A Stable Graph-Based Representation for Object Recognition through High-Order Matching

A. Albarelli, F. Bergamasco, L. Rossi, S. Vascon and A. Torsello
Università Ca' Foscari Venezia
www.dais.unive.it

Abstract

Many Object recognition techniques perform some flavour of point pattern matching between a model and a scene. Such points are usually selected through a feature detection algorithm that is robust to a class of image transformations and a suitable descriptor is computed over them in order to get a reliable matching. Moreover, some approaches take an additional step by casting the correspondence problem into a matching between graphs defined over feature points. The motivation is that the relational model would add more discriminative power, however the overall effectiveness strongly depends on the ability to build a graph that is stable with respect to both changes in the object appearance and spatial distribution of interest points. In fact, widely used graph-based representations, have shown to suffer some limitations, especially with respect to changes in the Euclidean organization of the feature points. In this paper we introduce a technique to build relational structures over corner points that does not depend on the spatial distribution of the features.

1. Introduction

Object recognition can be performed through many different approaches, ranging from the spatial analysis of color distribution [5] to contour matching [13]. In this paper we focus on techniques that work by matching sets of feature points that have been extracted from both a model image and a scene. For this kind of approach to succeed, the interest points should be repeatable, that means that the sets of features extracted under moderately different conditions of pose and lighting should exhibit a reasonable overlap. To this end, Harris operator [7], Maximally Stable Extreme Regions [12] or Differences of Gaussians [11] are widely used in literature. Moreover, a robust and distinctive descriptor, such as SIFT [10] or SURF [8] must be computed for each interest point in order to ensure an accurate correspondence. While techniques that rely only on the descriptor for point-wise matching can be very effective [9] and efficient [14], the adoption of spatial constraints has proven to enhance the overall reliability of the correspondences. Such constraints can be expressed, for instance, as a precise class of transformations to which the points matched must adhere [2] or as a set of relations between features that must be preserved through the injection that connects model to data points [16]. The latter class of constraints casts the search for a correspondence to a graph-matching problem, that can be solved (heuristically) with general or specialized techniques [4]. However, it is not always straightforward to choose a proper relational model for the data points. Delaunay triangulation [3] and k-nearest-neighbour graphs (where each point is connected to the k nearest points) are popular choices since they are simple to compute and exhibit a meaningful planar organization. Unfortunately, with these approaches the connectivity is usually settled by the relative position of the points in the Euclidean space, which happens to be a property that is not resilient to changes in pose and occlusions.

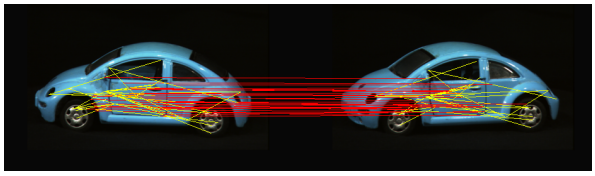

Figure 1. Matching two stable graphs

In the following section we introduce a relational structure that does not depend on the position of the points and that can be built directly on Harris corners without requiring the expensive computation of a feature descriptor. Additionally, we propose a game-theoretic matching schema that allows to find correspondences that respect the topology of the graph and that are also coherent with respect to geometrical constraints (see Fig. 1).

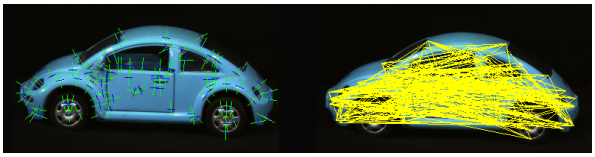

Figure 2. Building a stable graph

2. Stable Graph Creation

The vertices of the proposed graph, defined as the set $V = \{v_1, v_2 ... v_n\}$, are exactly the n image pixels that exhibit the stronger response with respect to the Harris operator [7], after non-maximal suppression and above a given threshold. For each vertex v, a simple descriptor is computed as:

$$\delta(v) = (\mathbf{e}', \mathbf{e}'', \lambda', \lambda'') \quad (1)$$

where $\mathbf{e}'$,$\mathbf{e}''$ and λ',λ'' are respectively the normalized eigenvectors and the eigenvalues of the covariance matrix of the square image patch centered on the vertex v with a given side w_c. The main idea behind the proposed representation is that the connectivity between two vertices should depend only on the descriptors involved, rather than on the positions of the vertices. To this end we define the edge set as E as:

$$E = \{(v_1, v_2) \in V \times V | \mathbf{e'_1} \cdot \mathbf{e'_2} < \alpha_d\} \quad (2)$$

this means that an edge insists between two vertices if and only if the respective main eigenvectors are orthogonal enough. From an image processing point of view this makes sense, in fact this rule avoids to connect local clusters of points that share the same edge and, by contrast, connects vertices that represent parts of the objects well apart. In the following section we will make clear how these conditions help in obtaining reliable matches. In Fig. 2 the graph building process is shown. The vertices are shown in the first half of the image along with the direction of the corresponding eigenvectors. In the second part of the same figure the stable graph built upon those vertices is displayed. Specifically, a value of α_d of 0.04 (that corresponds to about 88.5 degrees) was used. Visually, the graph seems to be very dense, however this is due to the fact that the very nature of the proposed approach tends to build edges spanning from one side of the object to the other. Indeed, the average degree of the graph shown is slightly below 4. By changing respectively the Harris response threshold and the value of α_d, graphs with a different number of vertices and with different edge densities can be obtained. Of course, the usual trade-off between the number of features and their distinctiveness applies. Fig. 3 displays a comparison between the structures produced over the same vertex set extracted from two slightly different images, respectively by the Delaunay triangulation, the knn-graph and our method. While topological changes are apparent with the first two techniques, it is not easy to tell how much of the non-planar structure is preserved by the stable graph. However, this will be assessed in the following as the three graph-based representations will be used for object recognition.

3. Game-Theoretic Matching

The Matching between graphs is performed using the game-theoretic framework presented in [1] and [2]. According to this technique it is possible to cast the correspondence problem in an optimization process that is guaranteed to converge under payoff-monotonic evolutionary dynamics [15]. Two steps are needed to exploit such framework: the preparation of a set of matching candidates that will be selected through the evolutionary process and the definition of a payoff function between pairs of candidates to drive the evolution toward the selection of a globally coherent set of matches.

3.1 Candidate Selection

Differently from [2], where correspondences are searched between well characterized features, we are dealing with rather poor descriptors (see Eq. 1). Since we trust the graph to be repeatable, we propose to create matching candidates between edges rather than vertices, thus each high-order putative match is defined by the quadruple $((v_a, v_b), (v_1, v_2))$ where (v_a, v_b) is an edge in the model and (v_1, v_1) is an edge in the data. For each edge in the model a maximum of k candidates are produced. Those candidates are selected by choosing those that exhibit a better alignment of the principal eigenvectors and similar eigenvalues between the corresponding vertices (see Fig. 4-1).

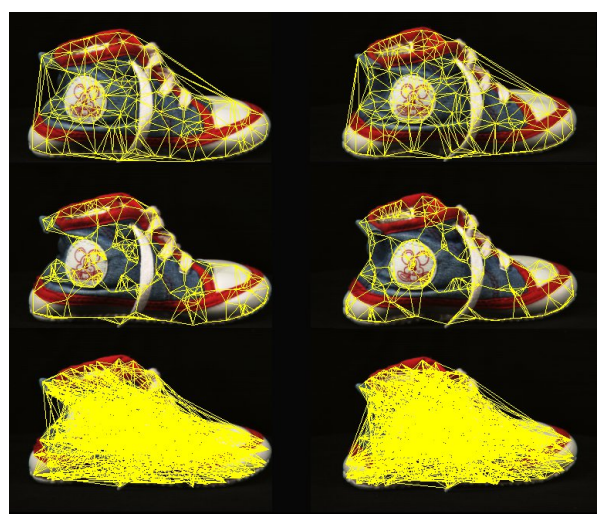

Figure 3. Comparison between graphs

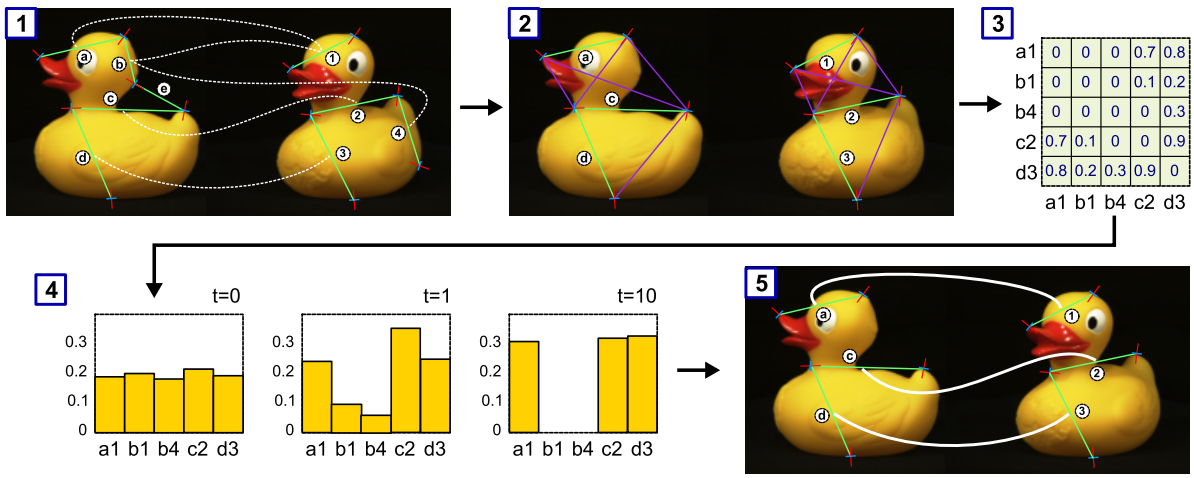

Figure 4. The proposed pipeline for stable graph matching (see text for details).

3.2 Payoff Definition

The payoff between two candidates expresses the compatibility between them. A high payoff lets two candidate to thrive together, while a payoff of value 0 prevents the evolutionary process to select both the candidates in the final population. Since we want the match to be one-to-one, we set the payoff to 0 if two candidates share the same model or data edge. Additionally, we also want to enforce topological constraints, thus we set the payoff to 0 if the incidence relation between source and data edges is not consistent. This is the case, for instance, if the model edges in two candidates share a vertex, but the corresponding edges in the data do not. If both one-to-one and topological constraints are satisfied, the payoff is computed as a function that accounts for the relative direction and length of the edges:

$$\pi\begin{pmatrix}((v_a,v_b),(v_1,v_2)),\\ ((v_c,v_d),(v_3,v_4))\end{pmatrix} = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{v\begin{pmatrix}((v_a,v_b),(v_1,v_2)),\\ ((v_c,v_d),(v_3,v_4))\end{pmatrix}^2}{2\sigma^2}} \tag{3}$$

where σ is a term that regulates the selectivity of the match and v is a function that measures the compatibility between the scaling transforms induced by the 6 distances among the involved vertices (see Fig. 4-2):

$$v\begin{pmatrix}((v_a,v_b),(v_1,v_2)),\\ ((v_c,v_d),(v_3,v_4))\end{pmatrix} = \min_{i=1..6,j=1..6} log\frac{R_i}{R_j} \tag{4}$$

with the 6 distance ratios being:

$$\begin{array}{lll} R_1 = \frac{\|v_a-v_b\|}{\|v_1-v_2\|} & R_2 = \frac{\|v_c-v_d\|}{\|v_3-v_4\|} & R_3 = \frac{\|v_a-v_c\|}{\|v_1-v_3\|} \\ R_4 = \frac{\|v_b-v_d\|}{\|v_2-v_4\|} & R_5 = \frac{\|v_a-v_d\|}{\|v_1-v_4\|} & R_6 = \frac{\|v_b-v_c\|}{\|v_2-v_3\|} \end{array} \tag{5}$$

Since we compare ratios the approach is scale independent. The log in Eq. 4 makes the measure symmetric and of 0 average, the min takes the worst value, the Gaussian of Eq. 3 maximizes the payoff when all the ratios are preserved. The computation of an example payoff matrix is shown in Fig. 4-3. The payoff between $b1, b4$ or $a1, b1$ is 0 because of the one-to-one constraint, while the payoff between $a1, b4$ or $c2, b4$ is 0 because of the topology constraint (for instance a and b share a vertex while 1 and 4 do not). Among the non-zero values, we can observe that the compatibility between $a1, c2$, $c2, d3$ and $a1, d3$ is high because all the ratios are preserved fairly well (note that the actually computable ratios over $c2, d3$ are only 3 because of the degeneration induced by the shared vertex). By contrast, the payoff between the remaining pairs of candidates is low as some ratios are poorly maintained. In fact, after a few iterations of the evolutionary game (see Fig. 4-4), candidates $b1$ and $b4$ become extinct, while $a1$,$c2$ and $d3$ have been selected as the correct matches.

4. Experimental Evaluation

In this section we analyze the object recognition performance obtained using the described pipeline with the proposed stable graph, the Delaunay triangulation and the knn-graph. We also added in the comparison the original SIFT-based affine game-theoretic matcher [2] (AGT). The dataset has been produced using 10 objects from the ALOI dataset [6]. For each object a set of different points of view was selected and random affine transformations were applied. Additionally, half of the resulting images were cluttered by applying random square patches extracted from other images in the set. One at a time, each object in the dataset was used as a query and the results were sorted according to the average payoff of the surviving population (which is expected to be higher when the match is good). In the first half of Fig. 5 the result sets obtained by the compared methods for the same query are shown. In the upper-right corner of the same figure we display a single example that shows the match preservation of our

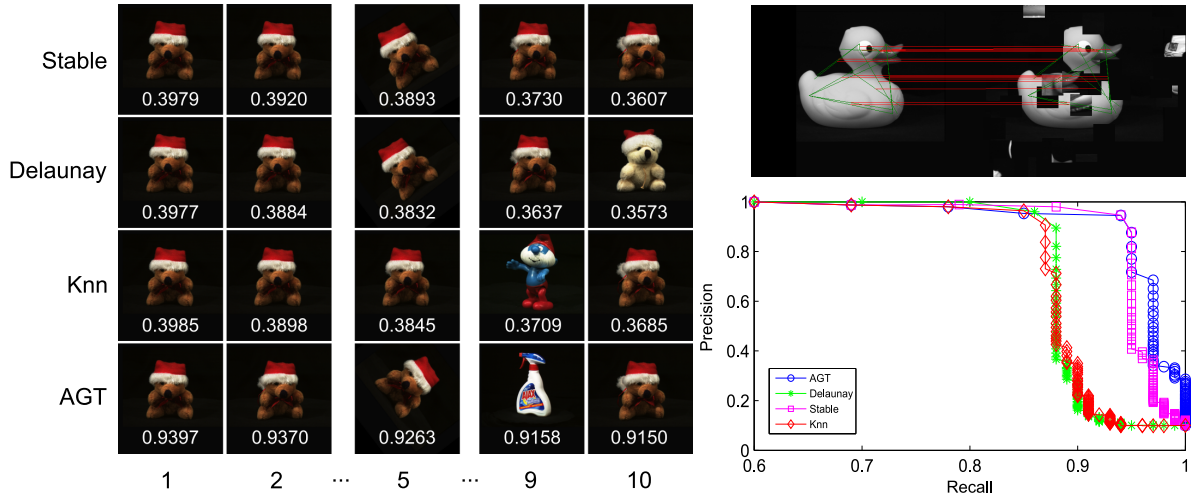

Figure 5. Experimental evaluation of the proposed method (see text for description).

method even when severe clutter is applied. Finally, we also plotted a precision/recall graph in order to supply a quantitative overview. While the stable graph offers the best relational representation for the proposed high-order matching, it can be observed that the original AGT performs sightly better for high recall rates. However, to be fair, it should be stressed that AGT adopts the robust SIFT descriptor, which requires several seconds to be computed over the tested images. Conversely, the very simple computation of the Harris response can be performed up to two orders of magnitude faster.

5. Conclusions

We introduced a novel graph-based representation that is built by connecting feature points in the space of their descriptor rather than in the Euclidean space. Despite the low distinctiveness of the characterization, good recognition results have been obtained adopting an edge-based high-order matching technique that enforces the topology of the proposed representation.

References

[1] A. Albarelli, S. R. Bulò, A. Torsello, and M. Pelillo. Matching as a non-cooperative game. In *ICCV: IEEE Intl. Conf. on Comp. Vis.* IEEE Comp. Society, 2009.

[2] A. Albarelli, E. Rodolà, and A. Torsello. Imposing Semi-Local geometric constraints for accurate correspondences selection in structure from motion: A Game-Theoretic perspective. *International Journal of Computer Vision*, pages 1–18, Mar. 2011.

[3] B. N. Boots. Delaunay triangles: An alternative approach to point pattern analysis. In *Proc. of the Ass. of American Geographers*, volume 6, pages 26–29, 1974.

[4] T. S. Caetano, T. Caelli, D. Schuurmans, and D. A. C. Barone. Graphical models and point pattern matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(10):1646–1663, Oct. 2006.

[5] S. Ekvall, D. Kragic, and F. Hoffmann. Object recognition and pose estimation using color cooccurrence histograms and geometric modeling. *Image Vision Comput.*, 23(11):943–955, Oct. 2005.

[6] J.-M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders. The amsterdam library of object images. *Int. J. Comput. Vision*, 61(1):103–112, 2005.

[7] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Fourth Alvey Vision Conference*, pages 147–151, 1988.

[8] T. T. Herbert Bay and L. V. Gool. SURF: Speeded up robust features. In *9th European Conference on Computer Vision*, volume 3951, pages 404–417, 2006.

[9] G. Kordelas and P. Daras. Robust sift-based feature matching using kendall's rank correlation measure. In *IEEE int. conf. on Image processing*, ICIP'09, pages 325–328, Piscataway, NJ, USA, 2009. IEEE Press.

[10] D. Lowe. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, volume 20, pages 91–110, 2003.

[11] D. Marr and E. Hildreth. Theory of edge detection. *Royal Soc. of London Proc. Series*, 207:187–217, 1980.

[12] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Comp.*, 22(10):761–767, 2004.

[13] J. Shotton, A. Blake, and R. Cipolla. Multiscale categorical object recognition using contour fragments. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(7):1270–1281, July 2008.

[14] S. Taylor, E. Rosten, and T. Drummond. Robust feature matching in 2.3s. In *CVPR: IEEE Computer Society Conf. on Comput.Vis. and Pat. Rec.*, pages 15–22, 2009.

[15] J. Weibull. *Evolutionary Game Theory*. MIT, 1995.

[16] K.-J. Yoon and M.-G. Shin. Reducing ambiguity in object recognition using relational information. In *Asian conf. on Computer vision*, ACCV'10, pages 293–306, Berlin, Heidelberg, 2011. Springer-Verlag.