A Hypergraph-based Approach to Affine Parameters Estimation

S. Rota Bulò, A. Albarelli, A. Torsello and M. Pelillo
Dipartimento di Informatica,
Università Ca' Foscari di Venezia, Italy
{srotabul,aalbarel,torsello,pelillo}@dsi.unive.it

Abstract

A problem commonly encountered in Computer Vision is the recovery of the transformation parameters between two affinely distorted images. In this paper, we propose a novel feature-based approach that casts the matching problem to the search of a maximum clique over an auxiliary hypergraph. We also introduce a continuous-based characterization of cliques in hypergraphs that allows us to handle the hard combinatorial problem using tools from the continuous domain. Finally, we present experimental result and comparisons with a state-of-the-art algorithm.

1. Introduction

The estimation of the transformation parameters between two affinely distorted images is a problem that is commonly encountered in many areas of Computer Vision. We can distinguish mainly two type of approaches: image-based and feature-based. The image-based approaches try to find a transformation that maximizes the overlap between the two images, usually by analyzing them in the frequency domain [2, 4, 6]. Conversely, feature-based approaches are characterized by two phases: Initially a set of features is extracted from each image and are then matched to estimate the affine transformation [3, 5, 12, 11]. In the literature, we find several feature detectors [8] and descriptors [7] that can be employed for the first phase. Among them the most widely used is the Scale-Invariant Feature Transform (SIFT) [5]. After the two sets of features have been extracted, a correspondence between them is established, and assuming that their coordinates are related by a parametric transformation, the system of equations defined by them is solved in the least square error sense to derive the transformation parameters. Unfortunately, the set of features found by real world detectors are not perfectly conserved under affine transformation, and also descriptors can lead to false matching candidates. Thereby, there is the need of filtering out wrong transformations and generally this process is accomplished by means of a generalized Hough transform, where the model parameters are quantized into bins, and each extracted correspondence between features votes for the best transformation.

The method presented in this paper falls in the class of feature-based approaches and focuses particularly on the solution of the second phase. A drawback of using a clustering mechanism to isolate the best affine transformation, is that the number of clusters is unknown a priori and a quantization of the parameter space can lead to imprecise solutions if the bins are too large, and to the dispersion of votes if they are too small. The solution that we propose overcomes these problems by casting the feature matching problem for the extraction of the affine transformation into a clique problem over an auxiliary hypergraph where the vertices correspond to feature associations and hyperedges correspond to groups of four associations that agree, within a desired tolerance, to the same affine transformation. In this way a maximum clique represents the largest group of features that agree on the same transformation. Since the maximum clique problem on hypergraphs is a relatively unexplored topic, a contribution of this paper is also to provide a continuous-based approach for it.

2. A Hypergraph Consistency Model

Before going into the details of the proposed approach, we need to introduce some notations and definitions regarding hypergraphs. Let A be a set and n a positive integer, with $\binom{A}{n}$ we denote the set of subsets of A of cardinality n. A *k-uniform hypergraph*, or simply a *k-graph*, is a pair $G = (V, E)$, where $V = \{1, \ldots, n\}$ is a finite set of *vertices* and $E \subseteq \binom{V}{k}$ is a set of *hyperedges*. Note that the concept of k-uniform hypergraph generalizes that of undirected graphs, in fact graphs are in a one-to-one relation with 2-graphs. The *complement* of a k-graph G is given by $\bar{G} = (V, \bar{E})$ where $\bar{E} = \binom{V}{k} \setminus E$. A subset of vertices $C \subseteq V$ is called

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

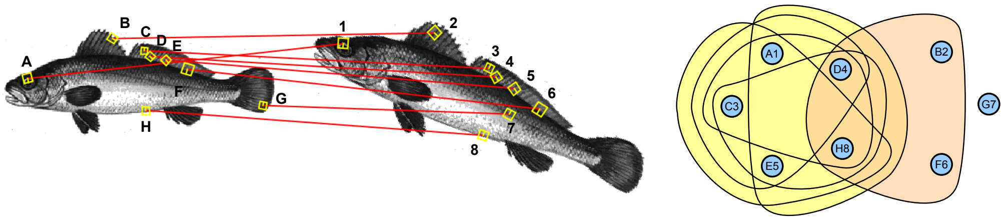


Figure 1. Example of auxiliary 4-graph for affine parameter estimation

a *clique* of G if $\binom{C}{k} \subseteq E$. A clique is said to be *maximal* if it is not contained in any other clique, while it is called *maximum* if it has maximum cardinality. We denote with $\mathcal{K}_n^k$ a complete k-graph with n vertices.

Given two sets of Euclidean features F_1 and F_2 extracted from the images, we build an auxiliary 4-graph $G = (V, E)$ where vertices are associations from $F_1 \times F_2$ and edges are groups of four correspondences that agree on the same affine transformation up to a given tolerance. In this way, a set of image features that agree, up to the specified tolerance, to the same affine transformation, form a clique of G. Thereby the problem of finding a robust set of correspondences that can be used to estimate the transformation is reduced to the problem of finding a clique on the auxiliary 4-graph with maximum cardinality. The method adopted to decide whether a set of four correspondences e agrees on a single affine transformation can be defined in many ways. Our solution consists in computing for every association $(\boldsymbol{x}, \boldsymbol{y}) \in e$ the affine transformation $(A, \boldsymbol{b})$ obtained from the remaining 3 associations and calculating the transformation error on $\boldsymbol{x}$ given by $\|A\boldsymbol{x} + \boldsymbol{b} - \boldsymbol{y}\|_2$. If the 4 distances are all below the desired threshold ε, then the hyperedge e is added to G. Note that by using this method to select edges, the user defined tolerance parameter ε is expressed in image units (e.g., pixels), which is arguably more intuitive than the parameter quantization scheme used in the generalized Hough transforms.

In order to obtain the set of feature correspondences from which to estimate the affine transformation we need to find the maximum clique in the auxiliary hypergraph G. However, in general a large maximal clique is enough. Clearly, due to the way in which G was constructed, the associations in the clique will all agree on the found affine transformation within an error of ε pixels, allowing for a very robust estimation of the parameters.

In Figure 1 we show a simplified auxiliary 4-graph generated from a very small set of SIFT features extracted from the two affinely distorted images. Each vertex represents a correspondence between features in the two images, i.e., node $A1$ represents a match between feature A of the first fish and feature 1 of the second. Note that, while correspondences $B2$, $D4$, $F6$, and $H8$ agree on the same transformation, the best group of coherent matches is represented by the set $A1$, $C3$,$D4$,$E5$, and $H8$.

3. Finding cliques in k-graphs

In 1965, Motzkin and Straus introduced a continuous characterization of cliques in graphs [9]. This result was generalized to hypergraphs in [10], where it is shown that maximal cliques of a k-graph G are in one-to-one correspondence with the local solutions of the following program:

$$\min_{\boldsymbol{x} \in \Delta} L_{\bar{G}}(\boldsymbol{x}) + \frac{1}{k(k-1)} \sum_{i=1}^{n} x_i^k \tag{1}$$

where $\Delta = \{\boldsymbol{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1,\ \boldsymbol{x} \geq \mathbf{0}\}$ is the *standard simplex* and $L_{\bar{G}}(\boldsymbol{x}) = \sum_{e \in \bar{E}} \prod_{i \in e} x_i$ is the *Lagrangian* of $\bar{G}$. More precisely, $\boldsymbol{x}$ is a local/global solution of (1) if and only if it is the characteristic vector of a maximal/maximum clique C of G, i.e. $x_i = |C|^{-1}$ for all $i \in C$ and 0 elsewhere.

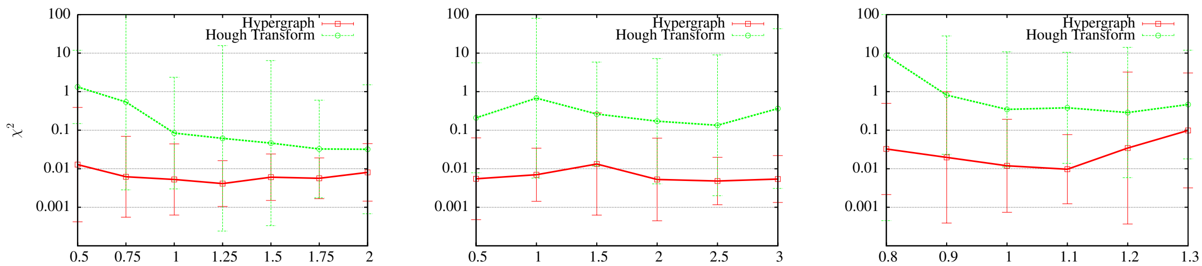
This result permits to cast clique problems on k-graphs in a continuous optimization setting. To solve Program 1 we turn it into an equivalent maximization of a homogeneous polynomial P with nonnegative coefficients over the standard simplex, where

$$P(\boldsymbol{x}) = \frac{1}{k(k-1)} \left[\left(\sum_{i=1}^{n} x_i \right)^k - \sum_{i=1}^{n} x_i^k \right] - L_{\bar{G}}(\boldsymbol{x}).$$

The function is then maximized using the discrete dynamics $x_j \leftarrow \alpha x_j \partial_j P(\boldsymbol{x})$, which, by means of the Baum-Eagon theorem [1], can be shown to be a growth transformation for P in Δ. Here, ∂_j denotes partial derivative with respect to x_j and α is the normalizing constant that projects $\boldsymbol{x}$ on Δ. By unfolding the partial derivative, we obtain

$$x_j \leftarrow \alpha x_j \left[\frac{1 - x_j^{k-1}}{k-1} - \partial_j L_{\bar{G}}(\boldsymbol{x}) \right]. \tag{2}$$

Since in our experiments $|E| \ll |\bar{E}|$, the computation of $\partial_j L_{\bar{G}}(\boldsymbol{x})$ is expensive, however, we can express



(a) Scale (b) Rotation (c) Aspect Ratio

Figure 2. Performace of the algorithms

this partial derivative in terms of G using the fact that $\partial_j L_{\bar{G}}(\boldsymbol{x}) = \partial_j L_{\mathcal{K}_n^k}(\boldsymbol{x}) - \partial_j L_G(\boldsymbol{x})$. Restricting our attention to 4-graphs, we have

$$\partial_j L_{\mathcal{K}_n^4}(\boldsymbol{x}) = \frac{(1-x_j)^3}{6} - \frac{1-x_j}{2} \sum_{\substack{i=1 \\ i \neq j}}^{n} x_i^2 - \frac{1}{3} \sum_{\substack{i=1 \\ i \neq j}}^{n} x_i^3.$$

Since the stable stationary points of the discrete dynamics (2) are in one-to-one correspondence with characteristic vectors of maximal cliques of G, they can be used as heuristic for the maximum clique problem on k-graphs, giving an approach with complexity $O(m(|V| + |E|))$ where m is the average number of iteration required by the dynamics for convergence.

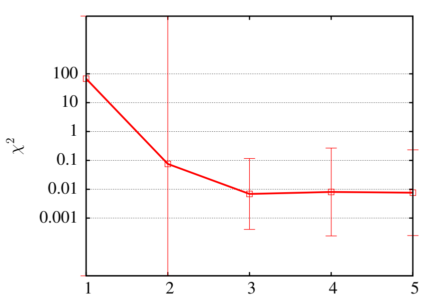


Figure 3. Sensitivity to ϵ

4. Experimental results

In order to assess the precision improvement of our approach over current techniques, we compared our algorithm with the Hough transform approach proposed by Lowe in [5] for clustering pose votes. The experiments were conducted on the fishes database used in [4]. Since Lowe's method is more robust with deformations that preserve the aspect ratio, we evaluated the approaches on both aspect ratio preserving as well as aspect ratio deforming transformations. We run the algorithms on 20 randomly selected images under affine transformations at varying scales, rotations, aspect ratios and translations. We used the SIFT algorithm to extract features both from the original as well as the deformed images. The set of feasible associations, and hence, the vertex set of the auxiliary hypergraph, was generated by keeping the 100 best associations ranked according to their Euclidean distance in the feature space. For Lowe's algorithm we used the modified version of the best-bin-first algorithm as suggested in his paper.

We performed tests on transformations with scaling factors varying from 0.5 to 2, rotation angles from 0.5 to 3 radians, aspect ratios from 0.8 to 1.3 and random translations. Further, to assess the effect of the tolerance ϵ on our approach, we iterated the set of experiments with values of ϵ varying from 1 to 5 pixels. The quality of the affine transformations were evaluated by calculating the χ^2 distance between the true parameters and the estimated ones.

In Figures 2 the performance of the two approaches are compared. Figures 2(a) and 2(b) refer to aspect ratio preserving experiments, and plot the average error at varying scale and rotation respectively. These were calculated averaging over all the values of the other aspect preserving parameters and all the images. Conversely, Figure 2(c) presents the error at varying aspect ratio, averaging this time over all parameters and images. The performance of our approach is roughly one order of magnitude better than that obtained using the Hough transform. This is mainly because our hypergraph-based formulation does not need to define bins, or to select good initial matches, as it is able to capture the largest set of coherent associations in a more accurate way. In Figure 3 we show how the performance of our approach varies with respect to the parameter ϵ. We can see that the algorithm is robust with respect to this parameter, in fact the accuracy obtained by using a tolerance of 3, 4 or 5 pixels does not change much and even with very relaxed constraints it is able to find highly coherent cliques. By the converse, setting very small values of ϵ is not only unnecessary, but also counterproductive: By setting a transformation constraint of 2 pixels the algorithm begins to reduce the number of hy-

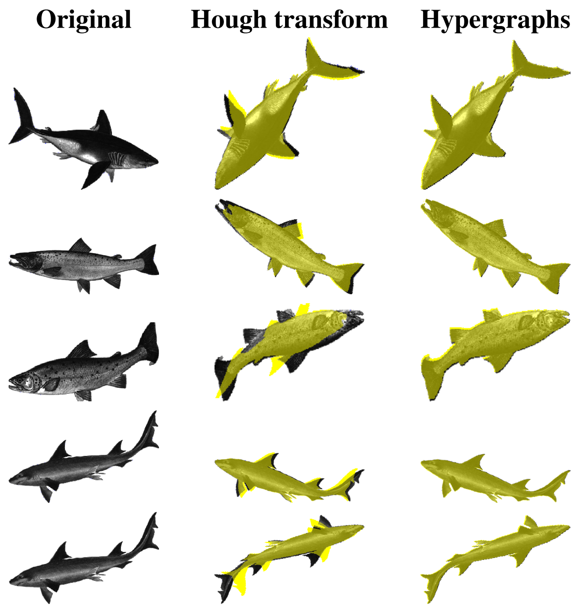


Table 1. Some example results.

peredges in G, yielding smaller cliques and, thus a less reliable estimation of the transformation parameters.

To illustrate how the differences in the χ^2 measure impact the estimation of the affine transformation, we show some examples in Table 1. Each row presents a different example, the first column shows the original fish, the second displays the target fish (black) and the original fish distorted with the transformation estimated using Lowe's approach (yellow). Finally, the last column shows the target fish (black) and the original fish distorted with the transformation estimated using our approach (yellow). From the images we can see that the estimation obtained by our method is indistinguishable from the target image, while the performances obtained using the Hough transform deviates from the target images significantly. Note that in the last two rows the same fish was transformed using respectively an aspect-ratio invariant transformation and an affine transformation that modified the aspect-ratio. As mentioned before, the solution proposed by Lowe is less robust with the latter transformation and, in fact, the error in the estimation of the last transformation is fairly high, while our approach performs very well in both cases.

5. Conclusions

In this paper we presented a new feature-based approach for the estimation of the transformation parameters between affinely distorted images. This is done by casting the problem to the search for a maximum clique in an auxiliary hypergraph, and then turning this into a continuous optimization problem over the standard simplex. Finally, we compared our approach with a state-of-the-art algorithm on a database of affinely distorted fish images. The tests show that our approach outperforms the competition in terms of precision of the estimation.

Acknowledgments. We thank Maria Petrou for kindly providing the data used in this paper.

We acknowledge the financial support of the Future and Emerging Technology (FET) Programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open project SIMBAD grant no. 213250.

References

[1] L. E. Baum and J. A. Eagon. An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. *Bull. Amer. Math. Soc.*, 73:360–363, 1967.

[2] J. Ben-Arie and Z. Wang. Pictorial recognition of objects employing affine invariance in the frequency domain. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 20(6):604–618, 1998.

[3] A. Goshtasby. Registration of images with geometric distortion. *IEEE Trans. Geosc. and Rem. Sens.*, 26(1):60–64, 1998.

[4] A. Kadyrov and M. Petrou. Affine parameter estimation from the trace transform. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 28(10):1631–1645, 2006.

[5] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. of Comp. Vis.*, 20:91–110, 2003.

[6] L. Lucchese. A frequency domain technique based on energy radial projection for robust estimation of global 2d affine transformations. *Comp. Vis. and Image Underst.*, 81:72–116, 2001.

[7] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 27(10):1615–1630, 2005.

[8] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *Int. J. of Comp. Vis.*, 65(1/2):43–72, 2005.

[9] T. S. Motzkin and E. G. Straus. Maxima for graphs and a new proof of a theorem of Turán. *Canad. J. Math.*, 17:533–540, 1965.

[10] S. Rota Bulò and M. Pelillo. A continuous characterization of maximal cliques in k-uniform hypergraphs. In *Proc. Lear. and Intell. Opt.*, 2008. (To appear).

[11] R. Yang and Y. Gao. Line-based affine invariant object location using transformation space decomposition. *Int. Conf. on Patt. Rec.*, 2:646–649, 2006.

[12] Z. Yang and F. S. Cohen. Cross-weighted moments nd affine invariants for image registration and matching. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 21(8):804–814, 1999.