2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Affine Properties of the Relative Position PHI-Descriptor

Pascal Matsakis
School of Computer Science
University of Guelph
Ontario, Canada
pmatsaki@uoguelph.ca

***Abstract* — Colour, texture, shape, and relative position descriptors are fundamental visual descriptors. In particular, a relative position descriptor is a quantitative representation of the relative position of two spatial objects, and a basis from which models of spatial relationships (like *inside*, *above*, *around, near*) can be derived. The affine properties of visual descriptors have been the subject of much attention. Here, we focus on the PHI-descriptor, introduced in recent work. We show that it is a relative position descriptor with remarkable affine properties, and we illustrate these properties with two experiments.**

Keywords — visual descriptor; relative position; spatial relationship; affine transformation; invariance; computer vision

I. Introduction

Colour, texture, shape, and relative position descriptors are fundamental visual descriptors. While a texture descriptor carries information about the spatial arrangement of colours, a relative position descriptor carries information about the spatial arrangement of shapes. The affine properties of visual descriptors have been the subject of much attention. For example, affine invariance is often seen as a desirable property. Indeed, the image formation process involves projective transformations, which are often approximated by affinities under the assumption of weak perspective; affine invariance makes the descriptor less sensitive to the position of the camera with respect to the photographed scene.

There is an extensive literature on colour, texture, and shape descriptors. In particular, a very large number of shape descriptors have been proposed, and many are affine invariant [12]. Comparatively, the literature on relative position descriptors is much smaller [6]. This is not surprising, as it is only natural to talk about shapes before talking about the spatial arrangement of shapes. Nonetheless, relative position descriptors have found a variety of applications (e.g., graphical symbol retrieval [9], human-robot interaction [11], geospatial information retrieval and indexing [10], map-to-image conflation [1]). Knowing the spatial arrangement of an object's components often aids the recognition of the object—and of its components. Likewise, knowing the arrangement of objects in a scene aids the understanding of the scene—and the identification of the objects.

The focus here is on relative position descriptors — or *descriptors*, for short. We say that the descriptor Δ is *affine invariant* if for any affinity *aff* (e.g., rotation, scaling, shear) and for any spatial objects A and B we have $\Delta^{aff(A)aff(B)} = \Delta^{AB}$, where Δ^{AB} is a quantitative representation of the position of A relative to B. If Δ is not affine invariant, it may be possible, nonetheless, to find $\Delta^{aff(A)aff(B)}$ knowing *aff* and Δ^{AB}; we then say that the descriptor solves the *direct affine problem*. It may also be possible to find *aff* (up to a translation) knowing Δ^{AB} and $\Delta^{aff(A)aff(B)}$; we then say that the descriptor solves the *inverse affine problem*. According to a recent review [6], there is only one affine invariant descriptor, most descriptors do not solve the direct affine problem, and none solves the inverse problem.

In [5], we introduced a new descriptor—the Φ-descriptor —which has many advantages over its competitors: it can handle raster objects and vector objects, whatever their topology (e.g., connected or disconnected, without or with holes), and whether they are disjoint or not; it is much faster to compute than most descriptors; it can be used to develop qualitative and quantitative models of a large number of spatial relationships (like *inside*, *above*, *around, near*). For example, the well-known RCC8 topological relations are definable in terms of the descriptor, and they can be fuzzified based on the descriptor [4].

In this paper, we show that the Φ-descriptor solves both the direct and inverse affine problems. Moreover, it can be normalized so as to become affine invariant. The descriptor is reviewed briefly in Section II. Its affine properties are investigated in Section III. They are illustrated, tested and validated in Sections IV and V. Conclusions and future work are discussed in Section VI.

II. Φ-Descriptor

We give here an informal definition of the Φ-descriptor. It is very incomplete, but complete enough for the purpose of the paper. In the following, a *direction* is a unit vector; an *object* is a nonempty bounded regular closed set of the Euclidean plane; A and B denote two objects.

Consider a direction θ and two points p and q. The segment [p, q] is an *interaction segment* of A and B in direction θ if p and q belong to $A \cup B$ and to a line L in direction θ that intersects A and B but does not intersect $A \cup B$ outside [p, q]. The union of all the interaction segments of A and B in direction θ is the *region of interaction* of A and B in direction θ (Fig. 1*a*). The area of that region is denoted by $F_0^{AB}(\theta)$ and its average width —i.e., the average length of all the interaction segments—is denoted by $G_0^{AB}(\theta)$.

For every θ, the region of interaction in direction θ is partitioned into a finite number of subregions. Each subregion is delimited by the boundaries of the objects and lines in direction θ (Fig. 1*b*). The subregions fall into 10 categories. Two func-

tions from the set of directions to the set of real numbers are attached to each category: one records the areas (or half the areas) of the corresponding subregions and the other records their average widths. We therefore have 10 area functions, F_1^{AB}, F_2^{AB}, …, F_{10}^{AB}, and 10 length functions, G_1^{AB}, G_2^{AB}, …, G_{10}^{AB}, in addition to the functions F_0^{AB} and G_0^{AB}. See [5] for details, and Table I for an example.

The Φ-descriptor attached to the pair *(A,B)* can be defined as $\Phi^{AB} = (area(A), area(B), F_0^{AB}, F_1^{AB}, ..., F_{10}^{AB}, G_0^{AB}, G_1^{AB}, ..., G_{10}^{AB})$. This 24-tuple represents the position of *A* relative to *B*, and encapsulates a great amount of qualitative and quantitative spatial relationship information [4] [5]. The last piece of knowledge we need to share with the reader is the following: consider the function $F_{sum}^{AB} = \Sigma_{i=1}^{10} F_i^{AB}$; for any θ, we have $F_0^{AB}(\theta) = F_0^{AB}(-\theta) = F_{sum}^{AB}(\theta) + F_{sum}^{AB}(-\theta)$. Moreover, unless $F_0^{AB}(\theta)$ is 0, the values $F_{sum}^{AB}(\theta)$ and $F_{sum}^{AB}(-\theta)$ are rarely equal.

III. Affine Properties

An *affinity* (or *affine transformation*) is a one-to-one mapping from the Euclidean plane to itself that preserves lines and proportions on lines (e.g., shear). A *similitude* is an affinity that preserves all proportions (e.g., uniform scaling). An *isometry* is a similitude that preserves distances (e.g., rotation). The set of all affinities is a group under the operation of composition of functions; it is the *affine group*. The set of all similitudes and the set of all isometries are two subgroups of the affine group.

Consider an affinity *aff* and two objects *A* and *B*. The descriptor $\Phi^{aff(A)\,aff(B)}$ can be retrieved from Φ^{AB} and *aff* (Section III.A). Moreover, *aff* can be retrieved from Φ^{AB} and $\Phi^{aff(A)\,aff(B)}$ (Section III.C). This second result follows from the fact that the Φ-descriptor can be normalized so as to become affine invariant (Section III.B).

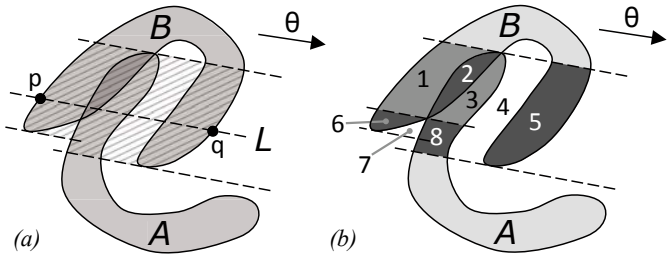

Fig. 1. *(a)* Two intersecting U-shaped objects and their region of interaction in direction θ (diagonal line pattern). *(b)* Partition of the region of interaction.

TABLE I. Areas Recorded by F_1^{AB}, F_2^{AB}, …, F_{10}^{AB}

Correspondence between the subregions in Fig. 1*b* and the areas recorded by the Φ-descriptor. For instance, the area of the subregion 1 is $F_4^{AB}(-\theta)$ and the total area of the subregions 5 and 6 is $2F_{10}^{AB}(-\theta)$, which is equal to $2F_{10}^{AB}(-\theta)$. Note that any function value at θ or –θ is 0 if not listed below (e.g., $F_2^{AB}(\theta) = F_3^{AB}(-\theta) = 0$).

1	$F_4^{AB}(-\theta)$	3	$F_3^{AB}(\theta)$	5∪6	$2F_{10}^{AB}(-\theta) = 2F_{10}^{AB}(\theta)$
2	$F_2^{AB}(-\theta)$	4	$F_1^{AB}(\theta)$	7	$F_1^{AB}(-\theta)$
				8	$2F_9^{AB}(-\theta) = 2F_9^{AB}(\theta)$

A. Retrieving $\Phi^{aff(A)\,aff(B)}$ from Φ^{AB} and aff

The affinity *aff* can be written as the composition of a translation with an invertible linear transformation aff_ω (an affinity such that $aff_\omega(\omega)=\omega$, where ω is the *origin*, i.e., an arbitrary point of the plane). It is a common convention to see linear transformations as 2×2 matrices and vectors as 2×1 column matrices. Consider the unit vector, i.e., the direction,

$$\theta' = \left(aff_\omega^{-1} \cdot \theta\right) / \left|aff_\omega^{-1} \cdot \theta\right| = \overline{aff_\omega^{-1} \cdot \theta} \quad (1)$$

where the dot denotes matrix multiplication, the vertical bars vector norm and the overbar vector normalization. For any integer *i* in 0..10, we have:

$$area(aff(A)) = |\det(aff_\omega)|\ area(A) \quad (2)$$

$$area(aff(B)) = |\det(aff_\omega)|\ area(B) \quad (3)$$

$$F_i^{aff(A)\,aff(B)}(\theta) = |\det(aff_\omega)|\ F_i^{AB}(\theta') \quad (4)$$

$$G_i^{aff(A)\,aff(B)}(\theta) = |aff_\omega \cdot \theta'|\ G_i^{AB}(\theta') \quad (5)$$

where $|\det(aff_\omega)|$ is the absolute value of aff_ω's determinant. These equalities are according to well-known results from linear algebra about how affinities transform directions, areas, lengths.

B. Normalizing the Φ-descriptor

Let *A′=aff(A)* and *B′=aff(B)*. The Φ-descriptor is not affine invariant, i.e., we usually have $\Phi^{A'B'} \neq \Phi^{AB}$. In this section, we show that it can be normalized such that $\overline{\Phi^{A'B'}} = \overline{\Phi^{AB}}$, where the overbar here indicates descriptor normalization.

Two directions, the *major direction* and the *minor direction*, are derived from the descriptor (Section III.B.1). They are used to define an invertible linear transformation: the *normalizing transformation* (Section III.B.2). Let θ_0 and θ_1 be the directions derived from Φ^{AB} and let θ_0' and θ_1' be those derived from $\Phi^{A'B'}$. We have:

$$\theta_0' = \overline{aff_\omega \cdot \theta_0} \text{ and } \theta_1' = \overline{aff_\omega \cdot \theta_1} \quad (6)$$

Put simply, the transformed major and minor directions of a descriptor are the major and minor directions of the transformed descriptor. Moreover, let *lin* be the normalizing transformation of Φ^{AB} and let *lin′* be that of $\Phi^{A'B'}$. We have:

$$\Phi^{lin'(A')\,lin'(B')} = \Phi^{lin(A)\,lin(B)} \quad (7)$$

The *normalized descriptor* $\overline{\Phi^{AB}}$ is defined as $\Phi^{lin(A)\,lin(B)}$ (Section III.B.3). Note that, in practice, descriptors should be compared through a similarity measure (Section III.B.4).

1) Major and minor directions: Let θ_{max} be a global maximum direction for F_0^{AB}:

$$\forall\theta,\ F_0^{AB}(\theta_{max}) \geq F_0^{AB}(\theta) \quad (8)$$

There is at least one other global maximum direction, $-\theta_{max}$, but in the general case we can expect that one and only one global maximum direction, θ_0, will maximize $F_{sum}^{AB}(\theta)$:

$$F_0^{AB}(\theta_0) = F_0^{AB}(\theta_{max}) \quad (9)$$

$$\forall\theta,\ [\ F_0^{AB}(\theta_0) = F_0^{AB}(\theta) \rightarrow F_{sum}^{AB}(\theta_0) \geq F_{sum}^{AB}(\theta)\ ] \quad (10)$$

Now, let θ_{min} be a global minimum direction for F_0^{AB}:

$$\forall\theta,\ F_0^{AB}(\theta_{min}) \leq F_0^{AB}(\theta) \quad (11)$$

Assume $F_0^{AB}(\theta_{\min}) \neq 0$, e.g., $F_0^{AB}(\theta_{\min}) \geq F_0^{AB}(\theta_{\max})/2$. There is at least one other global minimum direction: $-\theta_{\min}$. Again, we can expect that one and only one global minimum direction, θ_1, will maximize $F_{sum}^{AB}(\theta)$:

$$F_0^{AB}(\theta_1) = F_0^{AB}(\theta_{\min}) \tag{12}$$

$$\forall\theta, [\, F_0^{AB}(\theta_1) = F_0^{AB}(\theta) \rightarrow F_{sum}^{AB}(\theta_1) \geq F_{sum}^{AB}(\theta)\,] \tag{13}$$

There is no guarantee, of course, that $F_0^{AB}(\theta_{\min}) \neq 0$. In many cases (e.g., consider objects with nonintersecting minimum bounding rectangles), $F_0^{AB}(\theta_{\min}) = F_{sum}^{AB}(\theta_{\min}) = 0$, there is an infinite number of such global minimum directions, and we cannot single out any of them. A way to solve this issue is to replace (12) by (14):

$$F_0^{AB}(\theta_1) = \max\{\, F_0^{AB}(\theta_{\min}) \;,\; F_0^{AB}(\theta_{\max})/2 \,\} \tag{14}$$

From now on, we will assume the directions θ_0 and θ_1 are unique and distinct, as expected in the general case. The pair *(A, B)* is then *well-behaved*, θ_0 is its *major direction*, θ_1 is its *minor direction*, and (6) follows from (4).

2) Normalizing transformation: The normalizing transformation of Φ^{AB} is

$$lin = lin_6 \cdot lin_5 \cdot lin_4 \cdot lin_3 \cdot lin_2 \cdot lin_1 \tag{15}$$

where lin_1 to lin_6 are as shown below.
α, β and γ are in the real interval $(-\pi,\pi]$.

$$lin_1 = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}, \tag{16}$$

where α is the angle from θ_0 to the basis vector $[1\ \ 0]^t$;

$$lin_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ if } \beta>0 \text{ and } lin_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \text{ if } \beta<0, \tag{17}$$

where β is the angle from θ_0 to θ_1;

$$lin_3 = \begin{bmatrix} 1 & \cot\gamma - \cot|\beta| \\ 0 & 1 \end{bmatrix}, \tag{18}$$

$$\text{where } \gamma = \frac{F_0^{AB}(\theta_{\max}) - F_0^{AB}(\theta_1)}{F_0^{AB}(\theta_{\max}) - F_0^{AB}(\theta_{\min})}\frac{\pi}{2}; \tag{19}$$

$$lin_4 = \begin{bmatrix} 1 & (k-1)\cot\gamma \\ 0 & k \end{bmatrix}, \tag{20}$$

$$\text{where } k = \sin\gamma / \sin|\beta|; \tag{21}$$

$$lin_5 = \begin{bmatrix} 1 & (k_1-1)\cot\gamma \\ 0 & k_1 \end{bmatrix}, \tag{22}$$

$$\text{where } k_1 = \frac{F_0^{AB}(\theta_1)}{F_0^{AB}(\theta_0)}\frac{G_0^{AB}(\theta_0)}{G_0^{AB}(\theta_1)}; \tag{23}$$

$$lin_6 = \begin{bmatrix} k_0 & 0 \\ 0 & k_0 \end{bmatrix}, \tag{24}$$

$$\text{where } k_0 = \frac{1}{\sqrt{k_1 k F_0^{AB}(\theta_0)}}. \tag{25}$$

Roughly, *lin*'s purpose is to align θ_0 and θ_1 with the basis vectors $[1\ \ 0]^t$ and $[0\ \ 1]^t$. However, aligning θ_1 with $[0\ \ 1]^t$ only makes sense if $F_0^{AB}(\theta_1) = F_0^{AB}(\theta_{\min})$. For example, if $F_0^{AB}(\theta_{\min}) = 0$ and $F_0^{AB}(\theta_1) = F_0^{AB}(\theta_{\max})/2$, then a better option is to align θ_1 with $[\cos(\pi/4)\ \sin(\pi/4)]^t$. For a smooth transition between the two cases, θ_1 is aligned with the vector $[\cos\gamma\ \sin\gamma]^t$, using lin_1 to lin_4 (Fig. 2). The first transformation, lin_1, is a rotation; lin_2 is either the identity function or a reflection; lin_3 is a shear; lin_4 is a non-uniform scaling that preserves lengths in direction $[1\ \ 0]^t$ but multiplies them by k in direction $[\cos\gamma\ \sin\gamma]^t$. The vector $\theta_1^3 = lin_3 \cdot lin_2 \cdot lin_1 \cdot \theta_1$ is not a unit vector — its length is $\sin|\beta| / \sin\gamma$ — and must therefore be normalized; hence lin_4.

Equation (15) is equivalent to $lin \cdot \theta_0 = k_0\,[1\ \ 0]^t$ and $lin \cdot \theta_1 = k_0\, k_1\,[\cos\gamma\ \sin\gamma]^t$. The number k_1 controls relative scaling, i.e., it allows us to scale $lin \cdot \theta_1$ relative to $lin \cdot \theta_0$. For example, if lengths in direction $[\cos\gamma\ \sin\gamma]^t$ are multiplied by some value, then the length of $lin \cdot \theta_1$ must be divided by the same value to ensure that (7) holds—hence the presence of $G_0^{AB}(\theta_1)$ in (23). Because of (4) and (6), the ratio $F_0^{AB}(\theta_1)/F_0^{AB}(\theta_0)$ remains unchanged when the objects are affine transformed. Its presence in (23) is not required. It simply forces k_1 to be 1 when $F_0^{AB}(\theta_1)/F_0^{AB}(\theta_0) = G_0^{AB}(\theta_1)/G_0^{AB}(\theta_0)$. As for k_0, which controls absolute scaling, it forces $|\det(lin)|$ to be $1/F_0^{AB}(\theta_0)$. In other words, and according to (4) and (9), it forces the maximum value of the function $F_0^{lin(A)\,lin(B)}$ to be 1. Normalizing transformations satisfy (7), and this is largely due to (6).

3) Normalized descriptor: Let *lin* be the normalizing transformation of Φ^{AB}. The elements of the tuple $\Phi^{lin(A)\,lin(B)}$ can be derived from *lin* and the elements of Φ^{AB} using (1) to (5). The descriptor $\Phi^{lin(A)\,lin(B)}$ is denoted by $\overline{\Phi^{AB}}$. The pair *(lin(A), lin(B))* is the *normalized object pair* and $\overline{\Phi^{AB}}$ is the *normalized descriptor*. More precisely, $\overline{\Phi^{AB}}$ is the result of Φ^{AB}'s normalization with respect to the affine group. For any affinity *aff*, we have:

$$\overline{\Phi^{aff(A)\,aff(B)}} = \overline{\Phi^{AB}} \tag{26}$$

This is according to (7). In other words, a relative position normalized with respect to the affine group does not change when the objects are transformed by a member of the group. Note that:

$$\overline{\overline{\Phi^{AB}}} = \overline{\Phi^{AB}} \tag{27}$$

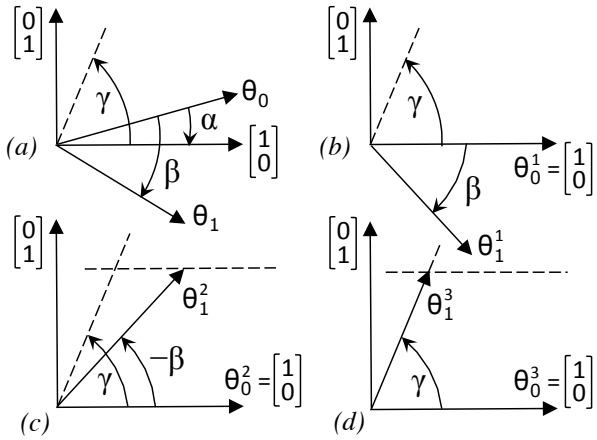

Fig. 2. Angles, directions, and linear transformations. lin_1 transforms *(a)* into *(b)*, lin_2 transforms *(b)* into *(c)*, and lin_3 transforms *(c)* into *(d)*.

Equations (6), (26) and (27) are illustrated by Fig. 3. The first object of each pair (the 'I') has two convex connected components. The second object (the 'A') is connected and has a hole. The two objects overlap.

The Φ-descriptor can also be normalized with respect to subgroups of the affine group. For example, replace lin_3 with the identity transformation, and set k and k_1 to 1: the normalizing transformation *lin* is then a similitude, and (26) holds for similitudes only. Or, replace lin_3, lin_4, lin_5 and lin_6 with the identity transformation: *lin* is then an isometry, and (26) holds for isometries only.

4) Comparing relative positions: In practice, relative positions are compared through a similarity measure, *SIM*, and (28) is used instead of (26):

$$SIM(\overline{\Phi^{AB}}, \overline{\Phi^{aff(A)\,aff(B)}}) \approx 1 \qquad (28)$$

The greater the left-hand side of the equation, the more similar the relative positions: 1 indicates complete similarity and 0 complete dissimilarity. See Fig. 4 and Table II. In that table, the similarity measure is the measure presented in Section V.A.

C. Retrieving aff from Φ^{AB} and $\Phi^{aff(A)\,aff(B)}$

Let (A_0,B_0) be a pair of objects. Solve for (A,B) the equation $\Phi^{AB} = \Phi^{A_0B_0}$. Any pair (A,B) such that $(A,B) \equiv (A_0,B_0)$, where ≡ means equality up to a translation, is a solution to this equation. Are there other solutions? We do not have an answer to this *recovery problem* yet. In practical situations, however, it is most reasonable to assume that the answer is negative.

Now, consider normalization with respect to the affine group. Let (A,B) and (A',B') be two well-behaved object pairs. Assume there exists an affinity *aff* such that

$$A' = aff(A) \text{ and } B' = aff(B). \qquad (29)$$

Is it possible to retrieve *aff* from Φ^{AB} and $\Phi^{A'B'}$? The answer is positive. Indeed, the normalizing transformations *lin* and *lin'* of Φ^{AB} and $\Phi^{A'B'}$ are such that $\Phi^{lin(A)\,lin(B)} = \Phi^{lin'(A')\,lin'(B')}$ (Section III.B). Therefore,

$$\text{lin'(A')} \equiv \text{lin(A)} \text{ and } \text{lin'(B')} \equiv \text{lin(B)} \qquad (30)$$

(subject to the assumption regarding the recovery problem). This implies that

$$A' \equiv lin'^{-1}(lin(A)) = (lin'^{-1} \circ lin)(A) \text{ and} \qquad (31)$$

$$B' \equiv lin'^{-1}(lin(B)) = (lin'^{-1} \circ lin)(B), \qquad (32)$$

where ∘ denotes function composition. In other words,

$$aff \equiv lin'^{-1} \circ lin. \qquad (33)$$

For example, consider the two object pairs in Fig. 4*ad*. Compute their Φ-descriptors, and from there the two normalizing transformations. Then, apply (33). You find the transformation we used to get Fig. 4*d* from Fig. 4*a*:

$$aff \equiv \begin{bmatrix} 0.12 & 0.86 \\ -1.10 & 0.64 \end{bmatrix}$$

Now, instead of (29), assume there exists an invertible transformation *t*, not necessarily affine, such that:

$$A' = t(A) \text{ and } B' = t(B). \qquad (34)$$

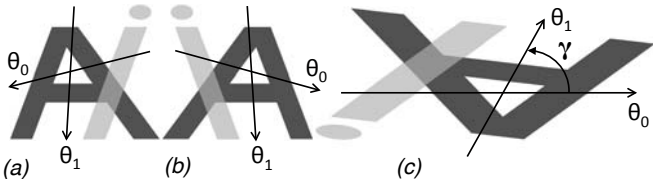

Fig. 3. Normalization wrt the affine group. *(a)(b)(c)* Three object pairs with their major and minor directions. Since the pairs are related by affinities, they have the same normalized Φ-descriptor, and they yield the same normalized pair — which is actually the pair shown in *(c)*. The normalizing transformation for that pair is the identity.

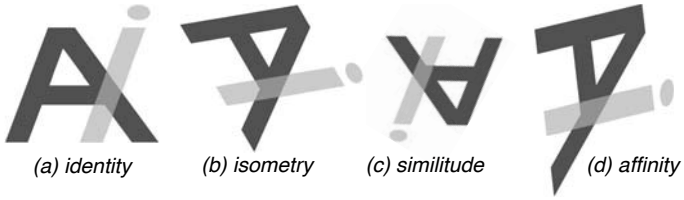

Fig. 4. The identity transformation, an isometry other than the identity, a similitude other than an isometry, and an affinity other than a similitude are applied to a pair of objects.

TABLE II. SIMILARITIES AND AFFINITIES.

Consider the four object pairs in Fig. 4. Normalized relative positions are compared through a similarity measure, SIM (Section V.A).

normalization with respect to	similarity between (a) and (a)	(b)	(c)	(d)
isometries	1.00	1.00	0.79	0.83
similitudes	1.00	1.00	1.00	0.83
affinities	1.00	1.00	1.00	1.00

$lin'^{-1} \circ lin$ may be seen as the affinity that best approximates *t*, and the similarity $SIM\left(\overline{\Phi^{AB}}, \overline{\Phi^{A'B'}}\right)$ between $\overline{\Phi^{AB}}$ and $\overline{\Phi^{A'B'}}$, where *SIM* denotes a similarity measure as discussed in Section III.B.4, can be used to assess the quality of the approximation. This will be illustrated in Section V. Note that normalization with respect to subgroups of the affine group leads to comparable results. For example, consider (34) again. Using normalization with respect to the group of all isometries, one can find the isometry that best approximates *t* and assess the quality of the approximation.

IV. TEST ONE

Let Δ be a relative position descriptor. Consider two views of a scene with five identical items (Fig. 5*ab*). Segment them automatically (Fig. 5*cd*). Can we match the objects in the two views correctly if the two sets of descriptors $\{\Delta^{A_iA_j}\}_{1\le i<j\le 5}$ and $\{\Delta^{B_iB_j}\}_{1\le i<j\le 5}$ are the only information available? This experiment was proposed in [3] to illustrate, test and validate the affine properties of another relative position descriptor—the force histogram. As shown in Section IV.A, the Φ-descriptor passes the test. It is, therefore, fairly robust to departures from the assumptions on the transformations being handled. Indeed, the image formation process involves projective transformations instead of affinities; besides, the objects here are 2-D representations of 3-D items and segmentation is seldom accurate. The affine properties of the Φ-descriptor and force histogram are compared in Section IV.B.

A. Which Can Is Which?

Consider the 10 descriptors $\Phi^{A_iA_j}$ and the 10 descriptors $\Phi^{B_iB_j}$, with $1 \le i < j \le 5$. There exist 10! one-to-one mappings from the first set to the second set of descriptors, i.e., 3,628,800 ways to match the two views. Of course, the search space could be drastically reduced since the descriptors in each set are not totally independent (they describe object pairs that share the same five objects). As in [3], this is a fact we choose to ignore. Let m be one of the 10! mappings. Consider some descriptor $\Phi^{A_iA_j}$. Assume $m(\Phi^{A_iA_j}) = \Phi^{B_kB_l}$. The affinity that best transforms (A_i,A_j) into (B_k,B_l) can be derived from the two descriptors—see (33), where normalization is with respect to the direct affine group (the group of affinities that preserve orientation). Ten affinities,

$$\begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix}$$ with $1 \le i \le 10$, can therefore be attached to m.

Compute the sum $a+b+c+d$, where a is the mean deviation of $a_1, a_2, \ldots, a_{10}$, where b is the mean deviation of $b_1, b_2, \ldots, b_{10}$, etc. The one true mapping defined by $m(\Phi^{A_iA_j}) = \Phi^{B_iB_j}$ should give the lowest of the 3,628,800 sums—and it does.

B. Φ-Descriptor vs. Force Histogram

The force histogram offers a solution to the problem examined in Section III.A, but only normalization with respect to the group of similitudes can be achieved, and there is no solution to the problem addressed in Section III.C [7]. A computationally expensive optimization algorithm was therefore used in [3] to compute each affinity. Moreover, the one true mapping could not be found when using only the retrieved affinities: similarities between force histograms had to be computed as well and used together with the affinities.

V. Test Two

Assume we are looking—within a database of objects—for objects that are arranged in space in a given way. For example, find five buildings in Fig. 6 that are arranged as specified by Fig. 7*a*. In this section, we present a solution based on the Φ-descriptor, with the aim of illustrating its affine properties. At the core of the search algorithm is a measure of similarity between two *configurations*, i.e., tuples of objects. This measure is introduced in Section V.B. It relies on *SIM*, which was mentioned in Section III, and an example of which is given in Section V.A. Experimental results are in Section V.C.

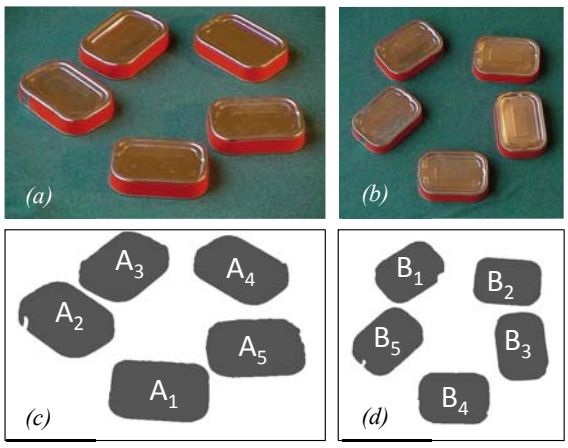

Fig. 5. *(a)(b)* Two views of the same scene. Which can is which? Use only relative position information. *(c)(d)* After segmentation. Note that A_1 corresponds to B_1, that A_2 corresponds to B_2, etc.

A. Comparing Relative Positions

In this paper, the similarity $SIM(\Phi^{AB}, \Phi^{A'B'})$ between two descriptors Φ^{AB} and $\Phi^{A'B'}$ is set to a weighted average of the similarities $sim(F_i^{AB}, F_i^{A'B'})$ between corresponding area functions, where the integer i belongs to 1..10. The weight attached to $sim(F_i^{AB}, F_i^{A'B'})$ is $\Sigma_k \left| F_i^{AB}(\theta_k) + F_i^{A'B'}(\theta_k) \right|$. Moreover,

$$sim(F_i^{AB}, F_i^{A'B'}) = 1 - \frac{\Sigma_k \left| F_i^{AB}(\theta_k) - F_i^{A'B'}(\theta_k) \right|}{\Sigma_k \left| F_i^{AB}(\theta_k) + F_i^{A'B'}(\theta_k) \right|}. \quad (35)$$

θ_1, θ_2, etc., are the directions that are actually considered, and (35) comes from [8]. We assume 0/0=0. See Fig. 4 and Table II.

B. Comparing Configurations

Here, relative positions are understood as being invariant under direct similitudes. The similarity between two configurations $(A_1, A_2, \ldots, A_n)$ and $(B_1, B_2, \ldots, B_n)$ may be set to the average of the similarities between corresponding normalized Φ-descriptors, where normalization is with respect to the group of all direct similitudes:

$$S((A_1,A_2,\ldots,A_n),(B_1,B_2,\ldots,B_n)) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} SIM(\overline{\Phi^{A_iA_j}}, \overline{\Phi^{B_iB_j}}) \quad (36)$$

Equation (36), however, is not satisfactory. For example, we may have $SIM(\overline{\Phi^{A_iA_j}}, \overline{\Phi^{B_iB_j}}) = SIM(\overline{\Phi^{A_kA_l}}, \overline{\Phi^{B_kB_l}}) = 1$ while the angles δ_{ij} and δ_{kl} or the scale factors σ_{ij} and σ_{kl} of the similitudes that transform (A_i, A_j) into (B_i, B_j) and (A_k, A_l) into (B_k, B_l) are completely different. (These similitudes are derived from the normalizing transformations, as explained in Section III.C.) Consider the multiset $\{\delta_{ij}\}_{i<j}$. Its median δ can be defined as the value δ_{uv} that gives the smallest mean deviation:

$$\delta = \arg\min_{u,v} d_{rot}(\delta_{uv}), \text{ where} \quad (37)$$

$$d_{rot}(\delta_{uv}) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left| \delta_{ij} - \delta_{uv} \right|. \quad (38)$$

However, since we are dealing with angles, i.e., circular data [2], Equation (38) must be replaced with (39):

$$d_{rot}(\delta_{uv}) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\pi - \left| \pi - \left| \delta_{ij} - \delta_{uv} \right| \right| \right) \quad (39)$$

Angles are in $(-\pi,\pi]$. The δ-angle rotation may be seen as the rotation that best transforms $(A_1, A_2, \ldots, A_n)$ into $(B_1, B_2, \ldots, B_n)$. The lower each $\pi - \left| \pi - \left| \delta_{ij} - \delta \right| \right|$, and the more similar the two configurations. Likewise, and since we are dealing with scale factors, let σ be the logarithmic median of the multiset $\{\sigma_{ij}\}_{i<j}$:

$$\sigma = \arg\min_{u,v} d_{sca}(\sigma_{uv}), \text{ where} \quad (40)$$

$$d_{sca}(\sigma_{uv}) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left| \log\sigma_{ij} - \log\sigma_{uv} \right|. \quad (41)$$

The lower each $\left| \log\sigma_{ij} - \log\sigma \right|$, i.e., each $\exp\left| \log\sigma_{ij} - \log\sigma \right| = \exp\left| \log(\sigma_{ij}/\sigma) \right| = (\sigma_{ij}/\sigma)^{\pm 1}$, the more similar the configurations. In the end, the similarity between $(A_1, A_2, \ldots, A_n)$ and $(B_1, B_2, \ldots, B_n)$ is defined as:

$$S((A_1,A_2,\ldots,A_n),(B_1,B_2,\ldots,B_n)) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ij}\, SIM(\overline{\Phi^{A_iA_j}}, \overline{\Phi^{B_iB_j}}), \quad (42)$$

where $w_{ij} = w_{rot}(\pi - \left|\pi - \left|\delta_{ij} - \delta\right|\right|)\, w_{sca}(\exp\left|\log(\sigma_{ij} / \sigma)\right|)$ (43)
and w_{rot} and w_{sca} are weighting functions as in Fig. 8.

C. Experimental Results

Consider the reference objects in Fig. 6 and a query configuration (A_1, A_2, A_3, A_4, A_5) as in Fig. 7*a*. The result of the query is the tuple of reference objects (B_1, B_2, B_3, B_4, B_5) with the highest similarity $S((A_1, A_2, A_3, A_4, A_5), (B_1, B_2, B_3, B_4, B_5))$.

We trust the results do not go against the reader's perception. Note that the exact shapes of the objects do not matter, since the focus is on their arrangement in space. For instance, the result of the query in Fig. 7*a* is Fig. 7*d*, even though there are many 5-tuples of perfectly rectangular reference objects (see, e.g., top-right corner of the map). Likewise, the result of the query in Fig. 7*c* is Fig. 7*f*, even though there is a perfect hollow rectangle in the set of reference objects (left side of the map).

VI. Conclusion

The affine properties of visual descriptors have been the subject of much attention. Here, we have focused on the Φ-descriptor—a relative position descriptor introduced in recent work. It is much faster to compute than most of its competitors, and it carries way more spatial relationship information. We have shown in this paper that the Φ-descriptor has remarkable affine properties, which give it unique capabilities. For example, consider the relative positions Φ^{AB} and $\Phi^{aff(A)aff(B)}$ of two spatial objects, before and after the affinity *aff* is applied to the objects. Any two of these elements allow the third one to be recovered, without relying on any computational optimization technique. Moreover, the Φ-descriptor can be normalized so as to become invariant under the affine group, or under a subgroup of the affine group (like the group of similitudes). No other relative position descriptor has comparable properties. Finally, our preliminary experiments tend to show that the Φ-descriptor reacts well to nearly-affine transformations, is robust to variations in shapes, and can be used to define perceptually relevant measures of relative position similarity between spatial configurations. In future work, we will conduct thorough experiments to corroborate and refine these findings, we will describe how to detect and handle ill-behaved object pairs (Section III.B.1), and we will investigate the recovery problem (III.C).

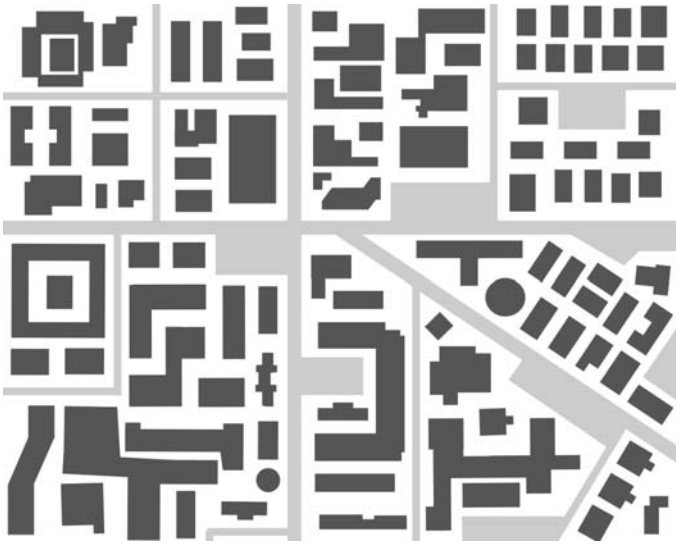

Fig. 6. A map with roads, parking lots, and about 100 buildings.

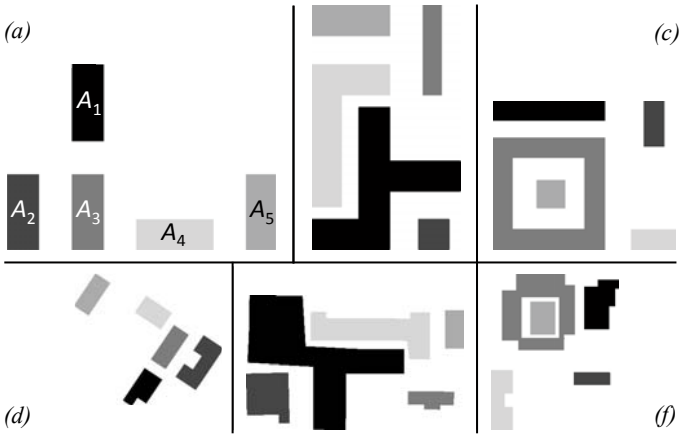

Fig. 7. Top row: queries. Bottom row: top matches from the map in Fig. 6. Note the independence from scale and orientation.

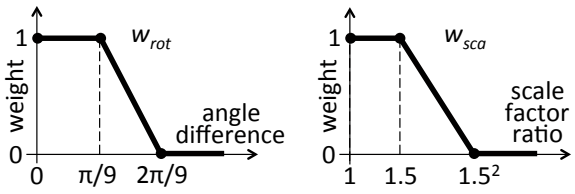

Fig. 8. The weighting functions w_{rot} and w_{sca}.

References

[1] A.R. Buck, J.M. Keller, "A memetic algorithm for matching spatial configurations with the histogram of forces," *IEEE Trans. on Evolutionary Computation*, 17(4):588-604, 2013.

[2] N.I. Fisher, *Statistical Analysis of Circular Data*, Cambridge University Press, New York, 1993.

[3] P. Matsakis, J. Keller, O. Sjahputera, J. Marjamaa, "The use of force histograms for affine-invariant relative position description," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(1):1-18, 2004.

[4] P. Matsakis, M. Naeem, "Fuzzy models of topological relationships based on the Φ-descriptor," *Proc. of the 2016 IEEE Int. Conf on Fuzzy Systems (FUZZ-IEEE)*, in press.

[5] P. Matsakis, M. Naeem, F. Rahbarnia, "Introducing the Φ-descriptor: A most versatile relative position descriptor," *Proc. of the 2015 Int. Conf. on Pattern Recognition Applications and Methods (ICPRAM)*, 87-98.

[6] M. Naeem, P. Matsakis, "Relative position descriptors: A review," *Proc. of the 2015 Int. Conf. on Pattern Recognition Applications and Methods (ICPRAM)*, 286-95.

[7] J. Ni, P. Matsakis, "An equivalent definition of the histogram of forces: Theoretical and algorithmic implications," *Pattern Recognition*, 43(4): 1607-17, 2010.

[8] C. Pappis, N. Karacapilidis, "A comparative assessment of measures of similarity of fuzzy values," *Fuzzy Sets and Systems*, 56:171-4, 1993.

[9] K.C. Santosh, B. Lamiroy, L. Wendling, "Symbol recognition using spatial relations," *Pattern Recognition Letters*, 33(3):331-41, 2012.

[10] C.R. Shyu, M. Klaric, G.J. Scott, A.S. Barb, K. Palaniappan, "GeoIRIS: Geospatial information retrieval and indexing system: Content mining, semantics modeling, and complex queries," *IEEE Trans. on Geoscience & Remote Sensing*, 45(4):839-52, 2007.

[11] M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, D. Brock, "Spatial language for human-robot dialogs," *IEEE Trans. on Systems, Man, and Cybernetics (Part C)*, 34(2):154-67, 2004.

[12] M. Yang, K. Kpalma, J. Ronsin, "A survey of shape feature extraction techniques," *Pattern Recognition*, 43-90, 2008.