -descriptor.

Affine invariant descriptors play an important role in computer vision. Examples of affine invariant colour, texture, and shape descriptors abound in the literature. The Φ -descriptor can be normalized to obtain affine invariance and has many interesting related properties. Let aff be an invertible affine transformation. Areas under an affine transformation are scaled by the absolute value of the determinant of the matrix that represents the linear part of the affine transformation. As a result, $\Phi^{\text{aff}(A)\text{aff}(B)}$ can be easily derived from aff and Φ^{AB} . In other words, the

behaviour of the Φ -descriptor under affine transformations is known.

We have developed a normalization procedure $\Phi^{AB} \mapsto \overline{\Phi^{AB}}$ with the two following properties. Except for particular object pairs (i.e., object pairs that are not *well-behaved*), there exists a unique invertible linear transformation lin such that:

$$\overline{\Phi^{AB}} = \Phi^{\text{lin}(A)\text{lin}(B)} \quad (19)$$

Moreover, for any well-behaved object pair and for any invertible affine transformation aff we have:

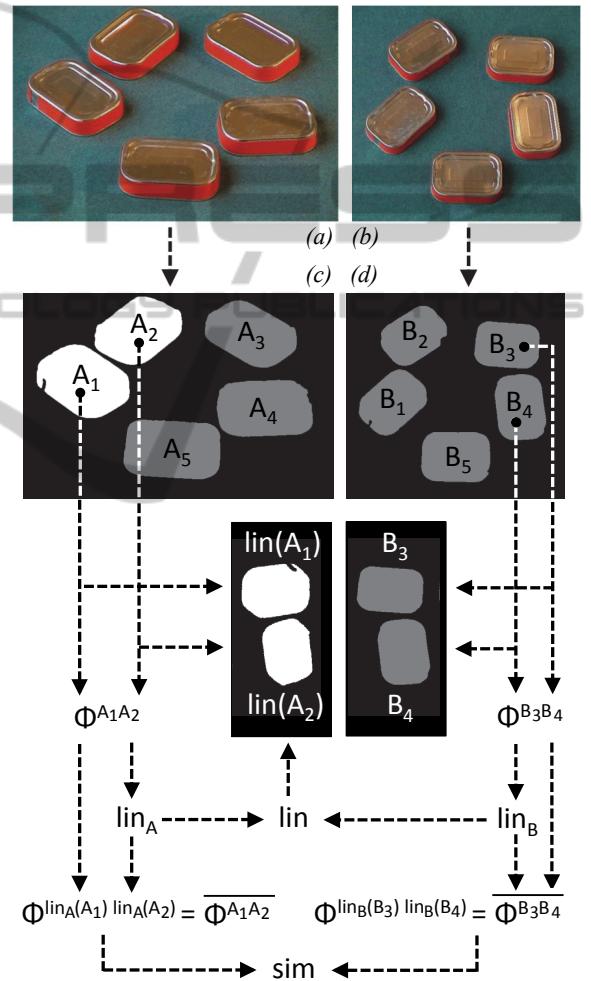


Figure 12: (a)(b) Two RGB pictures. (c)(d) The pictures after segmentation. According to the Φ -descriptor, the best match for (A_1, A_2) is (B_3, B_4) , and the linear transformation that best changes (A_1, A_2) into (B_3, B_4) is lin .

$$\Phi^{\text{aff}(A)\text{aff}(B)} = \overline{\Phi^{AB}} \quad (20)$$

In other words, the normalized Φ -descriptor is affine invariant. The idea behind Φ -descriptor

normalization is to derive from Φ^{AB} a vector basis intrinsic to the pair (A,B) . The uniqueness of lin comes from the uniqueness of the transformation used to change a vector basis into another. Note that the normalization procedure involves the length histogram $\overline{F_w^{AB}}$ and uses the fact that the behaviour of the Φ -descriptor under affine transformations is known. Consider Fig. 12 again: $\Phi^{A_1A_2}$ is derived from A_1 and A_2 ; then, lin_A is derived from $\Phi^{A_1A_2}$; finally, $\overline{\Phi^{A_1A_2}}$ is derived from lin_A and $\Phi^{A_1A_2}$.

Now, let (A,B) and (A',B') be two well-behaved object pairs. If there exists an invertible affine transformation aff such that

$$A' = aff(A) \text{ and } B' = aff(B) \quad (21)$$

then aff can be easily retrieved (up to a translation) from Φ^{AB} and $\Phi^{A'B'}$, using the normalization procedure. Moreover, if there exists an invertible transformation t (not necessarily affine) such that

$$A' = t(A) \text{ and } B' = t(B) \quad (22)$$

then the linear transformation that best approximates t (up to a translation) can be found, and the quality of the approximation can be assessed. Consider Fig. 12 once again: lin_A is derived from $\Phi^{A_1A_2}$ and lin_B is derived from $\Phi^{B_3B_4}$; then lin is derived from lin_A and lin_B . In this case, however, (A_1,A_2) and (B_3,B_4) are not affine-related (although they are matching pairs): in photography, the image formation process involves projective transformations instead of affine transformations; besides, A_1 , A_2 , B_3 and B_4 are 2D representations of 3D cans. The linear transformation lin is, therefore, only an approximation of the non-affine transformation that changes (A_1,A_2) into (B_3,B_4) . Compare $(lin(A_1),lin(A_2))$ with (B_3,B_4) . To assess the quality of the approximation, one must compare the two normalized descriptors $\overline{\Phi^{A_1A_2}}$ and $\overline{\Phi^{B_3B_4}}$, i.e., calculate their similarity sim (see Section 4.1, Property P9). Note that the similarity between $\overline{\Phi^{A_1A_2}}$ and $\overline{\Phi^{B_3B_4}}$, where i and j belong to $\{1,2,3,4,5\}$, was found to be maximum for $i=3$ and $j=4$.

5 CONCLUSIONS

What are the most important properties that may be expected from a relative position descriptor? In the light of articles on these descriptors and their applications, we have identified 13 properties. Taken individually, the current descriptors meet only a few of them. In this paper, we have introduced a relative

position descriptor—the Φ -descriptor—that meets each and every one of the 13 properties.

While most descriptors reduce the study of the relative position between two objects to the study of the relative positions between elementary components of the objects (e.g., pixels, points, segments), the Φ -descriptor uses an original approach based on the categorization of pairs of consecutive boundary points on a line. Moreover, the Φ -descriptor consists of raw data that are easy to acquire and interpret. There is no time-consuming pre-processing, like force calculation, or membership degree calculation. More spatial relationship information is preserved and can be extracted.

We are now developing a library of crisp and fuzzy models of spatial relationships based on the Φ -descriptor. The next step will be to focus on 3D objects; the definition of the Φ -descriptor holds in any Euclidean space, but only 2D objects have been considered so far. As for fuzzy objects, they can be handled using, e.g., the double sum scheme by Dubois and Jaulent (1987), or the simple sum scheme by Krishnapuram et al. (1993). However, such generic schemes are computationally expensive. Since the elementary values of the Φ -descriptor are areas, there should be a much simpler and more efficient way to process fuzzy objects, based on the concept of the area of a fuzzy set. The idea will have to be validated.

REFERENCES

- J. F. Allen, 1983. "Maintaining Knowledge About Temporal Intervals," *Communications of the ACM*, 26(11): 832-43.
- A. Buck, J. M. Keller, M. Skubic, 2013. "A Memetic Algorithm for Matching Spatial Configurations with the Histograms of Forces," *IEEE Trans. on Evolutionary Computation*, 17(4):588-604.
- J. C.-W. Chan, H. Sahli, Y. Wang, 2005. "Semantic Risk Estimation of Suspected Minefields Based on Spatial Relationships Analysis of Minefield Indicators from Multi-Level Remote Sensing Imagery," *Detection and Remediation Technologies for Mines and Minelike Targets X, Proceedings of SPIE*, 5794(1):1071-9.
- A. G. Cohn, B. Bennett, J. Gooday, N. M. Gotts, 1997. "Qualitative Spatial Representation and Reasoning with the Region Connection Calculus," *GeoInformatica*, 1(3):275-316.
- D. Dubois, M.-C. Jaulent, 1987. "A General Approach to Parameter Evaluation in Fuzzy Digital Pictures," *Pattern Recognition Letters*, 6:251-59.
- M. J. Egenhofer, 2007. "Temporal Relations of Intervals with a Gap," *14th Int. Symposium on Temporal Representation and Reasoning, Proceedings*, 169-74.

- R. Krishnapuram, J. M. Keller, Y. Ma, 1993. "Quantitative Analysis of Properties and Spatial Relations of Fuzzy Image Regions," *IEEE Trans. on Fuzzy Systems*, 1(3): 222-33.
- H. Kwasnicka, M. Paradowski, 2005. "Spread Histogram — A Method for Calculating Spatial Relations Between Objects," *4th Int. Conf. on Computer Recognition Systems (CORES), Proceedings*, 30:249-56.
- P. Ladkin, 1986. *The Logic of Time Representation*, PhD Thesis, University of California at Berkeley.
- J. Malki, E.-H. Zahzah, L. Mascarilla, 2002. "Indexation et recherche d'image fondées sur les relations spatiales entre objets," *Traitement du Signal*, 18(4):235-51.
- P. Matsakis, M. Naeem, submitted. "Basic and Affinity-Related Properties of a Groundbreaking Relative Position Descriptor," *IEEE Trans. on Pattern Analysis and Machine Intelligence*.
- P. Matsakis, D. Nikitenko, 2005. "Combined Extraction of Directional and Topological Relationship Information from 2D Concave Objects," in M. Cobb, F. Petry, V. Robinson (Eds.), *Fuzzy Modeling with Spatial Information for Geographic Problems*, Springer-Verlag, 15-40.
- P. Matsakis, L. Wawrzyniak, J. Ni, 2010. "Relative Positions in Words: A System that Builds Descriptions Around Allen Relations," *Int. J. of Geographical Information Science*, 24(1):1-23.
- P. Matsakis, L. Wendling, 1999. "A New Way to Represent the Relative Position of Areal Objects," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(7):634-43.
- P. Matsakis, L. Wendling, J. Ni, 2010. "A General Approach to the Fuzzy Modeling of Spatial Relationships," in R. Jeansoulin, O. Papini, H. Prade, S. Schockaert (Eds.), *Methods for Handling Imperfect Spatial Information*, Springer-Verlag, 49-74.
- M. Naeem, P. Matsakis, 2015, "Relative Position Descriptors: A Review," *4th Int. Conf. on Pattern Recognition Applications and Methods, Proceedings*, in press.
- C. Pappis, N. Karacapilidis, 1993. "A Comparative Assessment of Measures of Similarity of Fuzzy Values," *Fuzzy Sets and Systems*, 56:171-4.
- N. Salamat, E.-H. Zahzah, 2012a. "Spatio-Temporal Reasoning by Combined Topological and Directional Relations Information," *Int. J. of Artificial Intelligence and Soft Computing*, 3(2):185-201.
- N. Salamat, E.-H. Zahzah, 2012b. "On the Improvement of Combined Fuzzy Topological and Directional Relations Information," *Pattern Recognition*, 45(4):1559-1568.
- N. Salamat, E.-H. Zahzah, 2012c. "Two-Dimensional Fuzzy Spatial Relations: A New Way of Computing and Representation," *Advances in Fuzzy Systems*, 2012: 1-15.
- K.C. Santosh, L. Wendling, B. Lamiroy, 2010. "Unified Pairwise Spatial Relations: An Application to Graphical Symbol Retrieval," in J.-M. Ogier, W. Liu, J. Lladós (Eds.), *Graphics Recognition. Achievements, Challenges, and Evolution*, Springer-Verlag, 163-74.
- C.-R. Shyu, M. Klaric, G. J. Scott, A. S. Barb, C. H. Davis, K. Palaniappan, 2007. "GeoIRIS: Geospatial Information Retrieval and Indexing System—Content Mining, Semantics Modeling, and Complex Queries," *IEEE Trans. on Geoscience and Remote Sensing*, 45(4):839-52.
- O. Sjahputera, J. M. Keller, 2007. "Scene Matching Using F-Histogram-Based Features with Possibilistic C-Means Optimization," *Fuzzy Sets and Systems*, 158(3):253-69.
- M. Skubic, P. Matsakis, G. Chronis, J. M. Keller, 2003. "Generating Multi-Level Linguistic Spatial Descriptions from Range Sensor Readings Using the Histogram of Forces," *Autonomous Robots*, 14(1):51-69.
- M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, D. Brock, 2004. "Spatial Language for Human-Robot Dialogs," *IEEE Trans. on Systems, Man, and Cybernetics (Part C)*, 34(2):154-67.
- C. Vaduva, D. Faur, I. Gavat, 2010. "Data Mining and Spatial Reasoning for Satellite Image Characterization," *8th Int. Conf. on Communications (COMM), Proceedings*, 173-176.
- Y. Wang, F. Makedon, A. Chakrabarti, 2004. "R*-Histograms: Efficient Representation of Spatial Relations between Objects of Arbitrary Topology," *12th Annual ACM Int. Conf. on Multimedia (MM), Proceedings*, 356-59.
- Y. Wang, F. Makedon, J. Ford, L. Shen, D. Goldin, 2004. "Generating Fuzzy Semantic Metadata Describing Spatial Relations from Images Using the R-Histogram," *4th ACM/IEEE Joint Conf. on Digital Libraries, Proceedings*, 202-11.
- W. Wang, B. Xiong, H. Sun, H. Cai, Y. Jiang, G. Kuang, 2012. "An Affine Invariant Relative Attitude Relationship Descriptor for Shape Matching Based on Ratio Histograms," *EURASIP J. on Advances in Signal Processing*, 2012(1):1-10.