Robust Figure Extraction on Textured Background: a Game-Theoretic Approach

Andrea Albarelli, Emanuele Rodolà, Alberto Cavallarin, and Andrea Torsello
Dipartimento di Informatica - Università Ca' Foscari
via Torino, 155 - 30172 Venice Italy
http://www.dsi.unive.it

Abstract

Feature-based image matching relies on the assumption that the features contained in the model are distinctive enough. When both model and data present a sizeable amount of clutter, the signal-to-noise ratio falls and the detection becomes more challenging. If such clutter exhibits a coherent structure, as it is the case for textured background, matching becomes even harder. In fact, the large amount of repeatable features extracted from the texture dims the strength of the relatively few interesting points of the object itself. In this paper we introduce a game-theoretic approach that allows to distinguish foreground features from background ones. In addition the same technique can be used to deal with the object matching itself. The whole procedure is validated by applying it to a practical scenario and by comparing it with a standard point-pattern matching technique.

1. Introduction

Given its central role in many computer vision tasks, image matching and registration is a widely investigated topic in literature. Several approaches exploit global properties of the images, ranging from the many techniques based on cross-correlation [6] to those that work in the frequency domain [4] or adopt the mutual information as a similarity measure [10]. While successful in many scenarios, the global nature of those techniques makes them little robust to changes in illumination and to the presence of clutter. Feature-based approaches partially solve those problems. Attributed feature points are extracted from images using detectors [8, 9, 7] and descriptors [5, 2] that are locally invariant to illumination, scale and rotation. Usually, the model features are matched with those obtained from the target image by means of some RANSAC-based approach that can exploit the prior given by the descriptors [3]. Critical to the success of this kind of technique is of course the distinctiveness of the extracted features. Unfortunately, when dealing with textured clutter, this distinctiveness comes short and the number of very repeatable but irrelevant features overshadows those coming from the foreground object. To avoid false matches it is mandatory to recognize and ignore the background. In this paper we cope with both the filtering of the background features and the recognition task by tailoring the matching framework introduced in [1]. Specifically we model the filtering step as a self-matching game, where features that show high mutual similarity in the same image are deemed not distinctive enough and thus screened away. By converse, the recognition step is performed as a matching game between the model and a data image, where a set of highly coherent pairs of corresponding features is seeked.

2. The Matching Game

Evolutionary game theory [11] considers an idealized scenario where pairs of individuals are repeatedly drawn at random from a large population to play a two-player game. Each player obtains a payoff that depends only on the strategies played by him and its opponent. Players are not supposed to behave rationally, but rather they act according to a pre-programmed behavior, or mixed strategy. It is supposed that some selection process operates over time on the distribution of behaviors favoring players that receive larger payoffs. More formally, let $O = \{1, \cdots, n\}$ be the set of available strategies (*pure strategies* in the language of game theory) and $C = (c_{ij})$ be a matrix specifying the payoff that an individual playing strategy i receives against someone playing strategy j. A *mixed strategy* is a probability distribution $\mathbf{x} = (x_1, \ldots, x_n)^T$ over the available strategies O.

Being probability distributions, mixed strategies are constrained to lie in the n-dimensional standard simplex

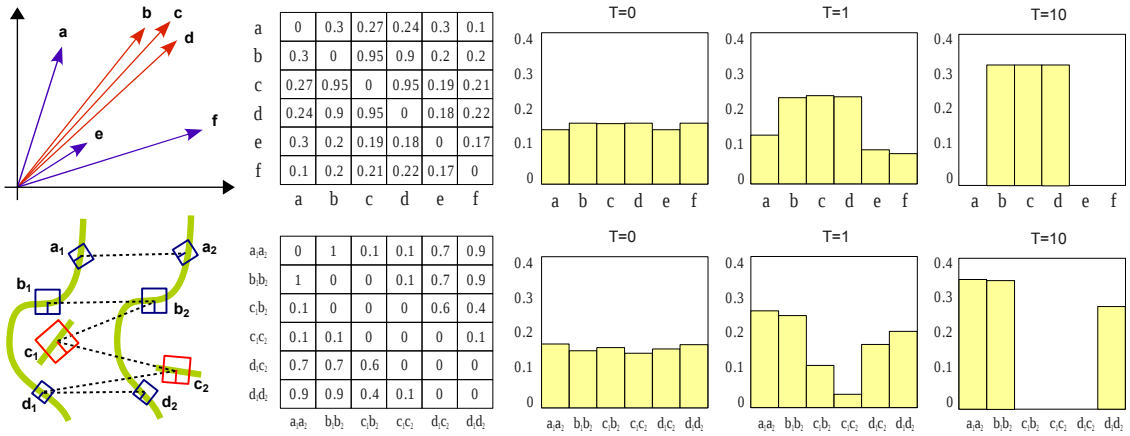


Figure 1. Examples of the two evolutionary matching games proposed

$\Delta^n = \{\mathbf{x} \in \mathbb{R}^n : \forall i \in 1 \dots n\ x_i \geq 0,\ \sum_{i=1}^n x_i = 1\}$. The *support* of a mixed strategy $\mathbf{x} \in \Delta$, denoted by $\sigma(\mathbf{x})$, is defined as the set of elements chosen with non-zero probability: $\sigma(\mathbf{x}) = \{i \in O \mid x_i > 0\}$. The expected payoff received by a player choosing element i when playing against a player adopting a mixed strategy $\mathbf{x}$ is $(C\mathbf{x})_i = \sum_j c_{ij} x_j$, hence the expected payoff received by adopting the mixed strategy $\mathbf{y}$ against $\mathbf{x}$ is $\mathbf{y}^T C\mathbf{x}$. The *best replies* against mixed strategy $\mathbf{x}$ is the set of mixed strategies

$$\beta(\mathbf{x}) = \{\mathbf{y} \in \Delta \mid \mathbf{y}^T C\mathbf{x} = \max_{\mathbf{z}}(\mathbf{z}^T C\mathbf{x})\}\,.$$

A strategy $\mathbf{x}$ is said to be a *Nash equilibrium* if it is the best reply to itself, i.e., $\forall \mathbf{y} \in \Delta$, $\mathbf{x}^T C\mathbf{x} \geq \mathbf{y}^T C\mathbf{x}$. This implies that $\forall i \in \sigma(\mathbf{x})$ we have $(C\mathbf{x})_i = \mathbf{x}^T C\mathbf{x}$; that is, the payoff of every strategy in the support of $\mathbf{x}$ is constant. A strategy $\mathbf{x}$ is said to be an *evolutionary stable strategy* (ESS) if it is a Nash equilibrium and

$$\forall \mathbf{y} \in \Delta \quad \mathbf{x}^T C\mathbf{x} = \mathbf{y}^T C\mathbf{x} \Rightarrow \mathbf{x}^T C\mathbf{y} > \mathbf{y}^T C\mathbf{y}\,.$$

This condition guarantees that any deviation from the stable strategies does not pay. The search for a stable state is performed by simulating the evolution of a natural selection process. Under very loose conditions, any dynamics that respect the payoffs is guaranteed to converge to Nash equilibria [11] and (hopefully) to ESS's; for this reason, the choice of an actual selection process is not crucial and can be driven mostly by considerations of efficiency and simplicity. We chose to use the replicator dynamics, a well-known formalization of the selection process governed by the following equation

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t)\frac{(C\mathbf{x}(t))_i}{\mathbf{x}(t)^T C\mathbf{x}(t)}$$

where $\mathbf{x}_i$ is the i-th element of the population and C the payoff matrix. Once the population has reached a local maximum, all the non-extincted pure strategies (i.e., $\sigma(\mathbf{x})$) can be considered selected by the game.

2.1. Filtering a Textured Background

When dealing with textures, we can expect a large number of features that exhibit very similar descriptors. This is a very unfortunate condition for matching: in fact, this high level of conguence can easily distract any matcher from the foreground object. Paradoxically we use this property to screen out background features. Following [1], we model each feature as a strategy in a matching game where the payoff matrix is defined by:

$$C(i,j) = e^{-\alpha|d_i - d_j|} \tag{1}$$

where d_i and d_j are the descriptor vectors associated to features i and j, and α is a parameter that controls the level of selectivity. Clearly, features that are similar will get a large mutual payoff and thus are more likely to be selected by the evolutive process. A simplified (but numerically correct) example of such evolution is shown in the first row of Fig. 1. Here, six descriptors of dimensionality 2 are labeled from a to f. Vectors b,c and d get high values in the payoff matrix since they are close in the descriptor space. Other descriptors get lower mutual payoffs, according to their respective distances. We start the replicator dynamics ($T = 0$) near the barycenter of Δ^6, which is sligthly perturbed to help avoiding local minima. After just one iteration ($T = 1$), strategies b,c and d get a significant evolutionary boost over the others, and after ten iterations ($T = 10$) they are the only strategies left in the support. We can then classify those features as background and filter them out.

2.2. Matching Model and Data

In order to match model and data points we need to define a slightly different matching game. In this con-

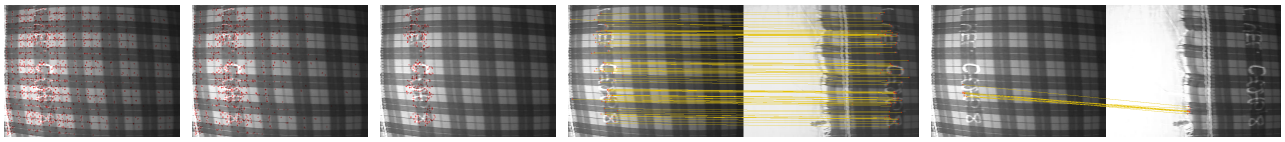
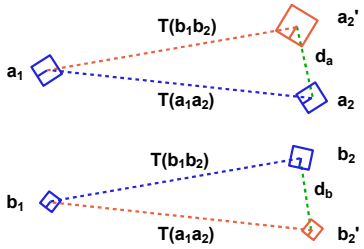


Figure 2. Background filtering and feature matching (best viewed in color)

text, each strategy models a pair of features (a_1, a_2) that belong respectively to the model and the data. We define a payoff among strategies that is proportional to the compatibility of the affine transformation estimated by the descriptor used (for instance, SIFT [5] or SURF [2]). Specifically, we are able to associate to each strategy (a_1, a_2) an affine transformation, which we call $T(a_1, a_2)$.



When this is applied to a_1 it produces the point a_2, but when it is applied to the model point b_1 it will give a point b_2' that is near to b_2 if $T(a_1, a_2)$ is similar to $T(b_1, b_2)$. Given two strategies (a_1, a_2) and (b_1, b_2) and their associated transformations $T(a_1, a_2)$ and $T(b_1, b_2)$ we calculate their reciprocal reprojected virtual points as: $a_2' = T(b_1, b_2)a_1$ and $b_2' = T(a_1, a_2)b_1$. Given virtual points a_2' and b_2' we are finally able to define the payoff between (a_1, a_2) and (b_1, b_2) as:

$$C((a_1, a_2), (b_1, b_2)) = e^{-\beta max(|a_2 - a_2'|, |b_2 - b_2'|)} \quad (2)$$

where β is a selectivity parameter that allows to operate a more or less selective matching game. Clearly, large groups of point pairs that are coherent with respect to an affine transformation will receive a large payoff and thus an evolutive advantage. In the second row of Fig. 1 we show an example of this matching game. Here, coherent strategies exhibit high payoff values (i.e., $C((a_1, a_2), (b_1, b_2)) = 1$), while less compatible pairs get lower scores (i.e., $C((a_1, a_2), (c_1, c_2)) = 0.1$). Note that strategies that share the same model or data point get payoff 0 to avoid one-to-many matching. Initially, the population is set to a slightly perturbed barycenter of Δ^6. After one iteration, (c_1, b_2) and (c_1, c_2) have lost a significant amount of support, while (d_1, c_2) and (d_1, d_2) are still played by a sizeable amount of population, despite being mutually exclusive. After ten iterations, (d_1, d_2) has finally prevailed over (d_1, c_2) and the final support has emerged.

3. Experimental Evaluation

We tested our game-theoretic approach by applying it to the detection of hand-written markers placed on textured fabric. This is a typical scenario for batch tracking in the textile industry, where barcodes or RFID tags are not viable solutions due to the harsh cloth processing conditions that would destroy them. The first three frames of Fig. 2 show the background filtering performance of our method. The first frame contains all the original SIFT features extracted, the second one shows those survived after applying our filter with selectivity parameter $\alpha = 10^{-4}$. By using $\alpha = 10^{-3}$ all the background is screened in the third frame. We observed that a larger value of α does not affect much the result, as foreground features are quite disjointed. The matcher performance has been evaluated by comparing its precision-recall curve with those obtained by using an optimized RANSAC-based technique. Specifically, we implemented a PROSAC [3] variant by using descriptor vectors as hints for the selection of transformation candidates in an affine point-pattern matching. In order to assess the effect of the background elimination step, we applied this RANSAC schema to both filtered and unfiltered frames.

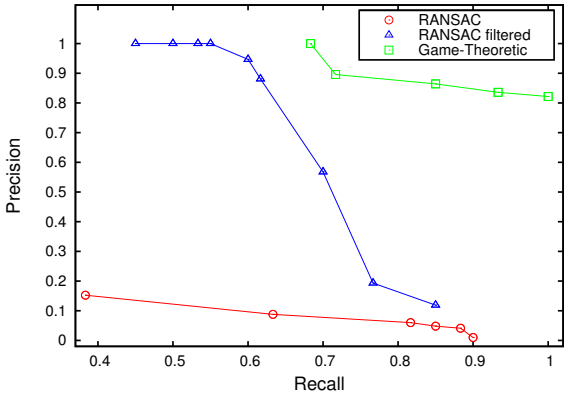


Figure 3. Comparison with RANSAC

The trade-off between precision and recall was adjusted respectively by means of parameter β and by using different thresholds for the consensus. Tests were performed with 20 markers and 15 different fabric patterns. The markers were present in 59 frames of a

30.000 frames long video sequence. Given the constant presence of a textured background, the poor results obtained with RANSAC and the unfiltered video were expected. Indeed, we were unable to reach a full recall without a complete loss of precision, and even when accepting a low recall most of the detected frames were false positives due to background matching. RANSAC performance increases dramatically after application of the filter. Nevertheless, it is not possible to obtain a high level of recall without losing precision. This is due to the presence of features that do not belong to the foreground marker and neither are part of a texture. This happens, for instance, with sewings, seams or dirt present in the fabric. In the right half of Fig. 2 we show an instance where our method obtains the correct match, while RANSAC is distracted by a junction in the fabric. The game-theoretic matcher (applied over filtered frames) obtains by far the best results. In fact, a perfect recall is obtained with a precision value above 0.8 ($\beta = 10^{-3}$) and, by using a more selective parameter ($\beta = 10^{-2}$) all the false positives are avoided while still obtaining a recall just slightly below 0.7. In some practical applications it is more important to guarantee a recall of 1 since a moderate number of false positives can be tolerated (and filtered bottomward in the pipeline), while a miss in the detection is not allowed. To measure the loss in precision with respect to noise, we corrupted both data and model with additive Gaussian noise. At each noise level (expressed with the standard deviation in Fig. 4) we tuned β to maintain a recall of 1 and measured the precision. While it was always possible to obtain a complete recall, we observed a linear decay of the precision. This is not a failure of the matcher itself, but an impaired effectiveness of the background filter due to the reduced similarity among the extracted descriptors. It should be noted, however, that in this experimental setup a precision of 0.3 with a recall of 1 corresponds to a fall-out of 0.006 (about 180 false positives over 30.000 tests).

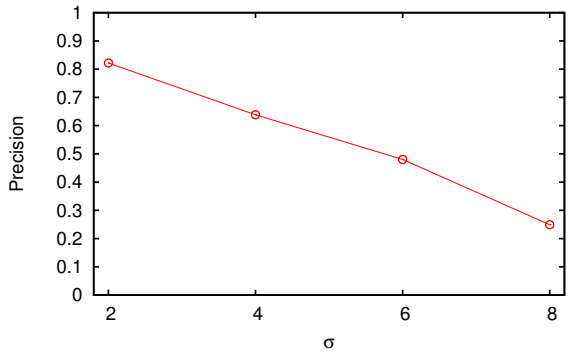


Figure 4. Effect of image noise

4. Conclusions

We presented a game-theoretic approach that allows to perform a robust feature-based matching even when the foreground is absorbed in a highly textured background. This is done by playing two different non-cooperative games: a filter game, that separates foreground from background, and a matching game, that performs the actual point-pattern matching. An experimental validation shows that both the steps concur to the improvement of the whole matching task and the obtained results outperform in terms of precision and recall an optimized RANSAC-based approach.

Acknowledgments

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] A. Albarelli, S. Rota Bulò, A. Torsello, and M. Pelillo. Matching as a non-cooperative game. In *ICCV 2009: Proceedings of the 2009 IEEE International Conference on Computer Vision*. IEEE Computer Society, 2009.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. J. V. Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[3] O. Chum and J. Matas. Matching with prosac - progressive sample consensus. In *CVPR 05: Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR05) - Volume 1*, pages 220–226, Washington, DC, USA, 2005. IEEE Computer Society.

[4] E. De Castro and C. Morandi. Registration of translated and rotated images using finite fourier transforms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(5):700–703, 1987.

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

[6] W. K. Pratt. *Digital image processing (2nd ed.)*. John Wiley & Sons, Inc., New York, NY, USA, 1991.

[7] E. Rosten, R. Porter, and T. Drummond. Faster and better: a machine learning approach to corner detection. *CoRR*, abs/0810.2434, 2008.

[8] J. Shi and C. Tomasi. Good features to track. In *1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pages 593 – 600, 1994.

[9] S. M. Smith and J. M. Brady. Susan—a new approach to low level image processing. *Int. J. Comput. Vision*, 23(1):45–78, 1997.

[10] P. Viola and W. M. Wells, III. Alignment by maximization of mutual information. *Int. J. Comput. Vision*, 24(2):137–154, 1997.

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