A Game-Theoretic Approach to the Enforcement of Global Consistency in Multi-View Feature Matching

Abstract. In this paper we introduce a robust matching technique that allows to operate a very accurate selection of corresponding feature points from multiple views. Robustness is achieved by enforcing global geometric consistency at an early stage of the matching process, without the need of ex-post verification through reprojection. Two forms of global consistency are proposed, but in both cases they are reduced to pairwise compatibilities making use of the size and orientation information provided by common feature descriptors. Then a game-theoretic approach is used to select a maximally consistent set of candidate matches, where highly compatible matches are enforced while incompatible correspondences are driven to extinction. The effectiveness of the approach in estimating camera parameters for bundle adjustment is assessed and compared with state-of-the-art techniques.

1 Introduction

The selection of 3D point correspondences from their 2D projections is arguably one of the most important steps in image based multi-view reconstruction, as errors in the initial correspondences can lead to sub-optimal parameter estimation. The selection of corresponding points is usually carried out by means of interest point detectors and feature descriptors. Salient points are localized with subpixel accuracy by general detectors, such as Harris Operator [2] and Difference of Gaussians [6], or by using techniques that are able to locate affine invariant regions, such as Maximally stable extremal regions (MSER) [7] and Hessian-Affine [8]. This latter 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 salient and well-identifiable points are found on each image, correspondences between the features in the various views must be extracted and fed to the bundle adjustment algorithm. To this end, each point is associated a descriptor vector with tens to hundreds of dimensions, which usually include a scale and a rotation value. Arguably the most famous of such descriptors are the Scale-invariant feature transform (SIFT) [4], the Speeded Up Robust Features (SURF) [3], and the Gradient Location and Orientation Histogram (GLOH) [9], and more recently the Local Energy based Shape Histogram (LESH) [10]. Features are designed so that similar image regions subject to similarity transformation exhibit descriptor vectors with small Euclidean distance. This property is used to match each point with a candidate with similar descriptor. However, if the descriptor is not distinctive enough this approach is prone to select many outliers since it only exploits local information.

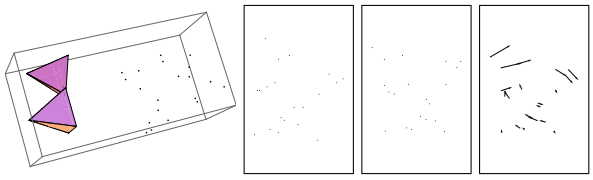


Fig. 1. Locally uniform 3D motion does not result in a locally uniform 2D motion. From left to right: 3D scene, left and right views, and motion estimation.

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, features that are close in one view are expected to be close in the other one as well. In addition, if a pair of features exhibit a certain difference of angles or ratio of scales, this relation should be maintained among their respective matches. This prior information about scene structure can be accounted for by using a feature tracker [5, 12] to extract correspondences, but this requires that the view positions be not far apart. Further, in the presence of strong parallax, a locally uniform 3D motion does not result in a locally uniform 2D motion, and for these reasons the geometric constraints can be enforced only locally (see Fig. 1 for an example). A common heuristic for the enforcement of global structure is to eliminate points that exhibit a large reprojection error after a first round of Bundle Adjustment [13]. Unfortunately this post-filtering technique requires good initial estimates to begin with.

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 reprojections. The approach selects feasible matches by enforcing global geometric consistency. Two geometric consistency models are presented. The first enforces that all pairs of correspondences between 2D views are consistent with a common 3D rigid transformation. Here, as is common in similar point-matching approaches, we assume that we have reasonable guesses for the intrinsic camera parameters and reduce the problem space to the search of a 3D rigid transformation from one image space to the other. This condition is in general underspecified, as a whole manifold of pairs of correspondences are consistent with a rigid 3D transformation. However, by accumulating mutual support through a large set of mutually compatible correspondences one can expect to reduce the ambiguity to a single 3D rigid transformation. In the proposed approach, high order consistency constraints are reduced to a second order compatibility where sets of 2D point correspondences that can be interpreted as projections of rigidly-transformed 3D points all have high mutual support. The reduction is obtained by making use of the scale and orientation information linked with each feature point in the SIFT descriptor [4] and a further reprojection that can be considered a continuous form of hypergraph clique expansion [15].

The second geometric consistency constraint assumes a weak perspective camera and matches together points whose maps are compatible with a common affine transformation. This allows us to extract small coherent clusters of points

all laying at similar depths. The locally affine hypothesis could seem to be an unsound assumption for general camera motion, and in effect cannot account for point inversion due to parallax, but in the experimental section we will show that it holds well with the typical disparity found in standard 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.

Once the geometric consistency contraints are specified, we can use them to drive the matching process. Following [14, 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 pair of matches that are consistent. 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 Pairwise Geometric Consistency

In what follows we will describe the two geometric constraints that will be used to drive the matching process. The first approach tries to impose that the points be consistent with a common 3D rigid transformation.

There are two fundamental hypotheses underlying the reduction to second order of this high-order 3D geometric consistency. First, we assume that the views have the same set of camera parameters, that we have reasonable guesses for the intrinsic parameters, and we can ignore lens distortion. Thus, the geometric consistency is reduced to the compatibility of the projected points with a single 3D rigid transformation related to the relative positions of the cameras. Second, we assume that the feature descriptor provides scale and orientation information and that this is related to actual local information in the 3D objects present in the scene. The effect of the first assumption is that the geometric consistency is reduced to a rigidity constraint that can be cast as a conservation along views of the distances between the unknown 3D position of the feature points, while the effect of the second assumption is that we can recover the missing depth information as a variation in scale between two views of the same point and that this variation is inversely proportional to variation in projected size of the

local patch around the 3D point and, thus, to the projected size of the feature descriptor.

More formally, assume that we have two points p_1 and p_2, which in one view have coordinates (u_1^1, v_1^1) and (u_2^1, v_2^1) respectively, while in a second image they have coordinates (u_1^2, v_1^2) and (u_2^2, v_2^2). These points, in the coordinate system of the first camera, have 3D coordinates $z_1^1(u_1^1, v_1^1, f)$ and $z_2^1(u_2^1, v_2^1, f)$ respectively, while in the reference frame of the second camera they have coordinates $z_1^2(u_1^2, v_1^2, f)$ and $z_2^2(u_2^2, v_2^2, f)$. Up to a change in units, these coordinates can be re-written as

$$p_1^1 = \frac{1}{s_1^1}\begin{pmatrix} u_1^1 \\ v_1^1 \\ f \end{pmatrix},\ p_2^1 = \frac{a}{s_2^1}\begin{pmatrix} u_2^1 \\ v_2^1 \\ f \end{pmatrix},\ p_1^2 = \frac{1}{s_1^2}\begin{pmatrix} u_1^2 \\ v_1^2 \\ f \end{pmatrix},\ p_2^2 = \frac{a}{s_2^2}\begin{pmatrix} u_2^2 \\ v_2^2 \\ f \end{pmatrix},$$

where f is the focal lenght and a is the ratio between the actual scales of the local 3D patches around points p_1 and p_2, whose projections on the two views give the perceived scales s_1^1 and s_1^2 for point p_1 and s_2^1 and s_2^2 for point p_2.

The assumption that both scale and orientation are linked with actual properties of the local patch around each 3D point is equivalent to having 2 points for each feature correspondence: the actual location of the feature, plus a virtual point located along the axis of orientation of the feature at a distance proportional to the actual scale of the patch. These pairs of 3D points must move rigidly going from the coordinate system of one camera to the other, so that given any two sets of correspondences with 3D points p_1 and p_2 and their corresponding virtual points q_1 and q_2, the distances between these four points must be preserved in the reference frames of every view (see Fig. 2).

Under a frontal-planar assumption for each local patch, or, less stringently, under small variation in viewpoints, we can assign 3D coordinates to the virtual points in the reference frames of the two images:

$$\begin{aligned} q_1^1 &= p_1^1 + \begin{pmatrix} \cos\theta_1^1 \\ \sin\theta_1^1 \\ 0 \end{pmatrix} & q_2^1 &= p_2^1 + a\begin{pmatrix} \cos\theta_2^1 \\ \sin\theta_2^1 \\ 0 \end{pmatrix} \\ q_1^2 &= p_1^2 + \begin{pmatrix} \cos\theta_1^2 \\ \sin\theta_1^2 \\ 0 \end{pmatrix} & q_2^2 &= p_2^2 + a\begin{pmatrix} \cos\theta_2^2 \\ \sin\theta_2^2 \\ 0 \end{pmatrix}, \end{aligned}$$

where θ_i^j is the perceived orientation of feature i in image j. At this point, given two sets of correspondences between points in two images, namely the correspondence m_1 between a feature point in the first image with coordinates, scale and orientation $(u_1^1, v_1^1, s_1^1, \theta_1^1)$ with the feature point in the second image $(u_1^2, v_1^2, s_1^2, \theta_1^2)$, and the correspondence m_2 between the points $(u_2^1, v_2^1, s_2^1, \theta_2^1)$ and $(u_2^2, v_2^2, s_2^2, \theta_2^2)$ in the first and second image respectively, we can compute a distance from the manifold of feature descriptors compatible with a single 3D rigid transformation as

$$\begin{aligned} d(m_1, m_2, a) = (||p_1^1 - p_2^1||^2 - ||p_1^2 - p_2^2||^2)^2 + (||p_1^1 - q_2^1||^2 - ||p_1^2 - q_2^2||^2)^2 + \\ (||q_1^1 - p_2^1||^2 - ||q_1^2 - p_2^2||^2)^2 + (||q_1^1 - q_2^1||^2 - ||q_1^2 - q_2^2||^2)^2 . \end{aligned}$$

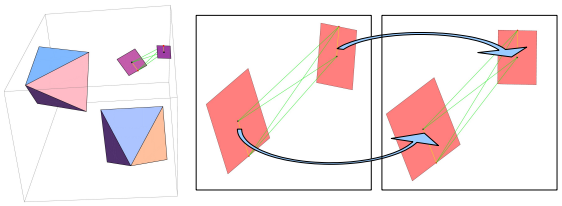


Fig. 2. Scale and orientation offer depth information and a second virtual point. the conservation of the distances in green enforces consistency with a 3D rigid transformation.

From this we define the compatibility between correspondences as $C(m_1, m_2) = \max_a e^{-\gamma d(m_1, m_2, a)}$, where a is maximized over a reasonable range of ratio of scales of local 3D patches. In our experiments a was optimized in the interval $[0.5; 2]$.

The second geometric consistency constraint assumes a weak perspective camera and matches together points whose maps are compatible with a common affine transformation. 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 $T(b_1, b_2)$. 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:

$$a_2' = T(b_1, b_2)a_1$$
$$b_2' = T(a_1, a_2)b_1$$

That is the virtual points obtained by applying to each source point the similarity transformation associated to the other match (see Fig 3). Given virtual points a_2' and b_2' we are finally able to calculate the payoff between (a_1, a_2) and (b_1, b_2) as:

$$\Pi((a_1, a_2), (b_1, b_2)) = e^{-\lambda \max(||a_2 - a_2'||, ||b_2 - b_2'||)} \quad (1)$$

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.

The rationale of the payoff function proposed in equation 1 is that, while by changing point of view the similarity relationship between features is not mantained (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.

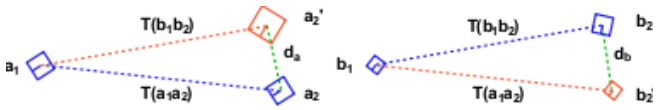


Fig. 3. The payoff between two matching strategies is inversely proportional to the maximum reprojection error obtained by applying the affine transformation estimated by a match to the other.

Each matching 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 significative residual groups start to emerge from the evolution. To avoid the selection of this spurious matches we fixed a minimum cardinality for each valid group.

3 Game-Theoretic Feature Matching

We model the matching process in a game-theoretic framework [1], 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. Clearly, it is in each player's interest to pick matches that are compatible with those the other players are likely to choose. It is supposed that some selection process operates over time on the distribution of behaviors favoring players that receive larger payoffs and driving all inconsistent hypotheses to extinction, finally 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. 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, thus lying in the n-dimensional standard simplex $\Delta^n = \{\mathbf{x} \in \mathbb{R}^n : \forall i \in 1 \ldots n \; x_i \geq 0, \; \sum_{i=1}^n x_i = 1\}$. 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 $\mathbf{y}^T C\mathbf{x}$. A strategy $\mathbf{x}$ is said to be a *Nash equilibrium* if it is the best reply to itself, i.e., $\forall \mathbf{y} \in \Delta, \; \mathbf{x}^T C\mathbf{x} \geq \mathbf{y}^T C\mathbf{x}$. 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 and (hopefully) to ESS's; for this reason, the choice of an actual selection process is not crucial and can be driven

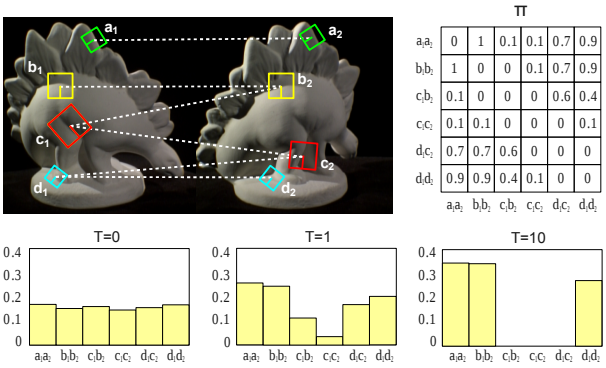


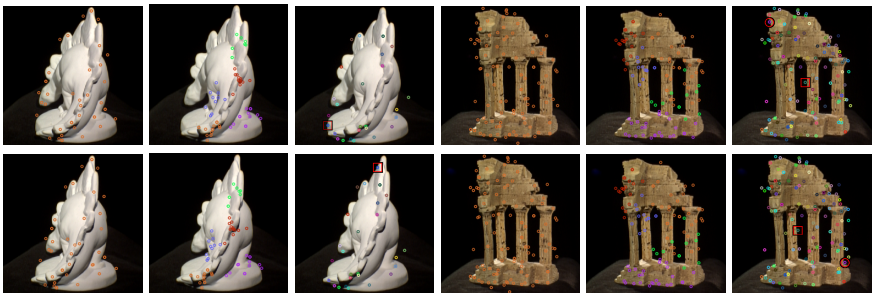
Fig. 4. An example of the evolutionary process. Four feature points are extracted from two images and a total of six matching strategies are selected as initial hypotheses. The matrix Π shows the compatibilities between pairs of matching strategies according to a one-to-one similarity-enforcing payoff 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 a 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 larger support than (d_1, d_2) since they are a little more coherent with respect to similarity.

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 recurrence $\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^t \frac{(C\mathbf{x}^t)_i}{\mathbf{x}^{tT} C \mathbf{x}^t}$, where $\mathbf{x}_i^t$ is the proportion of the population that plays the i-th strategy at time t. Once the population has reached a local maximum, all the non-extincted pure strategies can be considered selected by the game.

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 mutual zero payoff cannot belong to the support of an ESS (see [1]), thus any payoff function can easily be adapted to enforce one-to-one matching by setting to 0 the payoff of mates that share either the source or the destination point.

4 Experimental Results

To evaluate the performance of our proposals, we compared the results with those obtained with the keymatcher included in the structure-from-motion suite



GT-3Drigid GT-2Daffine Bundler GT-3Drigid GT-2Daffine Bundler

Dino sequence

	GT-3Drigid	GT-2Daffine	Bundler
Matches	262.5 ± 61.4	271.1 ± 64.2	172.4 ± 79.5
$\Delta\alpha$	0.0668 ± 0.0777	0.0497 ± 0.0810	0.0767 ± 0.1172
$\Delta\gamma$	0.4393 ± 0.4963	0.3184 ± 0.3247	0.6912 ± 0.8793

Temple sequence

	GT-3Drigid	GT-2Daffine	Bundler
Mathces	535.7 ± 38.7	564.3 ± 37.2	349.3 ± 36.2
$\Delta\alpha$	0.1326 ± 0.0399	0.0989 ± 0.0224	0.1414 ± 0.0215
$\Delta\gamma$	0.0809 ± 0.0144	0.0792 ± 0.0091	0.0850 ± 0.0065

Fig. 5. Results obtained with the Dino and Temple data sets.

Bundler [13]. For the first set of experiments we selected pairs of adjacent views from the "DinoRing" and "TempleRing" sequences from the Middlebury Multi-View Stereo dataset [11]; for these models, camera parameters are provided and used as a ground-truth. For all the sets of experiments we evaluated the differences in radians between the (calibrated) ground-truth and respectively the estimated rotation angle ($\Delta\alpha$) and rotation axis ($\Delta\gamma$). The "Dino" model is a difficult case in general, as it provides very few features; the upper part of Fig. 5 shows the correspondences produced by our game -theoretic matching approach with geometric constraints enforcing a 3D rigid transformation (GT-3Drigid), the approach with the weak perspective camera assumptions (GT-2Daffine), and the Bundler matcher (Bundler). The color of the points matched using GT-2Daffine relate to the extraction group, i.e., points with the same color have been matched at th same re-iteration of the game-theoretic matching process. The "Temple" model is richer in features and for visualization purposes we only show a subset of the detected matches for all three techniques. The Bundler matcher, while still achieving good results, provides some mismatches in both cases. This can be explained by the fact that the symmetric parts of the object, e.g. the pillars in the temple model, result in very similar features that are hard to disambiguate by a purely local matcher. Both our methods, on the other

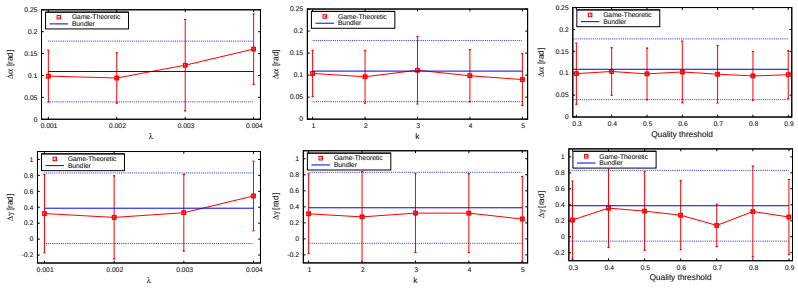


Fig. 6. Analysis of the performance of the approach with respect to variation of the parameters of the algorithm.

hand, by enforcing global consistency, can effectively disambiguate the matches. Looking at the results we can see that both our approaches extract around 50% more correspondences than Bundler. The first approach provides a slight increase in precision and reduction in variance of the estimates. Note, however, that the selected measures evaluate the quality of the underlying least square estimates of the motion parameters after a reprojection step, thus small variations are expected. The approach enforcing a global 2D affine transformation exhibits a larger increase in precision and reduction in variance. This can be explained by the fact that the adjacent views of the two sequences have very little parallax effects, thus the weak persective camera assumption holds quite well. In this context the stricter model is better specified and thus more discriminative.

Next, we analyzed the impact of the algorithm parameters over the quality of the results obtained. To this end, we investigated three parameters: the similarity decay λ, the number k of candidate mates per features, and the *quality threshold*, that is the minimum support for a correspondence to be considered non-extinct, divided by the maximum support in the population. Figure 4 reports the results of these experiments. The goal of these experiments was to show the sensitivity to the matcher's parameters, not to choose between constraints, so only the 3D geometric constraint was used. Overall, these experiments show that almost all reasonable values of the parameters give similar values for the match, thus those parameters have little influence over the quality of the result, with the Game-Theoretic approach achieving better average results and smaller standard deviation than the Bundler matcher.

5 Conclusions

In this paper we introduced a robust matching technique for feature points from multiple views. Robustness is achieved by enforcing global geometric consistency in a pairwise setting. Two different geometric consistency models are proposed. The first enforces the compatibility with a single 3D rigid transformation of the points. This is achieved by using the scale and orientation information offered by SIFT features and projecting what is left of a high-order compatibility problem into a pairwise compatibility measure, by enforcing the conservation of distances

between the unknown 3D positions of the points. The second model assumes a weak perspective camera model and enforces that points are subject to an affine transformation. This extracts only local groups at similar depths, but the matching process is repeated to cover the whole scene. In both cases, a game-theoretic approach is used to select a maximally consistent set of candidate matches, where highly compatible matches are enforced while incompatible correspondences are driven to extinction. Experimental comparisons with a widely used technique show the ability of our approach to obtain more accurate estimates of the scene parameters.

References

1. Andrea Albarelli, Samuel Rota Bulò, Andrea Torsello, and Marcello Pelillo. Matching as a non-cooperative game. In *ICCV 2009*, 2009.
2. C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Fourth Alvey Vision Conference*, pages 147–151, 1988.
3. Tinne Tuytelaars Herbert Bay and Luc Van Gool. Surf: Speeded up robust features. In *9th European Conference on Computer Vision*, volume 3951, pages 404–417, 2006.
4. D. Lowe. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, volume 20, pages 91–110, 2003.
5. Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.
6. D. Marr and E. Hildreth. Theory of Edge Detection. *Royal Soc. of London Proc. Series B*, 207:187–217, February 1980.
7. 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.
8. 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.
9. K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(10):1615–1630, 2005.
10. M. Saquib Sarfraz and Olaf Hellwich. Head pose estimation in face recognition across pose scenarios. In *VISAPP (1)*, pages 235–242, 2008.
11. Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR '06*, pages 519–528, 2006.
12. Jianbo Shi and Carlo Tomasi. Good features to track. In *CVPR*, pages 593–600, 1994.
13. Noah Snavely, Steven M. Seitz, and Richard Szeliski. Modeling the world from internet photo collections. *Int. J. Comput. Vision*, 80(2):189–210, 2008.
14. Andrea Torsello, Samuel Rota Bulò, and Marcello Pelillo. Grouping with asymmetric affinities: A game-theoretic perspective. In *CVPR '06*, pages 292–299, 2006.
15. J. Y. Zien, M. D. F. Schlag, and P. K. Chan. Multi-level spectral hypergraph partitioning with arbitrary vertex sizes. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 18:1389–1399, 1999.