2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Robust Joint Selection of Camera Orientations and Feature Projections over Multiple Views

Mara Pistellato, Andrea Albarelli, Filippo Bergamasco and Andrea Torsello
Dipartimento di Scienze Ambientali, Informatica e Statistica
Università Ca' Foscari Venezia
via Torino 155, 30172 Mestre(VE), Italy
mara.pistellato@gmail.com, albarelli@unive.it, filippo.bergamasco@unive.it, torsello@unive.it

***Abstract*—A number of critical factors arises when a complex 3D scene is to be reconstructed by means of a large sequence of different views. Some of them are related to the ability to recover the correct identity and the accurate projection of each observed feature. Other sources of error are tied to the reliability of the orientation estimate for each view. With this paper we propose a method that tries to solve both problems at the same time, while being also inherently resilient to outliers. At the core of the approach stands a widely adopted game-theoretical selection technique, which has already been successfully embraced to address similar tasks. The original inception, however, has been further refined to address a wider range of scenarios, as well as to offer a reduced memory consumption and computation complexity. By exploiting these enhancements, we were able to apply our technique to a large-scale setup involving several hundreds of view points and tens of thousands of independent observations.**

I. Introduction

Image-based 3D reconstruction relies on two major factors: the ability to match observations from different cameras and the knowledge of the relative pose between these cameras. Over the last decades countless approaches have been proposed to solve both problems.

Correspondences can be found by exploiting the local appearance of the scene, by means of concise descriptors capturing such information including SIFT [1], SURF [2], GLOH [3], BRISK [4], FREAK [5] and many others. While all these descriptors seek repeatability and distinctiveness, they are still prone to lead to false matching due to noise or similarities in appearances. Correspondence reliability can be improved, for instance, using high-level matching frameworks accounting for multi-feature consistency [6], [7], [8] or by ditching photometric descriptors in favor of intrinsically robust identification methods, including structured light or artificial markers [9].

Several methods also exist to estimate the pose between cameras. Some of them adopt special calibration targets characterized by a known model based on features exhibiting some projective invariant, such as squares [10] or circles [11]. Such approaches, however, are only feasible when dealing with fixed cameras that can be calibrated offline. Differently, self calibration methods allow to exploit directly the observed features. This usually happens by minimizing the overall reprojection error of feature points triangulated under the estimated poses [12]. The use of self calibration is especially useful when addressing scenarios including multiple cameras organized in a network [13], [14] or a sequence of frames generated by an unknown camera motion [15].

Unfortunately, when dealing with feature-based self calibration, points labeling and pose estimation are tightly intertwined. As a matter of fact, uncertainty about points localization on the image and wrong features labeling could result in large pose estimation error, due to the ill-posed reprojection function to be minimized. On the other hand, an inaccurate pose estimate could hinder any attempt to validate correspondences on the basis of the epipolar constraint. Of course many solutions have been proposed to deal with this quandary, such as RANSAC-based feature selection [16], restriction to affine transforms [17], [6] or enforcement of the transitive closure for the recovered poses [18].

With this paper, we propose a novel approach to the simultaneous selection of camera orientations and corresponding feature projections. The key idea underlying our method is to adopt a game-theoretical validation of mutually consistent observations. Indeed, Game theory has been successfully used as an inlier selection framework for problems ranging from image registration [19] and classification [20] to interest points detection [21]. Actually, game-theoretical methods are not novel even with respect to camera poses selection. In fact a dynamic optimal path selection for camera networks has already been introduced in [22], where abundant triangulation hypotheses, produced by means of different pairs of cameras, are validated according to mutual consistence.

This work advances such previous method by extending its scope and, at the same time, by reducing its computational complexity and memory footprint. Namely, the original approach was only able to select paths of rigid transforms connecting cameras observing the same physical point or a set of already labeled points. By contrast, our method works equally well with unlabeled points, making it possible to find correspondences between observations. Moreover, if some function is available to measure the compatibility between observations, it can be exploited to reduce the probability of mismatches. Finally, since we operate directly on the point images (rather than on their triangulations), the number of potential hypotheses to validate is reduced from $O(n^2)$ to $O(n)$, where n is the total number of observations among all the involved cameras.

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

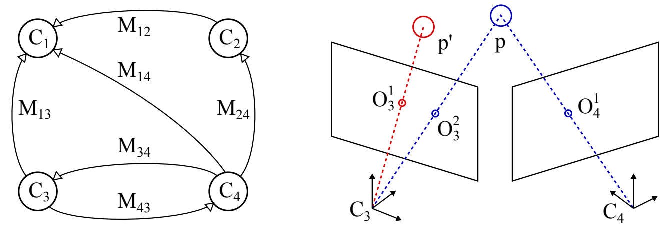

Fig. 1. The general scenario we are addressing includes several cameras, a (partial) graph of pairwise pose estimation and a set of observations of physical points (see the text for a detailed description).

II. General Context and Previous Work

The general context we are dealing with includes a network of k cameras $C = \{C_1, C_2 \dots C_k\}$. Each camera, referred by an unique label C_α, observes the same scene and captures projective images of material points in the scene as single observations O^u_α, where u is the sequential label of the observation for that particular camera. A set of rigid transformations $M = \{M_{11}, M_{12} \dots M_{kk-1}\}$ between pairs of cameras is available. Within this set each entry M_{ij} is a 4×4 roto-translation matrix which transforms 3D points expressed in the reference frame of C_j to point expressed in the reference frame of C_i. We make no assumption about how observations and pairwise transformations are obtained, thus each camera could observe just a partial subset of the points in the scene, each point could be observed with any amount of positional error on the image plane, it could be subject to wrong labeling, and the rigid transform between a pair of cameras might be available or not and, when available, could be subject to any amount of uncertainty (see Fig. 1). Within this context our goal is not to correct the error affecting the observations or the camera poses. Rather we are seeking, on a point-by-point basis, for the best combination of camera orientations and point projections to get the most reliable 3D reconstruction.

A. Game-Theoretical Inlier Selection

Following other inlier selection schemas (such as [23] or [24]), we ground our notion of reliability upon the mutual support between pair of hypotheses $H_i, H_j \in H$, expressed by some payoff-monotonic function $\pi(i, j) : H \times H \to \mathbb{R}_{\geq 0}$, where i and j are labels to hypotheses H_i and H_j. By properly selecting the hypotheses to compare and the payoff function it is possible to solve different types of problems by enforcing the selection of a mutually consistent set of hypotheses by means of an evolutionary process.

This evolutionary process is performed over a *population*, that is a discrete probability distribution $\vec{x} = (x_1, \dots, x_n)^T$ over the available strategies H. Any population vector is bound to lie within the n-dimensional standard simplex $\Delta^n = \{\vec{x} \in \mathbb{R}^n \; : \; x_i \geq 0 \text{ for all } i \in 1 \dots n, \; \sum_{i=1}^n x_i = 1\}$ The *support* of a population $\vec{x} \in \Delta^n$, denoted by $\sigma(\vec{x})$, is defined as the set of elements chosen with non-zero probability: $\sigma(\vec{x}) = \{i \in O \mid x_i > 0\}$. In order to find a set of mutually coherent hypotheses, we are interested in finding configurations of the population maximizing the average payoff. Given the matrix $\Pi = (\pi_{ij})$, where $(\pi_{ij}) = \pi(i, j)$, the total payoff obtained by hypothesis i within a given population $\vec{x}$ is $(\Pi\vec{x})_i = \sum_j \pi_{ij} x_j$, and the (weighted) average payoff over all the considered hypotheses is exactly $\vec{x}^T \Pi \vec{x}$. Unfortunately, it is not immediate to find the global maximum for $\vec{x}^T \Pi \vec{x}$, however local maxima can be obtained using a rather wide class of evolutionary dynamics called *Payoff-Monotonic Dynamics*. A quite common evolutionary process starts by setting an initial population $\vec{x}$ near the barycenter of the simplex and then proceeds by evolving it through the discrete-time replicator dynamic [25]:

$$x_i(t+1) = x_i(t) \frac{(\Pi\vec{x}(t))_i}{\vec{x}(t)^T \Pi \vec{x}(t)} \tag{1}$$

where x_i is the i-th element of the population and Π the payoff matrix. This family of dynamics are guaranteed to converge to an equilibrium where the support does not include strategies with mutual payoff zero. This means that hard constraints expressed by zero compatibility are actually guaranteed to be enforced. Finally, once the equilibrium is reached, the density of the population vector corresponding to each hypothesis can be used to assess its degree of participation to the inliers set.

B. Hypotheses and Payoff for the Optimal Path Selection

As noted in the introduction, our method builds directly over the Optimal Path Selection (OPS) approach [22]. The basic assumption of OPS is that a (mostly) correct labeling already exists, thus correspondences are already established and the optimal paths can be selected independently for each material point. Within OPS, the set H of hypothesis to be validated is made up of all the possible triangulations. To this end, it must include exactly two observations and two paths connecting the observing cameras to the world frame. Without any loss of generality we can assume the world frame aligned with C_1, this way each hypothesis is a quadruple $(M_{1x} \dots M_{yi}, O^a_i, M_{1w} \dots M_{vj}, O^b_j)$. Here $M_{1x} \dots M_{yi}$ and $M_{1w} \dots M_{vj}$ are paths that combine a sequence of rigid transformation matrices connecting cameras C_i and C_j to C_1. Accuracy of observations depends on several sources of noise, usually distributed as a zero-mean Gaussian. In addition, each wrong labeling could result in a totally misplaced point.

The payoff function is designed to account respectively for the consensus between triangulations coming from two hypotheses and their individual reliability. This is done by considering both the positions of the reconstucted points $\vec{x}(H_i)$ and the triangulation skewness $s(H_i)$. Even if these two error sources are not really independent, we can still reasonably approximate them as a bidimensional Gaussian function:

$$\pi'(H_i, H_j) = e^{-\frac{1}{2}\left(\frac{(|\vec{x}(H_i) - \vec{x}(H_j)|)^2}{\sigma_p^2} + \frac{max(s(H_i), s(H_j))^2}{\sigma_s^2}\right)} \tag{2}$$

Where σ_p and σ_s are two parameters that represent respectively the expected standard deviation of point position and of skewness. Note that $(|\vec{x}(H_i) - \vec{x}(H_j)|)^2$ is a pairwise measure. Differently, $s(H_i)$ and $s(H_j)$ are independent one from the other, thus the max operator is needed to account for them within the pairwise function π'.

While π' expresses the degree of consensus between hypotheses, we must also account for special cases where two hypotheses are not compatible regardless of the quality of triangulation. The first one is the case when the hypotheses include two different observations from the same camera. The other unfeasible case consists in the presence of two different paths to the same camera. As observed in Sec. II-A, these constraint can be enforced by explicitly setting a value of zero in the final payoff function:

$$\pi(H_i, H_j)) = \pi((P^u_\alpha, O^a_\alpha, P^{u'}_{\alpha'}, O^{a'}_{\alpha'}), (P^v_\beta, O^b_\beta, P^{v'}_{\beta'}, O^{b'}_{\beta'})) \tag{3}$$

$$= \begin{cases} 0 & \text{if} \quad \alpha = \beta \wedge (u \neq v \vee a \neq b) \\ & \quad\quad \alpha' = \beta' \wedge (u' \neq v' \vee a' \neq b') \\ & \quad\quad \alpha' = \beta \wedge (u' \neq v \vee a' \neq b) \\ & \quad\quad \alpha = \beta' \wedge (u \neq v' \vee a \neq b') \\ \pi'(i,j) & \text{otherwise} \end{cases}$$

where P^u_α represents a path connecting camera C_α to the world frame.

C. Shortcomings of the Optimal Path Selection Method

All the major shortcomings of OPS are due to the definition of the hypothesis set H. In fact, since the method works by validating triangulations, it needs to know in advance which projections to triangulate, thus a cadidate labeling must be provided as an input. While such labeling can be obtained with several methods, it is clear that this requirement limits the overall utility of the method, which can be used only to select the optimal path for each triangulation (hence the name).

The other main hurdle with OPS is related to its complexity, both in terms of memory and computational requirements. Since each hypothesis is based on a combination of two observations through two distinct paths, even with perfect labeling the overall number of hypotheses grows with the square of the feasible paths. Since the side of matrix Π is equal to the size of the hypotheses set $|H| \simeq |P|^2$, this means that each iteration of equation (1) is potentially $O(|P|^4)$. This is also true for the size in memory of matrix Π, albeit it is not mandatory to actually store the entries of Π in memory since they can be computed on the fly (although, increasing the overall number of computations to perform).

III. A Robust Observations Selection

The proposed enhancements over OPS are quite simple, still they represent a substantial upgrade of the original method both in terms of capabilities and scalability. The key idea is that the preliminary triangulation can be totally avoided by adopting a simpler (but less strict) hypotheses set.

A. Enforcing Transitive Closure on a Reduced Hypothesis Set

We propose to base the selection process on a set H where each hypothesis is just the line of sight (or *ray*) resulting from the combination of a path P and an observation O. Since this new hypothesis corresponds to a line in space, rather than to a point, the payoff functions (2) and (3) make no sense anymore, since the Euclidean distance between triangulated points can't be computed. Our guess, however, is that equally good results can be obtained using a reduced formulation that only accounts for the skewness of the rays:

$$\pi'(i,j) = \pi'((P^u_\alpha, O^a_\alpha), (P^v_\beta, O^b_\beta)) = e^{-\frac{1}{2}(\frac{s(H_i,H_j)^2}{\sigma_s^2})} \tag{4}$$

Of course it is much easier for two rays to exhibit low skewness by chance, in fact this happens for all the rays lying in the same epipolar plane for a pair of cameras. The rationale of the approach, however, is that for a large enough population of candidate rays, the probability of transitively exhibiting low skewness by chance among all the pairs is much lower. In principle this could happen for the epipolar subspace, but this is in general a different plane for different pairs of cameras. Of course, for this guess to be satisfying, it must withstand a proper experimental validation, which we perform in Sec. IV.

B. Joint Path Selection and Feature Labeling

Given the smaller size of H (which is reduced to the square root from OPS) and since no early triangulation is performed, there is nothing preventing to include more than one material point at the same time and letting the evolutionary process select the clusters of rays belonging to the same bundle. Note that some a-priori information about the likelihood of two observations to be related might be available (feature descriptors, tracking, etc.). If such information is available, without loss of generality, it can be used to define a compatibility function $C(H_i, H_j) : H \times H \rightarrow \{0, 1\}$, indicating the feasibility of the correspondence according to the a-priori information. With this function at hand, the complete payoff can be defined as:

$$\pi(H_i, H_j)) = \begin{cases} 0 & \text{if } C(H_i, H_j) = 0 \text{ or } \alpha = \beta \\ \pi'(i,j) & \text{otherwise} \end{cases} \tag{5}$$

where, in addition to th compatibility function C, two observations are automatically not corresponding if they came from the same camera.

The first iteration of the evolutionary process based on (5) will yield a single material point and the same label can be applied to all the supporting rays. Once the bundle is obtained two actions must be undertaken:

- the point must be triangulated according to the rays that are not extinct in the final population. This can be done in several ways, the simpler being by just triangulating the two rays with the higher density. A more sensible approach (which we will use in the evaluation) is to find the point that minimizes the squared distance from all the rays, weighted according to the population density of each ray;
- the non extinct rays must be removed from the hypotheses set H since they have already been assigned to a material point.

Once H has been reduced, additional iterations of the evolutionary process can be performed, until all the observations have been labeled or a satisfying number of material points has been reconstructed.

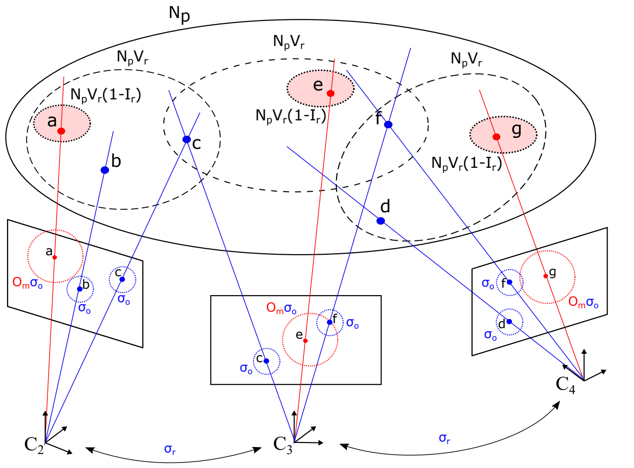

Fig. 2. A graphical depiction of the 6 main factors that are accounted for in our synthetic experimental setup (see the text for a detailed description).

IV. Experimental Validation

Beside the reduced complexity, there is unfortunately no technical or theoretical reason that guarantees the proposed enhancement to be superior (or even on par) with respect the original OPS method. To this end, the only sensible way to assess the merits of this novel technique is to perform an extensive set of comparisons over different experimental conditions. This has been done using synthetically generated data. The rationale of the synthetic analysis is based on the need of a well-defined ground truth to be able to finely assess the performance of the technique under a range of controlled conditions.

A. Synthetic experiments setup

The final goal of each reconstruction technique is to accurately recover the geometry of the scene regardless of the error sources. For this reason, within our synthetic experiments, we adopted as a measure of the quality of the obtained result the RMS error between the ground-truth model and the reconstructed points. It should be noted that some methods we compare against might change the reference system due to modifications of the camera poses. In order to be invariant with respect to such effects, the RMS has always been computed after a best-fit alignment between the obtained points and the model [26]. We call this quality measure P_{rms}.

To be able to study different aspects of the proposed method, we devised a general synthetic experimental framework which allows to control a wide range of different hindrances that could affect the process. Specifically, we think that it is critical to be able to control the overall density of material points, the visibility achieved by each camera over the scene, the different noise sources and the amount of outliers, i.e. incorrectly labeled correspondences. A graphical overview of the simulated scenario is presented in Fig.2. The control parameters, also depicted in the figure, are the following:

- N_p (Number of points): the number of material points that have been generated in the model observed by the cameras. Such points are generated in a cube box of side 10 and captured by cameras with focal length 1;
- V_r (Visibility ratio): the ratio of material points that can be observed by each camera. This parameter is needed to model complex networks (especially under motion scenarios) where only a portion of the points are captured by a given camera;
- σ_r (Orientation rotational error): the standard deviation of the rotational noise added to the orientation of each camera in the network. The rationale of adding only rotational noise (as opposite to translational) is to allow a single measure of noise intensity while still adressing the most influential factor;
- σ_o (Observation positional error): the standard deviation of the positional noise added to the projection of each observed material point;
- I_r (Inlier ratio): the ratio of observations that can be classified as inliers;
- O_m (Outlier multiplier): the multiplier to be applied to σ_o in order to amplify the positional error for outliers.

According to this parameterization, each camera observes exactly N_pV_r material points from the scene. Among these points, the projections of exactly $N_pV_rI_r$ are subject to an additive positional error with standard deviation σ_o, while the remaining, being outliers, are displaced by a (larger) positional error of standard deviation σ_oO_m.

Following [22] we created different network topologies, namely: grid, hemisphere and line (see Fig. 3). We first generated the exact relative motion between couple of adjacent cameras and then we added rotation noise distributed according to a Gaussian of standard deviation σ_r. All the evaluation presented have been computed by creating a ground truth of points using the specified parameters and then averaging the P_{rms} obtained over the three topologies.

Using this setup we first performed an analysis of the sensitivity of the method to its parameters, then we compared the new technique with the original OPS method, as well as with the same two baseline approaches adopted in [22].

B. Sensitivty Analysis

The Joint Observation and Projection Selection (JOPS) methods depends on a single parameter σ_s. To assess how sensitive is the method with respect to this parameter we measured the average P_{rms} obtained for different values of σ_s with fixed experimental conditions ($N_p = 10, V_r = 0.9, \sigma_r = 0.018, \sigma_o = 7 \cdot 10^{-3}, I_r = 0.9, O_m = 10$). As shown in

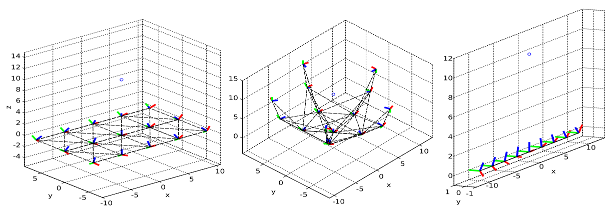

Fig. 3. The three different network topologies used for synthetic experiments.

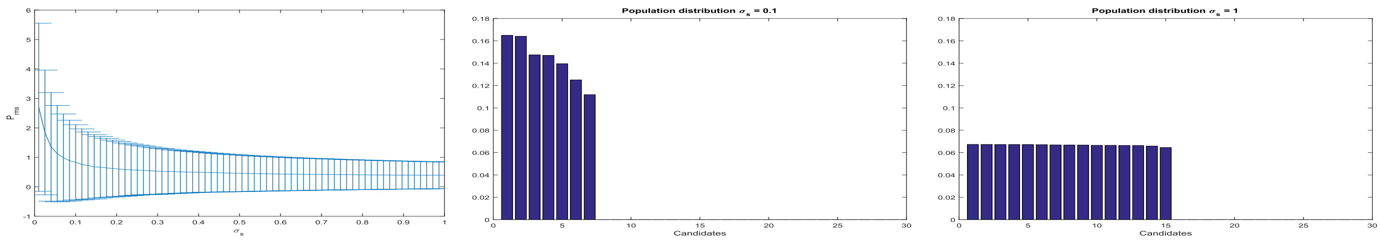

Fig. 4. Study of the behaviour of our approach with respect to the sensitivity parameter σ_s.

the first plot of Fig. 4, JOPS is not very sensitive to σ_s as long as it is large enough to avoid a too strict inlier selection. This is a behaviour similar to the one exhibited by OPS in [22], however being dependent on a single parameter makes it easier to tune the method. Notice also that, while increasing the value of σ_s above a threshold of about 0.5 has a limited effect on P_{rms}, the shape of the selected population can be very different. As shown in the second and third plot of Fig. 4, stricter values lead to a tighter selection with less successful hypotheses. This could result in a less stable triangulation (due to the reduced number of samples contributing to the average) and on a duplicate detection of the same physical point, which could appear in subsequent evolutionary processes, due to the lower number of rays culled out.

C. Comparison with Optimal Path Selection

We compared the performance of JOPS with the original OPS, Bundle Adjustment (SBA) and Dual Quaternion averaging [18]. Since the other methods cannot be used to perform feature matching, in order to be able to compare JOPS with them we had to reduce the scope of our method. We defined a labeling l according to the actually observed points and a compatibility function to be used with JOPS:

$$C(H_i, H_j) = \begin{cases} 1 \text{ if } & l(H_i) = l(H_j) \\ 0 \text{ if } & l(H_i) \neq l(H_j) \end{cases} \tag{6}$$

Regarding the other methods, they have been feeded directly with the correct matches.

The sensitivities to different error sources have been studied separately. This has been done by exploring the obtained P_{rms} by varying one error source and keeping the others fixed. The results are shown in Fig. 5. The performance with increasing levels of noise on observations (first plot) and on orientations (second plot) are on par with OPS and are superior to both SBA and Dual Quaternions (which has already been observed in [22]). An accuracy comparable to OPS is also obtained for the tested ranges of visibility ratio (V_r) and inlier ratio (I_r). These latter experiments, however, highlight the shortcomings of SBA, which is unable to reconstruct the correct orientations where not enough correct observations are available and of Dual Quaternion, which is unable to avoid outliers since it just deals with orientations.

D. Applications to Large-Scale Scenarios

The final and most important test has been designed to simulate a general scenario with about 200 cameras before a scene of about 1000 material points producing unlabeled observations. The observations are characterized by an hypothetical descriptor vector that can be used to filter out unfeasible correspondences. For this hypotetical descriptor we want to simulate different levels of repeatability and distinctiveness. To this end we define the following compatibility function:

$$C(H_i, H_j) = \begin{cases} 1 \text{ with prob. } R_f, & \\ 0 \text{ with prob. } 1 - R_f \text{ if } & l(H_i) = l(H_j) \\ 0 \text{ with prob. } D_f, & \\ 1 \text{ with prob. } 1 - D_f \text{ if } & l(H_i) = l(H_j) \end{cases} \tag{7}$$

The value R_f stands for *Repeatability Factor* and models the probability that the descriptors computed over two projections of the same physical points are considered corresponding. By contrast, $1 - R_f$ expresses the probability to be unable to recognize a matching feature. The value D_f stands for *Distinctiveness Factor* and represents the probability that the projections of two different material points are actually considered not corresponding. According to these parameters, compatibility function (6) corresponds to $R_f = 1$ and $D_f = 1$, by contrast, a totally worthless descriptor corresponds to $R_f = 0$ and $D_f = 0$.

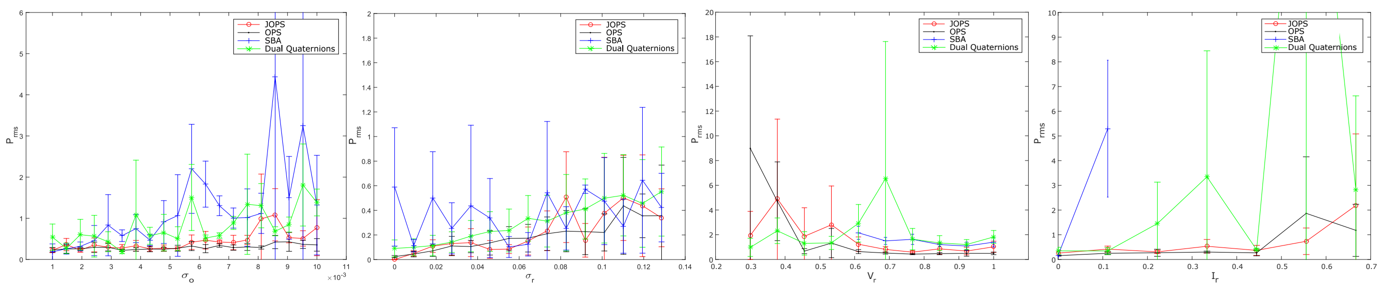

Fig. 5. Comparison of our approach with OPS and other widely adopted methods.

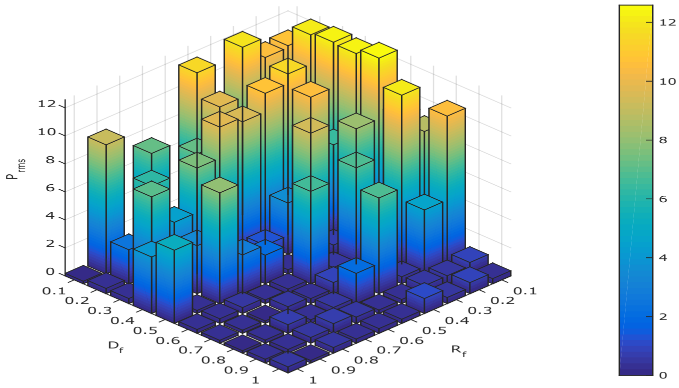

Fig. 6. Performance of JOPS for different degrees of (simulated) feature descriptor repeatability (R_f) and distinctiveness (D_f).

In Fig. 6 we show the average P_{rms} obtained for different combinations of R_f and D_f. We can observe that JOPS behaves very well also with substantially low values for R_f and D_f (which rarely occurs with well designed descriptors). Moreover, when the repeatability is high enough, the distinctiveness is less critical since, as long as correct hypotheses are available in H, outliers can be easily phased out by the evolutionary process because of lack of geometrical consistency.

V. Conclusions

With this paper we are proposing a game-theoretical approach that is able to jointly perform the selection of correspondences between point projections and of the optimal camera orientations required to triangulate such observations. This approach can be adopted over a wide range of scenarios, including camera networks, sequences of frames or collection of images coming from the web. In fact, we make no assumption about the method used to obtain the initial pose estimation. Neither we make any assumption as to the technique used to capture material point images, however, if some kind of feature descriptor or similarity assessment is available, it can be easily exploited. The experimental evaluation of the method has shown that it is able to yield a performance level comparable to a similar approach characterized by a narrower application range and some major shortcomings. Finally, when adopted to solve the more general problem of simultaneous observations and paths selection, it exhibited a strong resilience even when dealing with feature descriptors characterized by low distinctiveness and repeatability.

References

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346 – 359, 2008, similarity Matching in Computer Vision and Multimedia.

[3] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, Oct 2005.

[4] S. Leutenegger, M. Chli, and R. Y. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *2011 International Conference on Computer Vision (ICCV)*, Nov 2011, pp. 2548–2555.

[5] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast retina keypoint," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 510–517.

[6] A. Albarelli, E. Rodol, and A. Torsello, "Imposing semi-local geometric constraints for accurate correspondences selection in structure from motion: A game-theoretic perspective," *International Journal of Computer Vision*, vol. 97, no. 1, pp. 36–53, 2012.

[7] B. Jiang, J. Tang, B. Luo, and L. Lin, "Robust feature point matching with sparse model," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5175–5186, 2014.

[8] H. Chen, Y. Lin, and B. Chen, "Robust feature matching with alternate hough and inverted hough transforms," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2762–2769, 2013.

[9] A. Kolahi, M. Hoviattalab, T. Rezaeian, M. Alizadeh, M. Bostan, and H. Mokhtarzadeh, "Design of a marker-based human motion tracking system," *Biomedical Signal Processing and Control*, vol. 2, no. 1, pp. 59–67, 2007.

[10] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, Nov 2000.

[11] J. Heikkila, "Geometric camera calibration using circular control points," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1066–1077, Oct 2000.

[12] M. I. Lourakis and A. A. Argyros, "Sba: A software package for generic sparse bundle adjustment," *ACM Transactions on Mathematical Software (TOMS)*, vol. 36, no. 1, p. 2, 2009.

[13] T. Svoboda, D. Martinec, and T. Pajdla, "A convenient multicamera self-calibration for virtual environments," *Teleoperators and virt. environments*, vol. 14, no. 4, pp. 407–422, 2005.

[14] R. K. Kumar, A. Ilie, J.-M. Frahm, and M. Pollefeys, "Simple calibration of non-overlapping cameras with a mirror," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–7.

[15] P. H. Torr and A. Zisserman, "Feature based methods for structure and motion estimation," in *Vision Algorithms: Theory and Practice*. Springer, 1999, pp. 278–294.

[16] D. Nistér, "Preemptive ransac for live structure and motion estimation," *Machine Vision and Applications*, vol. 16, no. 5, pp. 321–329, 2005.

[17] J. J. Koenderink and A. J. van Doorn, "Affine structure from motion," *J. Opt. Soc. Am. A*, vol. 8, no. 2, pp. 377–385, Feb 1991.

[18] A. Torsello, E. Rodola, and A. Albarelli, "Multiview registration via graph diffusion of dual quaternions," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 2441–2448.

[19] J. Banerjee, C. Klink, E. Peters, W. Niessen, A. Moelker, and T. van Walsum, "Fast and robust 3d ultrasound registration - block and game theoretic matching," *Medical Image Analysis*, vol. 20, no. 1, pp. 173–183, 2015.

[20] H. Aghababaee and M. Sahebi, "Game theoretic classification of polarimetric sar images," *European Journal of Remote Sensing*, vol. 48, pp. 33–48, 2015.

[21] A. Albarelli, E. Rodol, and A. Torsello, "Loosely distinctive features for robust surface alignment," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6315 LNCS, no. PART 5, pp. 519–532, 2010.

[22] M. Pistellato, F. Bergamasco, A. Albarelli, and A. Torsello, "Dynamic optimal path selection for 3d triangulation with multiple cameras," in *Image Analysis and Processing - ICIAP 2015*, 2015, pp. 468–479.

[23] A. Albarelli, S. R. Bul, A. Torsello, and M. Pelillo, "Matching as a non-cooperative game," in *2009 IEEE 12th International Conference on Computer Vision*, Sept 2009, pp. 1319–1326.

[24] A. Albarelli, E. Rodol, and A. Torsello, "Fast and accurate surface alignment through an isometry-enforcing game," *Pattern Recognition*, vol. 48, no. 7, pp. 2209–2226, 2015.

[25] P. Taylor and L. Jonker, "Evolutionarily stable strategies and game dynamics," *Mathematical Biosciences*, vol. 40, pp. 145–156, 1978.

[26] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. of the Optical Society of America. A*, vol. 4, no. 4, pp. 629–642, Apr 1987.