

RECOGNITION OF PLANAR OBJECTS OVER COMPLEX BACKGROUNDS USING LINE INVARIANTS AND RELEVANCE MEASURES

Serguei Startchik, Christian Rauber, Thierry Pun
(University of Geneva, CH-1211 Geneva 4, Switzerland)
E-mail: {startchi, rauber, pun}@cui.unige.ch; <http://cuiwww.unige.ch>

Abstract

This paper addresses the robust recognition of planar polygonal objects situated in 3D space over highly textured backgrounds, where each object is modeled by a set of “five-lines” projective invariants. The main contributions of this work are the following: the establishment of the discriminative ability of an indexing space based on five-lines invariants, the presentation of a robust mechanism for extracting relevant line segments on which invariants are computed, the use of these invariants for geometric hashing amongst possible objects from the model base, and the verification of hypotheses through a purposive search for missing line segments. Finally, experiments verify that the indexing complexity remains linear when the size of the model base increases linearly. Experiments are presented using a model base of 10 shapes, with about 10 views for each shape acquired from various points of view.

1. Introduction

Geometric invariants are, in theory at least, valuable features to characterize a geometric structure because of their stability under various imaging conditions [9]. Given the knowledge of the applied camera transformation, the type of extracted features and the geometric structure of the analyzed shape, various invariants can be defined [8]. For example, a planar shape characterized by five points and examined through perspective projection may provide two invariant values [5]. Features such as lines [12] or corners [1] offer a more robust and precise positional information; line invariants therefore appear more suited for reliable indexing of objects from a model-base. Other primitives like ellipses can be extracted (or fitted) to characterize the shape [3].

The fact that primitives have already been extracted to compute invariant measures, makes invariant-based object recognition systems usually based on geometric hashing techniques [6][11]. In this case, there are two major concerns. The first is that the indexing space must provide enough discrimination capability. The second is that the indexing function should remain simple enough so as not to excessively increase recognition time. This property would be of particular importance in applications such as object tracking [13].

The present paper concentrates on following problems. Section 2. discusses the learning of invariant representations, based on a “five-lines” invariant, for automatic construction of object model. Simultaneously, an indexing invariant space is built and it is shown that such space can offer a reasonable segregation between class shapes. Section 3. presents the use of a relevance measure to facilitate the extraction of line segments; this heuristic greatly contributes to bounding the complexity of the recognition process. Hypothesis indexing and verification is treated in Section 4., and consists of the following stages: invariant values calculation for a group of lines, indexation of a shape hypothesis from these values, and verification of the indexed hypothesis by a guided search for missing line segments.

2. Learning Invariant Object Models from Connected “five-lines” Groups

The images consist of planar shapes lying on highly textured backgrounds; the optical axis of the camera is not constrained to be perpendicular to the image plane (Figure 1). The model base comprises P polygonal objects $p = 1 \dots P$ (currently $P = 10$), each being composed of N_p line segments numbered $n = 1 \dots N_p$ (N_p varies from 6 to 12 with the current objects).

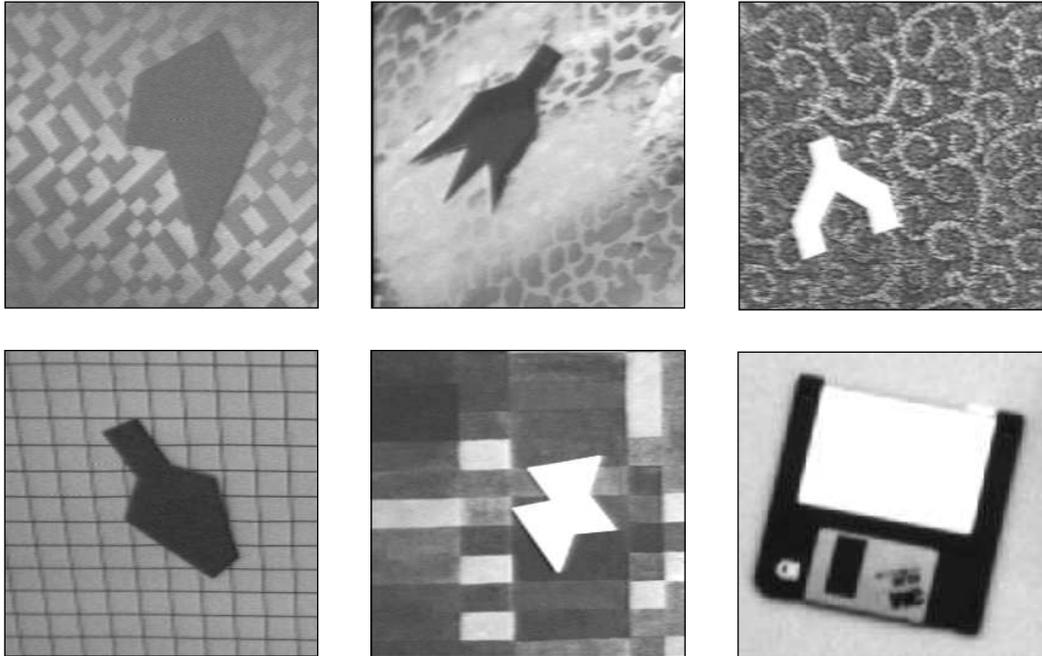


Fig. 1 Examples of planar polygonal objects on complex backgrounds. The total size of the database is currently 10 objects, with about 10 views for each object acquired on differing backgrounds, with differing orientations in 3D space.

Each group of five coplanar lines can be characterized by two scalar invariants [4] [5], so for each polygon p there are potentially $C_5^{N_p}$ possible five-lines combinations for which two invariant values can be evaluated. To avoid using such a high number of combinations for each shape representation, only groups of five *connected*, consecutive segments are used (two segments are connected if the intersection of their support lines lie near one of their extremities). This connectivity heuristic reduces the number of five-lines groups to be considered to N_p without diminishing robustness of the system in case of occlusions: topology being preserved under perspective projection, connected groups would be more likely to appear as a whole and therefore are more valuable for indexing. Thus, each polygonal object is characterized by a set of N_p points in the indexing space. Therefore, to a given group of five lines starting from the n -th segment on the contour of polygon p corresponds one point \mathbf{I}_n^p in the 2D *indexing space* [4]:

$$\mathbf{I}_n^p = \left[\frac{\det(M_{431}) \det(M_{521})}{\det(M_{421}) \det(M_{531})}, \frac{\det(M_{421}) \det(M_{532})}{\det(M_{432}) \det(M_{521})} \right], \quad (1)$$

where M_{ijk} is a matrix of three column vectors i, j, k , with f.ex. vector i describing line $n + i - 1$ by its three (non-independent) polar parameters $[\cos\theta, \sin\theta, -\rho]_{n+i-1}$.

During training, a random selection of K views from each object class p (i.e. same object, K different projections) are processed as described in Section 3., yielding polygonal contour shapes (Figure 2.a). Since to each view corresponds a set of N_p points in the invariant space, the whole selection of views yields KN_p points in this indexing space (Figure 2.b). Figure 2.b shows that these points are rather well segregated into

N_p clusters corresponding to the N_p possible five-lines groups (or *classes*), which is a very desirable behavior for indexing. Furthermore, the within-class scatter for each group is usually significantly lower than the average inter-class distance computed over all $\sum_p N_p$ classes, which confirms that this indexing space can be used for segregating between five-lines groups (Figure 2.c).

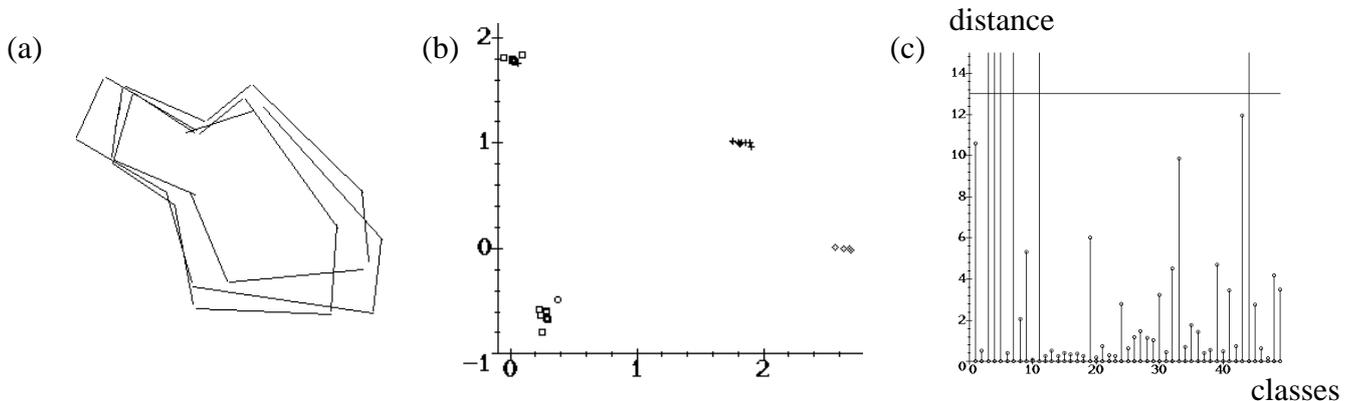


Fig. 2 (a) Polygonal outlines of several views of the “snakehead” object shown in Figure 1. (the extraction method is discussed in Section 3.). (b) The 2D indexing space spanned by the invariant values, where the clusters correspond to five-lines group from the example in (a). Points representing various instances of a given five-lines groups are drawn with the same symbol; each cluster usually corresponds to one particular five-lines group, or *class*. (c) Within-class scatter for each class (vertical bars) and average between-class distance (horizontal line), computed for all objects and all views from the model base.

3. Extracting Line-invariants using Relevance

The work described in this paper is part of a larger project whose goal is to build an object recognition system, based on the concept of relevance [2] [10], on a focus-of-attention mechanism [7], and on geometric hashing with invariant measures. The guiding idea is the transformation of a static image into several dataflows of image primitives (e.g. for line segments, for circular arcs, and for regions). The ranking of the primitives in each flow, that is the delay of each one with respect to the others, is a function of the quantitative *relevance* of each primitive for the recognition.

An original aspect of this system, not discussed in this paper, is a dynamic indexing scheme which benefits from the asynchronous “arrival” of primitives in the dataflow: the most relevant (early) primitives in the dataflow generate object hypotheses, whereas less relevant primitives are delayed in time rather than eliminated, and might be reconsidered after expiration of this delay.

Two problems arise when analyzing objects over complex backgrounds: how to distinguish pertinent from noisy information, and how to obtain primitives that are significant for recognition. This is where relevance plays a role, as it allows to segregate pertinent contours from useless ones. The relevance $\rho(\tau_i) \in [0, 1]$ of a contour token τ_i is defined by [2]:

$$\rho(\tau_i) = r(\tau_i) \cdot s(\tau_i) \tag{2}$$

where $r(\tau_i)$ is the *reliability* of τ_i and $s(\tau_i)$ its *significance*. Reliability values are the normalized averages of various attributes (e.g. length and contrast for segments) throughout the whole image; a high reliability indicates that the token is a meaningful entity, unlikely to have been generated by segmentation artifacts. The significance value measures the uniqueness of a token in the image; it is maximum when the attributes of τ_i , such as length and orientation for a segment, make τ_i unique over the whole image. Tokens with high significance are thus likely to originate from objects rather than from textured background patterns.

Multiplication rather than addition is used to combine reliability and significance since this guarantees that the measure is bounded by $[0, 1]$. Figure 3 shows the result of the relevance computation applied on one of the test images: the polygonal outline is clearly distinguished from the background. This relevance heuristic will therefore play a fundamental role in keeping within linear bounds the complexity of the indexing process.

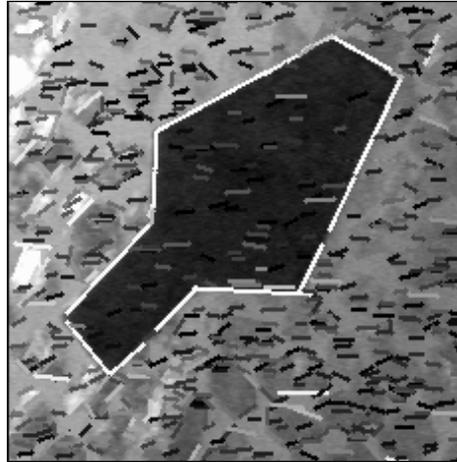


Fig. 3 Most relevant line segments (white for highest relevance), clearly corresponding to the contour of the polygonal shape.

4. Hypotheses Indexing and Verification

The indexing of object shapes and their verification is performed as follows. First, the image is segmented into contours, whose relevance is evaluated. Second, a search for a group of five connected segments is performed, using as seed the most relevant line segment. If the search initiated with the most relevant segment does not succeed, it iterates by using as seeds the other segments, by order of decreasing relevance.

Third, and this constitutes the indexing phase, the pair of invariant values \mathbf{I} is computed for the five-lines group that has been obtained. Using \mathbf{I} as an index into the learnt indexing space \mathbf{I}_n^p , it is possible to determine the most likely object hypothesis p and the five-lines group n to which the current group corresponds, as being the closest learnt class in terms of Euclidean distance (the complexity issue related to this indexing is discussed below).

Fourth, once a hypothesis p has been indexed, five lines are already matched to the n -th to $n+4$ -th polygon sides. The presence of the remaining line segments belonging to object p should be verified in the image; this constitutes the verification phase. Using the last four of those lines ($n+1 \dots n+4$) and the known (learnt) invariant \mathbf{I}_{n+1}^p corresponding to these four and the next connected line $n+5$ on the contour of object p , two equations can be obtained from (1):

$$\begin{cases} a_{n+5}k^a_1 + b_{n+5}k^b_1 + c_{n+5}k^c_1 = 0 \\ a_{n+5}k^a_2 + b_{n+5}k^b_2 + c_{n+5}k^c_2 = 0 \end{cases} \quad (3)$$

where the parameters $\{a_{n+5}, b_{n+5}, c_{n+5}\}$ characterize the fifth unknown line, and k^j_i are coefficients that depend only on the four known lines and two known invariant values. The line equation (3) defines a search zone, whose longitudinal boundaries are derived from the variance of invariant values (Figure 4.b).

The endpoints of each segment being searched are provided by the knowledge of the linear search zones from the previous and next segments (Figure 4.c). The hypothesis verification must therefore proceed in two passes, first by determining the supports of the line segments, then by using these supports to determine the exact length of the segments. When the search zone contains multiple, fragmented segments, a further

grouping operation is performed in order to combine them into a single line (Figure 4.c).

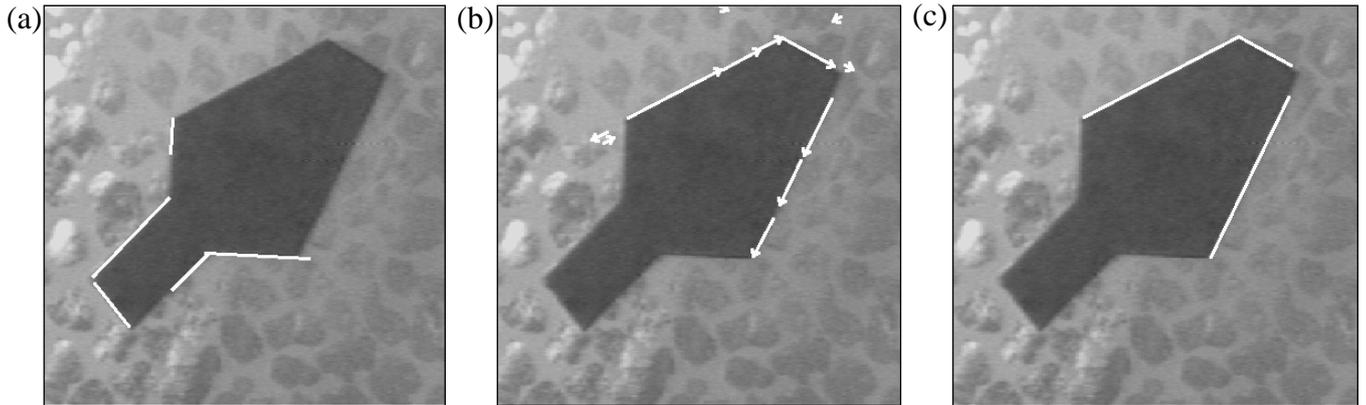


Fig. 4 (a) Initial group of five segments whose invariant is used as index. (b) Result of the first pass of the hypothesis verification by parametric search for missing segments: each found segment is represented by one arrow. Small spurious segments with the correct orientation are obtained; in addition, the two largest segments are broken. (c) Result after the second pass: only the correct segments remain, and the broken segments have been grouped.

In order to assess the increase in complexity of the indexing process when increasing the size of the model base, the following experiment was performed. An *indexing distance* measure has been defined, as the number of hypotheses that have to be verified before reaching the correct solution. Since indexing is made in the invariant space where a Euclidean distance is used, the indexing distance between a measured point I and the correct hypothesis I_n^p is given by the total number of learnt points that lie closest to I than I_n^p .

This total number is computed as the number of points falling in a rectangle of center I and half-diagonal $I - I_n^p$. The larger the indexing distance, the higher the number of hypotheses that will have to be verified.

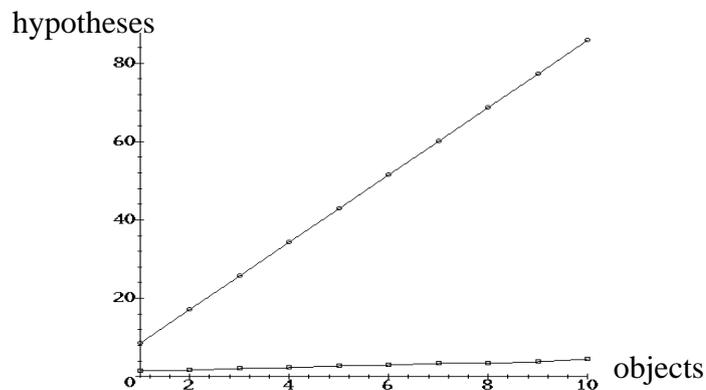


Fig. 5 Complexity evaluation when increasing the number of models. Upper line: number of hypotheses in the database as a function of the number of objects; lower line: number of false hypotheses that should be analyzed until a correct one is found.

The average indexing distance has been experimentally evaluated over all classes of the training set: for one class, the number of hypotheses to be tested was determined as the average number of points between one five-lines group and the same line group in different examples. This experiment was repeated for $P = 1$ to $P = 10$ objects in the model base. The upper curve in Figure 5 illustrates the increase in number of classes in the database, while the lower curve represents the growth of the indexing distance. These results indicate that the number of hypotheses to be evaluated only increases linearly with the number of objects in the model base; furthermore, the ratio of slopes is significantly less than one (about 0.037).

5. Conclusion

The role of the relevance measure and connectivity criterion in the drastic reduction of the number of indexed hypotheses has been shown, in the context of object recognition based on projective invariants. The indexing complexity scales up reasonably slowly (linearly) with respect to the number of invariant groups, when the number of models is increased. Future work will be directed towards the use of other geometric invariants (two lines-two points, five points, invariant frames for curves) as well as towards integration of photometric invariants (color) for hypotheses confirmation.

Acknowledgments

The authors thank J.-M. Bost and R. Milanese for their contributions, and gratefully acknowledge the support of the Swiss Priority Program in Computer Science, grant 5003-34278.

References

- [1] P.A. Beardsley, A. Zissermann, D.W. Murray, "Navigation using affine structure from motion", Proc. 3rd Europ. Conf. on Comp. Vision, Springer-Verlag, May 2-6, 1994, Stockholm, Sweden, Vol. 2, 85-96.
- [2] J.-M. Bost, R. Milanese and T. Pun, "Temporal precedence in asynchronous visual indexing", Proc. 5th Int. Conf. Comp. Anal. of Images and Patterns (CAIP'93), Springer-Verlag, Lecture Notes in C.S. 719, D. Chetverikov and W.G. Kropatsch, Eds., 1993, 468-475.
- [3] S. Carlsson, "Projectively invariant decomposition and recognition of planar shapes", Proc. 4th ICCV, Berlin, Germany, May 11-13, 1993, 471-475.
- [4] C. Coelho, A. Heller, J.L. Mundy, D. A. Forsyth, A. Zisserman, "An experimental evaluation of projective invariants", in: [9], 87-104.
- [5] D. Forsyth, J.L. Mundy, A. Zisserman, C. Coelho, A. Heller, C. Rothwell, "Invariant descriptors for 3-D object recognition and pose", IEEE Trans. Patt. Anal. Mach. Intell., 13, 10, October 1991, 971-991.
- [6] W.E.L. Grimson, Object Recognition by Computer: the Role of Geometric Constraints, MIT Press, 1990, 47-76, 447-454).
- [7] R. Milanese, H. Wechsler, S. Gil, J.-M. Bost, T. Pun, "Integration of bottom-up and top-down cues for visual attention using non-linear relaxation", Proc. IEEE - CVPR 94, Seattle, Washington, June 20-23, 1994, 781-785.
- [8] R. Mohr, "Projective geometry in computer vision", in: Handbook of Pattern Recognition and Computer Vision, C.H. Chen, L. F. Pau and P.S.P. Wang, Eds., World Scientific, 1993, 369-393.
- [9] J.L. Mundy, A. Zisserman, Eds., *Geometric Invariance in Computer Vision*, MIT Press, Cambridge Ma, 1992.
- [10] T. Pun, J.-M. Bost, R. Milanese, C. Rauber, S. Startchik, "Selecting relevant information and delaying irrelevant data for objects recognition", Proc. AAAI Fall Symp. Series, Relevance Workshop, New Orleans, LA, 4-6 Nov. 1994, 168-172.
- [11] C.A. Rothwell, "Hierarchical object description using invariants", Second Joint European-US Workshop on Application of Invariance in Computer Vision, Azores, Portugal, 1994, p. 397-414.
- [12] F.C.D. Tsai, "Geometric hashing with line features", Pattern Recognition, 27, 3, 1994, 377-389.
- [13] S. Vinther, R. Cipolla, "Active 3D object recognition using 3D affine invariants", Proc. 3rd Europ. Conf. on Comp. Vision, Springer-Verlag, May 2-6, 1994, Stockholm, Sweden, Vol. 2, 15-24.