Robust Game-Theoretic Inlier Selection for Bundle Adjustment

Andrea Albarelli, Emanuele Rodolà and Andrea Torsello
Dipartimento di Informatica - Università Ca' Foscari
via Torino, 155 - 30172 Venice Italy
albarelli@unive.it rodola@dsi.unive.it torsello@dsi.unive.it

Abstract

Bundle Adjustment is a widely adopted self-calibration technique that allows to estimate scene structure and camera parameters at once. Typically this happens by iteratively minimizing the reprojection error between a set of 2D stereo correspondences and their predicted 3D positions. This optimization is almost invariantly carried out by means of the Levenberg-Marquardt algorithm, which is very sensitive to the presence of outliers in the input data. For this reason many structure-from-motion techniques adopt some inlier selection algorithm. This usually happens both in the initial feature matching step and by pruning matches with larger reprojection error after an initial optimization. While this works well in many scenarios, outliers that are not filtered before the optimization can still lead to wrong parameter estimation or even prevent convergence. In this paper we introduce a novel stereo correspondences selection schema that exploits Game Theory in order to perform a robust inlier selection before any optimization step. The practical effectiveness of the proposed approach is confirmed by an extensive set of experiments and comparisons with state-of-the-art techniques.

1. Introduction

The selection of 2D point correspondences is arguably the most important step in image based multi-view reconstruction. As a matter of fact, differently from techniques augmented by structured light or known markers, wrong initial correspondences can lead to sub-optimal parameter estimation or, in the worst cases, to the inability of the optimization algorithm to obtain a feasible solution. For this reason reconstruction approaches adopt several specially crafted expedients to avoid as much as possible the inclusion of outliers. In the first place correspondences are not searched throughout all the image plane, but only points that are both repeatable and well characterized are considered. This selection is carried out by means of interest point detectors and feature descriptors. Salient points are localized with sub-pixel accuracy by general detectors, such as Harris Operator [2] and Difference of Gaussians [7], or by using techniques that are able to locate affine invariant regions, such as Maximally Stable Extremal Regions (MSER) [8] and Hessian-Affine [9]. The affine invariance property is desirable since the change in appearance of a scene region after a small camera motion can be locally approximated with an affine transformation. Once interesting points are found, they must be matched to form the candidate pairs to be fed to the bundle adjustment algorithm. Most of the currently used techniques for point matching are based on the computation of some affine invariant feature descriptor. Specifically, to each point is assigned a descriptor vector with tens to hundreds of dimensions, a scale and a rotation value. Among the most used feature descriptor algorithms are the Scale-Invariant Feature Transform (SIFT) [6, 5], the Speeded Up Robust Features (SURF) [3], the Gradient Location and Orientation Histogram (GLOH) [10] and more recently the Local Energy based Shape Histogram (LESH) [11]. In all of these techniques, the descriptor vector itself is robust with respect to affine trasformations: i.e., similar image regions exhibit descriptor vectors with small mutual Euclidean distance. This property is used to match each point with the candidate that is associated to the nearest descriptor vector. If the descriptor is not distinctive enough this approach is prone to select many outliers. A common optimization involves the definition of a maximum threshold over the distance ratio between the first and the second nearest neighbors. In addition, points that are matched multiple times are deemed as ambiguous and discarded (i.e., one-to-one matching is enforced). Another common heuristic for the elimination of erroneous matches is to exclude points that exhibit a large reprojection error after a first round of Levenberg-Marquardt optimization [4] (see for instance [14]). Unfortunately this afterthought is based upon an error estimation that depends on the point pairs chosen beforehand; this leads to a quandary that can only be solved by avoiding wrong matches from the start. In this paper we introduce a robust matching technique that allows to operate a very accurate inlier selection at an early stage of the process and without any need to rely on 3D reprojection. In the experimental section, to assess the advantages of our approach, we present a comprehensive set of comparisons between the results delivered by our technique and those

obtained with a reference implementation of the structure-from-motion system presented in [13] and [14].

2. Game-Theoretic Point Pairs Selection

The selection of matching points on behalf of the feature descriptor is only able to exploit local information. This limitation conflicts with the richness of information that is embedded in the scene structure. For instance, under the assumption of rigidity and small camera motion, intuition suggests that features that are close in one view cannot be too far apart in the other one. Further, if a pair of features exhibit a certain difference of angles or ratio of scales, this relation should be maintained among their respective matches. Our basic idea is to formalize this intuitive notion of consistency between pairs of feature matches into a real-valued utility function and to find a large set of matches that express a high level of mutual compatibility. Of course, the ability to define a meaningful pairwise utility function and a reliable technique for finding a consistent set as large as possible is paramount for the effectiveness of the approach. Following [15, 1], we model the matching process in a game-theoretic framework, where two players extracted from a large population select a pair of matching points from two images. The player then receives a payoff from the other players proportional to how compatible his match is with respect to the other player's choice, where the compatibility derives from some utility function that rewards pairs of matches that are consistent. In Section 2.2 such a function will be proposed, but in practice many different choices can be made: for instance it is possible to assign a high payoff to pairs of matches that preserve the distance between source and destination points and a low payoff otherwise. Clearly, it is in each player's interest to pick matches that are compatible with those the other players are likely to choose. In general, as the game is repeated, players will adapt their behavior to prefer matchings that yield larger payoffs, driving all inconsistent hypotheses to extinction, and settling for an equilibrium where the pool of matches from which the players are still actively selecting their associations forms a cohesive set with high mutual support. Within this formulation, the solutions of the matching problem correspond to evolutionary stable states (ESS's), a robust population-based generalization of the notion of a Nash equilibrium. In a sense, this matching process can be seen as a contextual voting system, where each time the game is repeated the previous selections of the other players affect the future vote of each player in an attempt to reach consensus. This way the evolving context brings global information into the selection process.

2.1. Non-cooperative Games

Originated in the early 40's, Game Theory was an attempt to formalize a system characterized by the actions of entities with competing objectives, which is thus hard to characterize with a single objective function [16]. According to this view, the emphasis shifts from the search of a local optimum to the definition of equilibria between opposing forces. In this setting multiple players have at their disposal a set of strategies and their goal is to maximize a payoff that depends also on the strategies adopted by other players. Evolutionary game theory originated in the early 70's as an attempt to apply the principles and tools of game theory to biological contexts. Evolutionary game theory considers an idealized scenario where pairs of individuals are repeatedly drawn at random from a large population to play a two-player game. In contrast to traditional game-theoretic models, 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 higher 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. Clearly, mixed strategies are constrained to lie in the n-dimensional standard simplex

$$\Delta^n = \left\{ \mathbf{x} \in \mathbb{R}^n \ : \ x_i \geq 0 \text{ for all } i \in 1 \ldots n, \ \sum_{i=1}^{n} x_i = 1 \right\} .$$

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

$$\forall \mathbf{y} \in \Delta \quad \mathbf{x}^T C \mathbf{x} = \mathbf{y}^T C \mathbf{x} \Rightarrow \mathbf{x}^T C \mathbf{y} > \mathbf{y}^T C \mathbf{y} . \qquad (1)$$

This condition guarantees that any deviation from the stable strategies does not pay.

2.2. Matching Strategies and Payoffs

Central to this framework is the definition of a *matching game*, which implies the definition of the strategies avail-

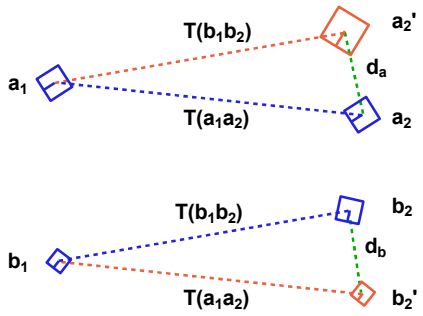


Figure 1. The payoff between two matching strategies is inversely proportional to the maximum reprojection error obtained by applying the affine transformation estimated by a match to the other.

able to the players and of the payoffs related to these strategies. Given a set M of feature points in a source image and a set D of potentially corresponding features in a destination image, we call a *matching strategy* any pair (a_1, a_2) with $a_1 \in M$ and $a_2 \in D$. We call the set of all the matching strategies S. In principle, all the features extracted by an interest point detector could be used to build the matching strategies set, thus leading to a size of the set S that grows quadratically with the average number of features detected in an image. In practice, however, in Section 2.3 we adopt some heuristics that allow us to obtain good overall results with a much smaller set. Once S has been selected, our goal becomes to extract from it the largest subset that includes only correctly matched points: that is, strategies that associate a feature in the source image with the same feature in the destination image. To this extent, it is necessary to define a payoff function $\Pi : S \times S \rightarrow \mathbb{R}^+$ that exploits some pairwise information available at this early stage (i.e. before estimating camera and scene parameters). Since scale and rotation are associated to each feature, it seems natural to try to use this information to enforce coherence between matching strategies. Specifically, we are able to associate to each matching strategy (a_1, a_2) one and only one similarity transformation, that we call $T(a_1, a_2)$. When this transformation is applied to a_1 it produces the point a_2, but when applied to the source point b_1 of the matching strategy (b_1, b_2) it does not need to produce b_2. In fact it will produce b_2 if and only if $T(a_1, a_2) = T(b_1, b_2)$, otherwise it will give a point b'_2 that is as near to b_2 as the transformation $T(a_1, a_2)$ is similar to $T(b_1, b_2)$. Given two matching strategies (a_1, a_2) and (b_1, b_2) and their respective associated similarities $T(a_1, a_2)$ and $T(b_1, b_2)$, we calculate their reciprocal reprojected points as:

$$\begin{aligned} a'_2 &= T(b_1, b_2)a_1 \\ b'_2 &= T(a_1, a_2)b_1 \end{aligned}$$

That is, the virtual points obtained by applying to each source point the similarity transformation associated to the other match (see Fig. 1). Thus, given virtual points a'_2 and b'_2, the payoff between (a_1, a_2) and (b_1, b_2) is:

$$\Pi((a_1, a_2), (b_1, b_2)) = e^{-\lambda max(|a_2 - a'_2|, |b_2 - b'_2|)} \quad (2)$$

where λ is a selectivity parameter that allows to operate a more or less strict inlier selection. If λ is small, then the payoff function (and thus the matching) is more tolerant, otherwise the evolutionary process becomes more selective as λ grows. We define 2 as a *similarity enforcing payoff function* and we call a *matching game* any symmetric non-cooperative game that involves a matching strategies set S and a similarity enforcing payoff function Π.

The rationale of the payoff function proposed in equation 2 is that, while by changing point of view the similarity relationship between features is not maintained (as the object is not planar and the transformation is projective), we can expect the transformation to be a similarity at least "locally". This means that we aim to extract clusters of feature matches that belong to the same region of the object and that tend to lie in the same level of depth. While this could seem to be an unsound assumption for general camera motion, in the experimental section we will show that it holds well with the typical disparity found in standard multiple view and stereo data sets. Further it should be noted that with large camera motion most, if not all, commonly used feature detectors fail, thus any inlier selection attempt becomes meaningless. One final note should be made about one-to-one matching. Since each source feature can correspond with at most one destination point, it is desirable to avoid any kind of multiple match. It is easy to show that a pair of strategies with zero mutual payoff cannot belong to the support of an ESS (see [1]), thus any payoff function Π can be easily adapted to enforce one-to-one matching by defining:

$$\Pi' = \begin{cases} \Pi((a_1, a_2), (b_1, b_2)) & \text{if } a_1 \neq b_1 \text{ and } a_2 \neq b_2 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

We can define 3 as a *one-to-one similarity enforcing payoff function*.

2.3. Building the Matching Strategies Set

From a theoretical point of view the total number of matching strategies can be as large as the Cartesian product of the sets of features detected in the images. Since most interest point detectors extract thousands of features from an image and the size of the payoff matrix grows quadratically with the number of matching strategies, this leads to problems too large to be managed in an efficient way. While the feature descriptor has not been used to define the payoff function Π, it could be useful to reduce the number of matching strategies considered. Specifically, for each source feature we can generate k matching strategies that connect it to the k destination features that are the nearest

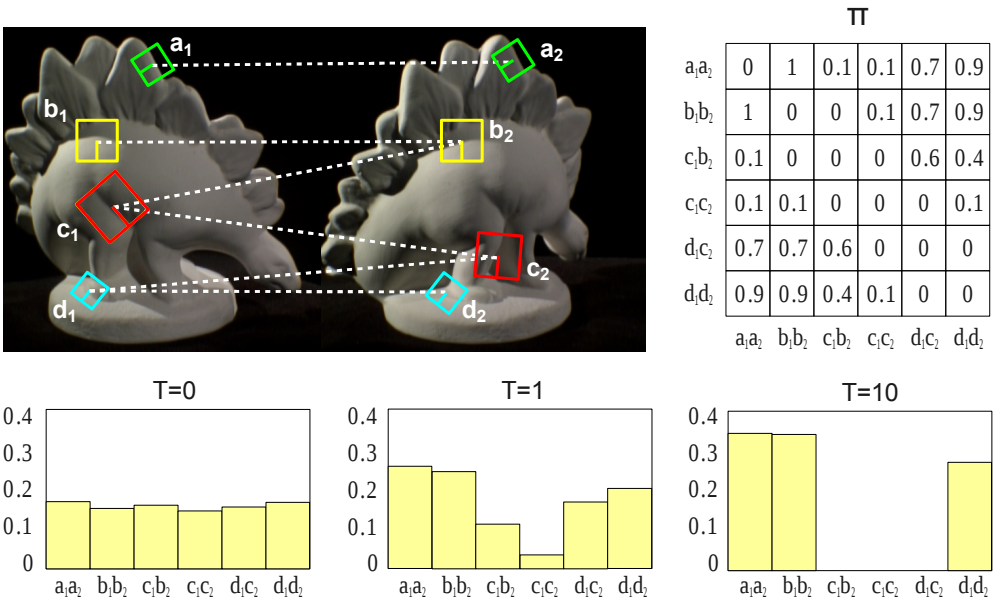


Figure 2. An example of the evolutionary process. Four feature points are extracted from two images and a total of six matching strategies are selected as initial hypotheses. The matrix Π shows the compatibilities between pairs of matching strategies according to a one-to-one similarity-enforcing payoff function. Each matching strategy got zero payoff with itself and with strategies that share the same source or destination point (i.e., $\Pi((b_1,b_2),(c_1,b_2)) = 0$). Strategies that are coherent with respect to similarity transformation exhibit high payoff values (i.e., $\Pi((a_1,a_2),(b_1,b_2)) = 1$ and $\pi((a_1,a_2),(d_1,d_2)) = 0.9$)), while less compatible pairs get lower scores (i.e., $\pi((a_1,a_2),(c_1,c_2)) = 0.1$). Initially (at T=0) the population is set to the barycenter of the simplex and slightly perturbed. After just 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 sizable amount of population. After ten iterations (T=10) (d_1,d_2) has finally prevailed over (d_1,c_2) (note that the two are mutually exclusive). Note that in the final population $((a_1,a_2),(b_1,b_2))$ have a higher support than (d_1,d_2) since they are a little more coherent with respect to similarity.

in terms of descriptor distance. Since our game-theoretic approach operates inlier selection regardless of the descriptor, we do not need to set any threshold with respect to the absolute descriptor distance or the distinctiveness between the first and the second nearest point. In this sense, the only constraint that we need to impose over k is that it should be high enough to allow the correct correspondence to be among the candidates a significative percentage of the times. In the experimental section we will analyze the influence of k over the quality of the matches obtained.

2.4. Evolving to an Optimal Solution

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 [16] 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

$$\mathbf{x}_i(t+1) = x_i(t)\frac{(C\mathbf{x}(t))_i}{\mathbf{x}(t)^T C\mathbf{x}(t)} \qquad (4)$$

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 mating strategies can be considered valid (see Fig. 2). In practice strategies are extincted only after an infinite number of iterations. Since we halt the evolution when the population ceases to change significantly, it is necessary to introduce some criteria to distinguish correct from non-correct matches. To avoid a hard threshold we chose to keep as valid all the strategies played by a population amount exceeding a percentage of the most popular strategy. We call this percentage *quality threshold*. As mentioned in Section 2.2, each evolution process selects a group of matching strategies that are coherent with respect to a local similarity transformation. This means that if we want to cover a large portion of the subject we need to iterate many times and prune the previously selected matches at each new start. Obviously, after all the depth levels have been swept, small and not significant residual groups start to emerge from the evolution. To avoid the selection of these spurious matches we fixed a minimum cardinality for each valid group. We call this cardinality *group size*.

3. Experimental Results

We conducted different sets of experiments. Our first goal was to analyze the impact of the algorithm parameters, namely λ, k, *quality threshold* and *group size*, over the quality of the results obtained. For this purpose we used a pair of

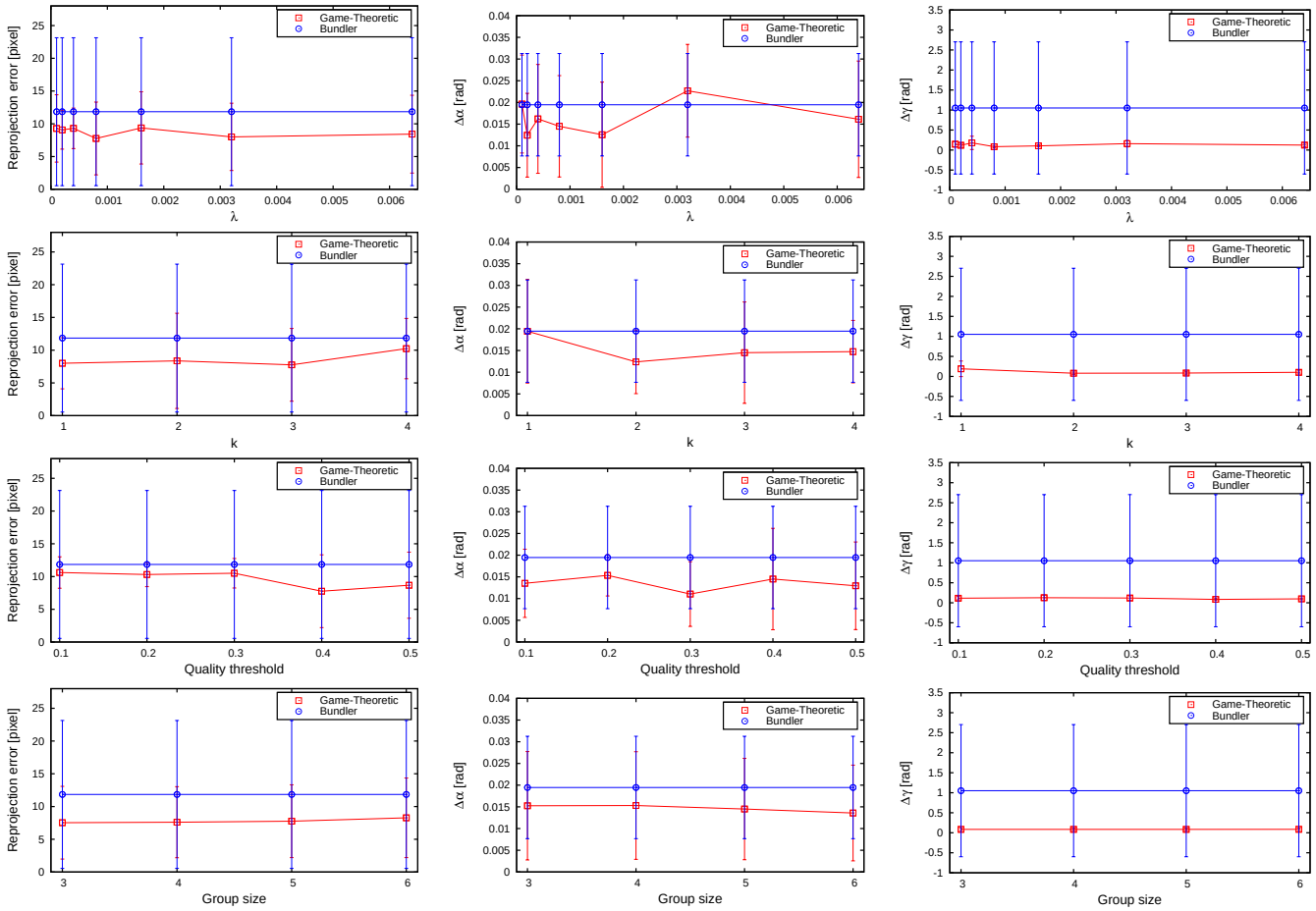
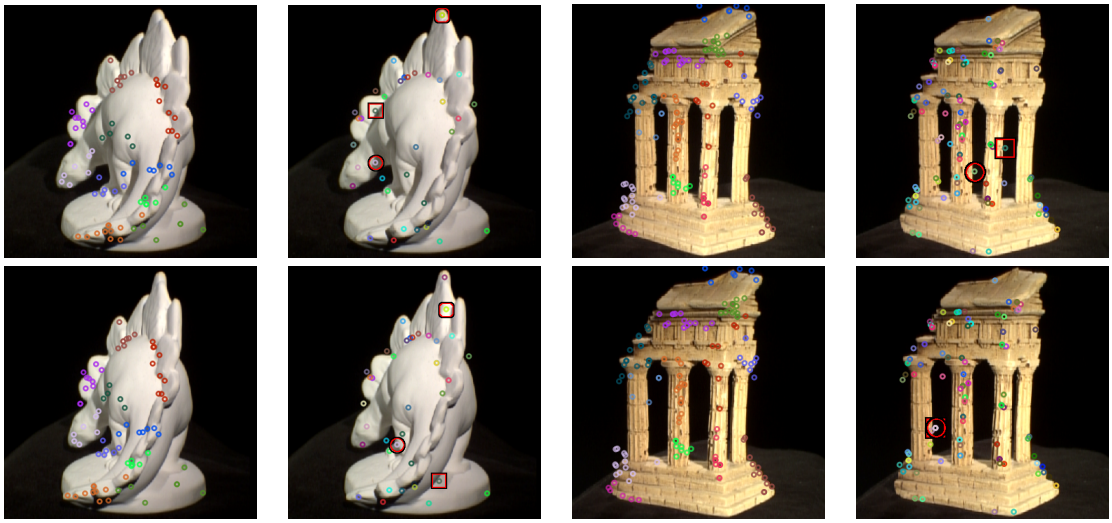


Figure 3. Analysis of the performance of the approach with respect to variation of the parameters of the algorithm.

cameras previously calibrated through a standard procedure and took stereo pictures of 20 different, isolated objects; then, we investigated the influence of the four parameters separately. For each test we evaluated three quality measures: the average reprojection error in pixels (ϵ) and the differences in radians between the (calibrated) ground-truth and respectively the estimated rotation angle ($\Delta\alpha$) and rotation axis ($\Delta\gamma$). In addition, each stereo pair was processed with the keymatcher included in the structure-from-motion suite Bundler [13, 14]. Finally, the correspondences produced by both the Bundler keymatcher and our technique were given as an input to the bundle adjustment procedure included in the suite. This allows to obtain a fair comparison of the two approaches, whose quality parameters can be directly compared, being the result of running the same optimizer on different inputs. In Fig. 3 we reported the results of these experiments. The first row shows the effect of the selectivity parameter λ. As expected both a too low and a too high value lead to less satisfactory results, mainly with respect to the estimation of the angle between the two cameras. This is probably due respectively to a too tight and a too relaxed enforcement of local coherence. The three rows below show the impact of the number of candidate matches for each source point, the quality threshold that a match must exceed to be considered feasible and the minimum size of a valid group. Overall, these experiments suggest that those parameters have little influence over the quality of the result, notwithstanding the Game-Theoretic approach achieves better results in nearly every case.

For the purpose of exploring further the differences between our technique and the Bundler keymatcher, we investigated in depth four cases. We will describe them here in two separate sets. The first set of unordered images comes from the "DinoRing" and "TempleRing" sequences from the Middlebury Multi-View Stereo dataset [12]; for these models, camera parameters are provided and used as a ground-truth. The second set is composed of two calibrated stereo scenes selected from the previously acquired collection, specifically a statue of Ganesha and a handful of screws placed on a table. It should be noted however that Bundler did not find a feasible matching for many stereo pairs in the collection. Again, for all the sets of experiments we evaluated both the rotation error of all the cameras in terms of angular distance and axis discrepancy, and



Dino sequence **Temple sequence**

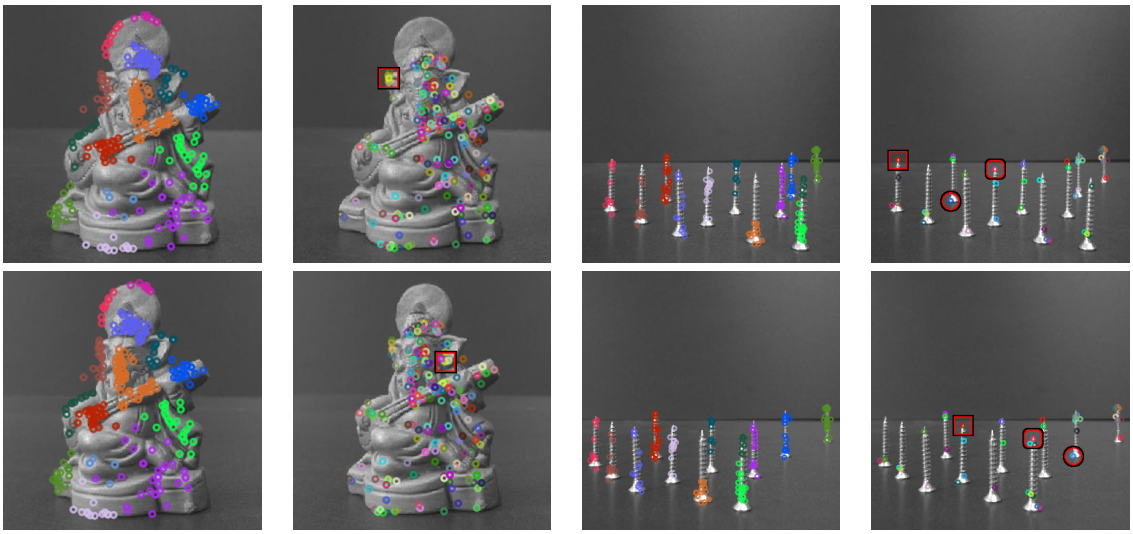
		Game-Theoretic	Bundler Keymatcher	Game-Theoretic	Bundler Keymatcher
Matches		14573	9245	25785	22317
ϵ	≤ 1 pix	24.83	6.49406	22.6049	24.6729
	≤ 5 pix	54.94	48.3659	62.7737	61.8957
	≥ 5 pix	20.21	45.1401	14.6214	13.4314
	Avg.	2.3086	4.5255	2.3577	2.3732
$\Delta\alpha$	Avg.	0.005751	0.005561	0.010514	0.009376
	S. dev.	0.003242	0.003184	0.005282	0.004646
	Max	0.012057	0.011475	0.021527	0.017016
$\Delta\gamma$	Avg.	0.008313	0.009561	0.014050	0.014079
	S. dev.	0.002948	0.006738	0.000511	0.000825
	Max	0.013449	0.030661	0.014692	0.015442
Avg. levels		8.42	-	9.27	-

Figure 4. Results obtained with two multiple view data sets (image best viewed in color).

the reprojection error of the detected keypoints. The average number of matching groups is also given for the Game-Theoretic method.

The "Dino" model is a difficult case in general, as it embodies very few features; the upper part of Fig. 4 shows the correspondences produced by our method (left column) in comparison with the other matcher (right column). A set of optimal parameters detected in the previous experiments was used for configuring our matcher. This resulted as expected in the detection of many correct matches organized in groups, each corresponding to a different level of depth, and visualized with a unique color in the figure. As can be seen, different levels of depth are properly estimated; this is particularly evident throughout the arched back going from the tail (in foreground) to the head of the model (in background), where clustered sets of keypoints follow one after the other. Furthermore, these sets of interest points maintain the right correspondences within the pair of images. The Bundler keymatcher on the other hand, while still achieving good results in the whole process, also outputs erroneous correspondences (marked in the figure).

The quality of reconstruction following the application of both methods can be visually compared by looking at the distribution of the reprojection error in the left half of Fig. 6. While most reprojections fall within 1-3 pixels of distance for the Game-Theoretic approach, the Bundler keymatcher exhibits a long-tail trend, reaching an error spread of 20 pixels. Differently from "Dino", the "Temple" model is quite rich of features; for visualization purposes we only show a subset of the detected matches for both the techniques. While the effectiveness of our approach is not negatively impacted by the model characteristics, mismatches are revealed with Bundler. In particular, the symmetric parts of the object (mainly represented by the pillars) result in very similar features and this causes the matcher to establish one-to-many pairings over them. However, it should be noted that for both the "Dino" and "Temple" models the two matchers deliver comparably good results when fed with a



Ganesha stereo **Screws stereo**

		Game-Theoretic	Bundler Keymatcher	Game-Theoretic	Bundler Keymatcher
Matches		280	200	211	46
ϵ	≤ 1 pix	98.2824	20	0	0
	≤ 5 pix	1.7175	80	34.7716	6.75676
	≥ 5 pix	0	0	65.2284	93.2432
	Avg.	0.321248	1.67583	5.86237	10.2208
$\Delta\alpha$		0.001014	0.007424	0.020822	0.030995
$\Delta\gamma$		0.048076	0.078715	0.106485	0.117885
Levels		14	-	12	-

Figure 5. Results obtained with two stereo view data sets (image best viewed in color).

whole set of views of the object.

In the calibrated stereo scenario, "Ganesha stereo" images are rich of distinctive features and should pose no difficulty to any of the methods. The Bundler keymatcher provides very good results, with only one evident false match out of a total of 200 matches (see Fig. 5). The resulting bundle adjustment is quite accurate, giving very small rotation errors and reprojection distances. Nevertheless, our method performs considerably better: reprojection errors dramatically decrease, with around 98 percent of the keypoints falling below one pixel of reprojection distance.

The second calibrated stereo scene, "Screws stereo", is an emblematic case and provides some meaningful insight. The images depict a dozen of screws standing on a table, placed by hand at different levels of depth. This configuration, together with the abundance of features in the objects themselves, should provide enough information for the two algorithms to extract significant matches. Indeed, the scene proves to be a difficult one due to the very nature of the objects depicted, which are all identical and highly symmetric, and diverse false matches are established by the Bundler keymatcher (see the last column of Fig. 5). This matching results nevertheless in a good estimation of the rigid transformation linking the two cameras, since erroneous pairings are removed *a posteriori* during the subsequent phases of bundle adjustment. By contrast, the Game-Theoretic approach outputs large and accurate sets of matches, roughly one per object, each corresponding to a level of depth; even moderately difficult cases, such as the left-right "swaps" due to the change of viewpoint taking place at the borders, are correctly dealt with. Again, a histogram of the reprojection error for this object is shown in Fig. 6.

Execution times for the matching steps of our technique are plotted in Fig. 7; the scatter plot shows a substantially linear growth of convergence time as the number of matching strategies increases, staying below half a second even with a large number of players.

4. Conclusions

In this paper we introduced a novel game-theoretic technique that performs an accurate feature matching between multiple views of the same subject as a preliminary step for bundle adjustment. Differently from other approaches, we do not rely on a first estimation of scene and camera param-

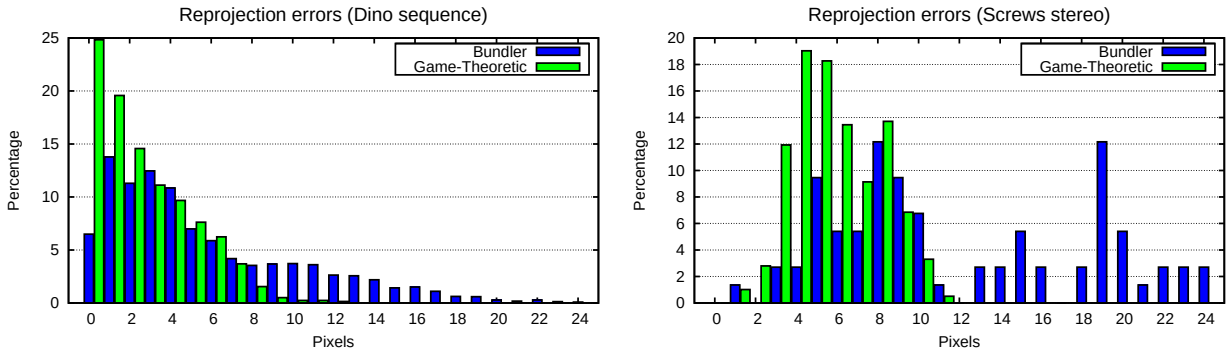


Figure 6. Distribution of the reprojection error on one multiple view (left) and one stereo pair (right) example.

eters in order to obtain a robust inlier selection. Rather, we enforce local compatibility of groups of features with respect to a common similarity transformation. By extracting one group at a time by means of an evolutive process, we are able to cover the entire subject. Experimental comparisons with a widely used technique show the ability of our approach to obtain a tighter inlier selection and thus a more accurate estimation of the scene parameters.

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, 3

[2] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Fourth Alvey Vision Conference*, pages 147–151, 1988. 1

[3] 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. 1

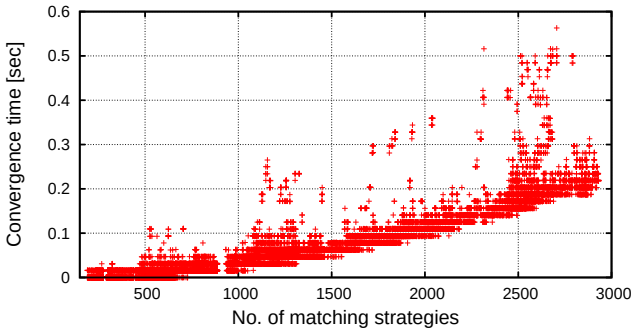


Figure 7. Plot of the convergence time of the replicator dynamics with respect to the number of matching strategies.

[4] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly Journal of Applied Mathmatics*, II(2):164–168, 1944. 1

[5] D. Lowe. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, volume 20, pages 91–110, 2003. 1

[6] D. G. Lowe. Object recognition from local scale-invariant features. In *Proc. of the International Conference on Computer Vision ICCV, Corfu*, pages 1150–1157, 1999. 1

[7] D. Marr and E. Hildreth. Theory of Edge Detection. *Royal Soc. of London Proc. Series B*, 207:187–217, Feb. 1980. 1

[8] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004. British Machine Vision Computing 2002. 1

[9] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I*, pages 128–142, London, UK, 2002. Springer-Verlag. 1

[10] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(10):1615–1630, 2005. 1

[11] M. S. Sarfraz and O. Hellwich. Head pose estimation in face recognition across pose scenarios. In *VISAPP (1)*, pages 235–242, 2008. 1

[12] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR '06*, pages 519–528, Washington, USA, 2006. IEEE Computer Society. 5

[13] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3d. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, pages 835–846, New York, NY, USA, 2006. ACM. 2, 5

[14] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *Int. J. Comput. Vision*, 80(2):189–210, 2008. 1, 2, 5

[15] A. Torsello, S. Rota Bulò, and M. Pelillo. Grouping with asymmetric affinities: A game-theoretic perspective. In *CVPR '06*, pages 292–299, Washington, USA, 2006. IEEE Computer Society. 2

[16] J. Weibull. *Evolutionary Game Theory*. MIT P., 1995. 2, 4