Dynamic Optimal Path Selection for 3D Triangulation with Multiple Cameras

Mara Pistellato, Filippo Bergamasco, Andrea Albarelli and Andrea Torsello

Dipartimento di Scienze Ambientali, Informatica e Statistica
Università Ca' Foscari Venezia, via Torino 155, Venice, Italy

Abstract. When a physical feature is observed by two or more cameras, its position in the 3D space can be easily recovered by means of triangulation. However, for such estimate to be reliable, accurate intrinsic and extrinsic calibration of the capturing devices must be available. Extrinsic parameters are usually the most problematic, especially when dealing with a large number of cameras. This is due to several factors, including the inability to observe the same reference object over the entire network and the sometimes unavoidable displacement of cameras over time. With this paper we propose a game-theoretical method that can be used to dynamically select the most reliable rigid motion between cameras observing the same feature point. To this end we only assume to have a (possibly incomplete) graph connecting cameras whose edges are labelled with extrinsic parameters obtained through pairwise calibration.

Keywords: Game Theory, 3D Reconstruction, Extrinsic Calibration

1   Introduction

Two points of view are in general enough to reconstruct 3D information from 2D projections. Nevertheless, in many practical scenarios, the adoption of multiple independent cameras is the preferred choice. This is the case, for instance, when people have to be tracked within large areas [1] and strong resilience to occlusion is sought [2].A collection of different points of view can also result from dynamic scenarios, where cameras are mounted on drones [3] or images are collected by different users on social networks and online services [4]. Camera grids can also be very helpful when the phenomenon to be studied is difficult to analyze from a single point of view. This happens with many Computer Vision tasks, ranging from human action recognition [5] to video surveillance [6] and tracking [7]. The adoption of multiple cameras could finally lead to improved accuracy with image-based surface reconstruction [8], especially when dealing with complex artifacts [9]. Of course, for any of these applications to be feasible, the geometry of the cameras must be known, at least to some extent. For this reason, a lot of calibration methods that can be used to recover intrinsic and extrinsic parameters have been proposed. Classic approaches use artificial targets with known geometry [10] to compute intrinsic parameters and simultaneously assess the relative pose of each camera [11]. Other methods perform pairwise calibrations that can be subsequently made consistent with respect to
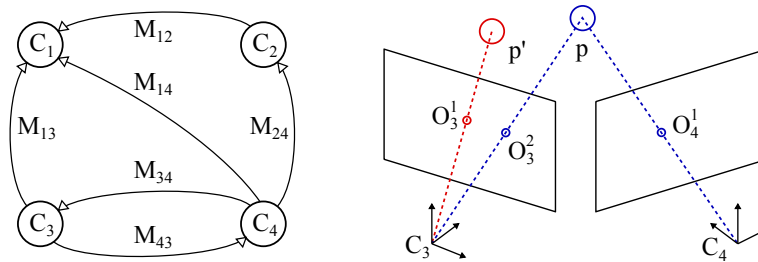
Fig. 1. Overview of the addressed scenario. Multiple cameras are loosely registered through a partially incomplete graph of extrinsic transformation. A physical point is observed by different cameras through an imaging process afflicted by random positional noise and a possibly wrong labelling.

a reference world [12]. It is also very common to exploit the scene itself to obtain simultaneous extrinsic calibration ad structure reconstruction [13]. Finally, many techniques have been proposed to calibrate a whole camera network at the same time [14, 15]. Regardless of the method of choice, any calibration procedure will result in some degree of inaccuracy. In addition, even very accurate calibrations could deteriorate over time, as a consequence of movements of the cameras, slight changes in the network topology or drift of intrinsic parameters. With this paper we are not proposing a new calibration method. Neither are we interested in how the calibration has been performed. We are just assuming that a (possibly large) set of cameras are available and that some process assessed their extrinsic position up to its best accuracy. Within this scenario, our goal is not to enhance the calibration or to correct its precision. Rather we are proposing a consensus-based approach to select the more reliable set of extrinsic parameters that can be used to perform triangulation given the current calibration. This happens dynamically and depends on several factors, including the position of the material points to be triangulated, the quality of the observations on the image plane and the quality of initial extrinsic estimate.

2 General Scenario and Main Contributions

The general scenario we are working with is well represented in Fig. 1. We are dealing with a network of cameras $C = \{C_1, C_2 \ldots C_n\}$, each referred by a unique label C_i. The typical size of such network can range from a few to several tenths of independent devices. The cameras have been previously calibrated for both intrinsic and extrinsic parameters. Within this paper, we are using the term *intrinsic calibration* to refer to all the parameters needed to convert the image acquired by the cameras to normalized image planes, as captured by an ideal pinhole imaging process. By contrast, the term *extrinsic calibration* is used to define the rigid transformations relating the camera reference frame to a common world frame. To avoid any loss of generality, we make no assumption about extrinsic calibration. Indeed, we model it as a set of rigid transformations $M = \{M_{11}, M_{12} \ldots M_{nn}\}$ where 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. Note that we are not assuming such matrices to be transitively consistent. In fact, if the calibration method that produced M is pairwise, we might easily have that, for instance $M_{13} \neq M_{12}M_{23} \neq M_{14}M_{43}$. Differently, if the calibration method is global, all the elements of M will be consistent with respect to the same world, still this does not mean that they exhibit the same accuracy, since it really depends on how they are built. For instance, if they are built by incrementally chaining pairwise registrations, errors can easily add up. Note that, although we are assuming $M_{ii} = I$, we are not pretending $M_{ij} = M_{ji}$ to be true, since this property also depends on how the extrinsic calibration was performed. Most of current literature aims to enforce consistency in the calibration graph $G = \{C, M\}$. This usually happens by means of some averaging process or by a-priori selection of optimal paths in G. With this paper we take quite the opposite approach. In fact, we propose to maintain the whole graph G and to select the optimal weighted subset of paths (extracted from $\mathcal{P}(M)$) for every triangulation to be performed. To this end we assume that two or more cameras from C are able to observe the same physical point p (see the right half of Fig. 1). This will result in a number of points on the image planes, which are subject to random observation uncertainty (usually modeled as zero-mean Gaussian additive noise). In addition, since a scene can comprise several physical points, the labelling process that identifies the same feature throughout all the observing cameras could fail, resulting in outliers related to wrongly labelled observations. In the following we will refer to observation a from camera i as O_i^a. An example is shown in Fig. 1. Here, cameras C_3 and C_4 are observing material point p. A single observation, labelled O_4^1 is reported by camera C_4. Differently, two observation are available from camera C_3, being O_3^2 generated by the correct point of interest p, and O_3^1 resulting from a wrong labelling of the p'. In order for the position of p to be recovered from its projections, four elements must be selected: two observations and two paths connecting the observing cameras to the global reference. For instance, if we assume C_1 to be the world reference for our example, two feasible triangulation hypotheses for p could be $(M_{14}M_{43}, O_3^2, M_{14}, O_4^1)$ and $(M_{13}, O_3^1, M_{12}M_{24}, O_4^1)$.

The ultimate goal of our method is to find the best way to triangulate p. This happens by modeling each possible pair of paths and observations as strategies in a non-cooperative game and by letting them play together until a consistent weighted subset emerges. Key advantages of this approach include the ability to select different paths for different regions for the Euclidean space. This makes sense, since, out of geometrical considerations, is easy to observe that different type of calibration errors will result in non-isotropic inaccuracies throughout the space. Moreover, calibration quality could change over time due to physical changes, making it desirable to dynamically select the optimal paths. Also our method works equally well if extrinsic calibration are consistent with a common space (i.e. transitively closed) and when only loosely related pairwise transformations are available. Finally, by properly defining the compatibility between alternative labelling, we automatically filter outliers from the final solution.

3 Game-Theoretical Optimal Path Selection

Game Theory was introduced in the early 40's by J. von Neumann to model the behavior of entities with competing objectives, and was further developed by J. Nash in the postwar period [16] through the introduction of the Nash equilibrium. According to this view, the emphasis shifts from the search for an optimum shared by the population, to the definition of equilibria between opposing forces. The main intuition is that competitive behavior between agents can be modeled as a *game* where a number of predefined *strategies* are available, and a fixed *payoff* is gained by two individuals when they play a given pair of strategies against each other. Evolutionary Game Theory [17] considers a scenario where pairs of individuals, each pre-programmed with a given strategy, are repeatedly drawn from a large population to play a game, and a selection process allows "fit" individuals to thrive, while driving "unfit" ones to extinction. The idea underlying the adoption of Evolutionary Game Theory for selection is to model each hypothesis as a strategy and let them be played one against the other, until a stable population emerges. We will cast this idea within our context by using paths and observations as hypotheses and the consistence of the related triangulations as their mutual payoff.

3.1 Hypotheses

We define $H = \{H_1, H_2, \cdots, H_n\}$ be the set of n available hypotheses. Each hypothesis is a possible triangulation. To this end, it must include exactly two observations and the 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 $(M_{1x} \ldots M_{yi}, O_i^a, M_{1w} \ldots M_{vj}, O_j^b)$. Here $M_{1x} \ldots M_{yi}$ and $M_{1w} \ldots M_{vj}$ are paths that combine a sequence of rigid transformation matrices connecting cameras C_i and C_j to C_1. The accuracy of such paths can vary according to the accuracy of the calibration. O_i^a and O_j^b are two observations of the (hopefully) same physical point on the image planes of camera C_i and C_j. Each hypotesis must include by construction two different cameras, each path must not comprise cycles and must be shorter to a maximum value $maxpath$. Accuracy of observations depends on several sources of noise, usually distributed as a zero-mean Gaussian. In addition, each wrong labelling could result in a totally misplaced point. Given these factors, the correctness of each triangulation depends on both calibration-related and observation-related error sources. It is very difficult, in general, to assess the quality of each hypothesis by looking just at the single 3D point obtained.

3.2 Payoff

As a matter of fact, we are not really interested in the reliability of each hypothesis alone. Rather, our focus is on the definition of a measure of how well two hypotheses play together. This measure is called payoff and should be high if the two hypotheses reliably converge on the same 3D point and low otherwise. The payoff is usually expressed as a 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. Since payoffs are defined

between all the pairs, an alternative notation is the payoff matrix $\Pi = (\pi_{ij})$, where $(\pi_{ij}) = \pi(i, j)$. Our goal is to define a payoff that can account for both the consensus between reconstruction and their reliability. From each hypothesis H_i it is possible to obtain a 3D point by means of triangulation. We call such point $\boldsymbol{x}(H_i)$. Our method is not affected by the actual triangulation technique used, as long as, in addition to the 3D points, it also return a *skewness* value $s(H_i)$ which measures the distance between the two rays used to recover the material point. The position of the point $\boldsymbol{x}(H_i)$ and the triangulation skewness $s(H_i)$ can be used to build a proper payoff function. In fact, two hypotheses can be deemed to be compatible if the associated 3D points are close. Moreover, each one of them can contribute to the overall reliability measure according to its skewness. Albeit these two error sources are not really independent, we can still reasonably approximate them as a bidimensional Gaussian function:

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

Where σ_p and σ_s are two parameters that represent respectively the expected standard deviation of point position and of skewness. Note that $(|\boldsymbol{x}(H_i) - \boldsymbol{x}(H_j)|)^2$ is a pairwise measure that is defined between pairs of hypotheses. 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'.
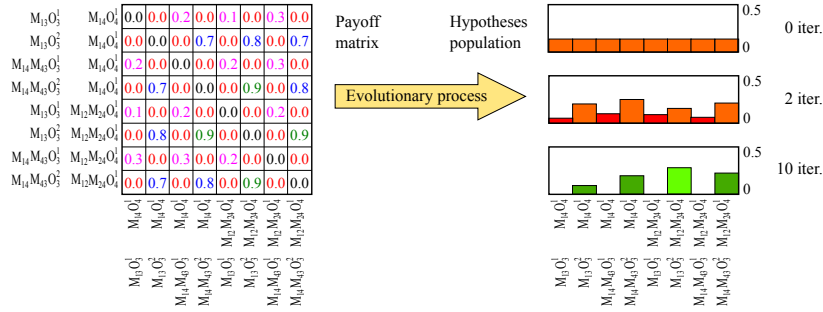
While π' expresses the degree of consensus between hypothesis, we must also account for special cases where two hypotheses are not compatible regardless of the quality of triangulation. This is the case when the hypotheses include two different observations from the same camera. Indeed this is not possible since we are tracking a single material point, thus one of the two observations must be a wrong labelling. The other unfeasible case consists in the presence of two different paths to the same camera. In fact, while different transformations might be optimal for different points, it makes no sense to use two alternative paths to the same camera for the same point, since it would break the reference to the common world. These constraint can be enforced by explicitly setting a value of zero in the final payoff function.

$$\pi(H_i, H_j)) = \pi((P_\alpha^u, O_\alpha^a, P_{\alpha'}^{u'}, O_{\alpha'}^{a'}), (P_\beta^v, O_\beta^b, P_{\beta'}^{v'}, O_{\beta'}^{b'})) \tag{2}$$

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

where P_α^u represents a path connecting camera C_α to the world frame.

3.3 Evolution

With the set of hypotheses and payoff at hand, we can perform the evolutionary process needed to select mutually consistent triangulations. To this end we need to define the concept of *population*. That is a discrete probability

$M_{13}O_3^1$ $M_{14}O_4^1$ | 0.0 | 0.0 | 0.2 | 0.0 | 0.1 | 0.0 | 0.3 | 0.0
$M_{13}O_3^2$ $M_{14}O_4^2$ | 0.0 | 0.0 | 0.0 | 0.7 | 0.0 | 0.8 | 0.0 | 0.7
$M_{14}M_{48}O_5^1$ $M_{14}O_4^3$ | 0.2 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | 0.3 | 0.0
$M_{14}M_{48}O_5^2$ $M_{14}O_4^4$ | 0.0 | 0.7 | 0.0 | 0.0 | 0.0 | 0.9 | 0.0 | 0.8
$M_{13}O_5^1$ $M_{12}M_{34}O_4^1$ | 0.1 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0
$M_{13}O_5^2$ $M_{12}M_{34}O_4^2$ | 0.0 | 0.8 | 0.0 | 0.9 | 0.0 | 0.0 | 0.0 | 0.9
$M_{14}M_{48}O_5^1$ $M_{12}M_{34}O_4^3$ | 0.3 | 0.0 | 0.3 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0
$M_{14}M_{48}O_5^2$ $M_{12}M_{34}O_4^4$ | 0.0 | 0.7 | 0.0 | 0.8 | 0.0 | 0.9 | 0.0 | 0.0

Payoff matrix — Hypotheses population — Evolutionary process

0 iter. / 2 iter. / 10 iter.

Fig. 2. Example of the selection process applied to the instance of the problem shown in Fig. 1. Please note that this is a very simplified example with a non perfectly accurate payoff matrix (albeit the sown evolution has been computed accurately). See the text for an explanation of this specific example.

distribution $\boldsymbol{x} = (x_1, \ldots, x_n)^T$ over the available strategies H. Any population vector is bound to lie within the n-dimensional standard simplex $\Delta^n = \{\boldsymbol{x} \in \mathbb{R}^n : x_i \geq 0 \text{ for all } i \in 1 \ldots n, \sum_{i=1}^{n} x_i = 1\}$ The *support* of a population $\boldsymbol{x} \in \Delta^n$, denoted by $\sigma(\boldsymbol{x})$, is defined as the set of elements chosen with non-zero probability: $\sigma(\boldsymbol{x}) = \{i \in O \mid x_i > 0\}$. In order to find a set of mutually coherent hypotheses, we are interested in finding configurations of the population maximizing the average payoff. Since the total payoff obtained by hypothesis i within a given population $\boldsymbol{x}$ is $(\Pi\boldsymbol{x})_i = \sum_j \pi_{ij}x_j$, the (weighted) average payoff over all the considered hypotheses is exactly $\boldsymbol{x}^T \Pi \boldsymbol{x}$. Unfortunately, it is not immediate to find the global maximum for $\boldsymbol{x}^T \Pi \boldsymbol{x}$, 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 $\boldsymbol{x}$ near the barycenter of the simplex and then proceeds by evolving it through the discrete-time replicator dynamic [18]:

$$x_i(t+1) = x_i(t)\frac{(\Pi\boldsymbol{x}(t))_i}{\boldsymbol{x}(t)^T \Pi \boldsymbol{x}(t)} \tag{3}$$

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 the constraints expressed in equation (2) are actually guaranteed to be enforced. This approach, while novel within the scenario covered by this paper, has shown to be very successful in addressing a wide range of problems, including feature-based matching [19], medical images segmentation [20] and registration [21], rigid [22] and non-rigid [23] 3D object recognition.

To give to the reader a better understanding of how the process works we illustrate a complete case in Fig. 2. While the example is kept intentionally simple, out of practical reasons, the evolution has been performed actually using the shown payoff matrix and the replication dynamic described by equation (3). Here, we are assuming the netwotk topology and observations shown in Fig. 1.

We are also assuming that all the pairwise extrinsic calibrations have been performed with good accuracy, with the exception of M_{14} which is afflicted by a larger error (for whatever reason). Observations O_3^2 and O_4^1 are correctly labeled, albeit subject to measurement error. Observation O_3^1 is an outlier and results from a wrong labeling. We are considering two cameras (i.e. $C = \{C_3, C_4\}$) and a total of four paths, two for each camera (i.e. $M = \{M_{13}, M_{14}M_{43}, M_{14}, M_{12}M_{24}\}$. This result in a total of eight hypotheses, shown in figure. The payoff matrix is shown with color coded entries. Namely we used red to highlight entries that are zero due to the constraints expressed by equation (2). For instance hypotheses $(M_{13}O_3^1, M_{14}O_4^1)$ and $(M_{13}O_3^2, M_{14}O_4^1)$ are not compatible since they include different observations from the same camera. Differently, hypotheses $(M_{13}O_3^1, M_{14}O_4^1)$ and $(M_{13}O_3^1, M_{12}M_{24}O_4^1)$ have zero mutual consensus since they are connecting camera C_4 to the world frame through two different paths. Hypotheses that receive low mutual payoff due to geometrical inconsistencies are highlighted in purple. Basically, they include all the pairs afflicted by the wrong labeling of O_3^1. Other pairwise payoffs are assigned according to the coherence between reconstructions. Blue payoffs are slightly lower than green ones due to the fact that they include the rigid transformation M_{14}, which we assumed to be less accurate. In the right part of the figure we show the actual evolution. The process starts from an uniform distribution, where every hypothesis has the same probability to thrive. After just 2 iterations of equation (3), hypotheses related to outliers start to decrease, due to they low average payoff. After just 10 iteration, only feasible hypotheses survived. Note that paths including the slightly inaccurate transformation M_{14} are less represented in the final population. Indeed, the final distribution can be used to produce a weighted average of the 3D point, rather than just selecting the hypothesis with higher score.

4   Experimental evaluation

To test our proposed optimal path selection approach we designed a set of synthetic camera networks resembling typical real-world topologies. Testing with a synthetic case allows us to factorize out all the unpredictable error sources and properly analyze the behavior with respect to erroneous observations and inaccurate positioning of the cameras.

We started by generating 3 different network topologies, namely: grid, hemisphere and line. The first is composed by a regular grid of 15 cameras spanning an area of 20×12 centered at the origin and lying onto the xy-plane (Fig 3, left). All the cameras were rotated so that the z-axis points toward the network center $\boldsymbol{c} = (0, 0, 10)^T$. We generated the exact relative motion between couple of cameras whose distance is less than 10, for a total of 88 graph edges. The second (hemisphere) is composed by 16 cameras disposed in a semi-sphere with center $\boldsymbol{c}$ and radius $r = 10$ (Fig 3, center). We placed the cameras reference frame origins with an uniformly distributed angular azimuth and elevation and pointed all the optical axes toward the center. Also in this case, a total of 90 edges describe the relative motions of camera pairs that are less than 10 units apart. Finally, we generated the line network topology as a set of 9 cameras lying
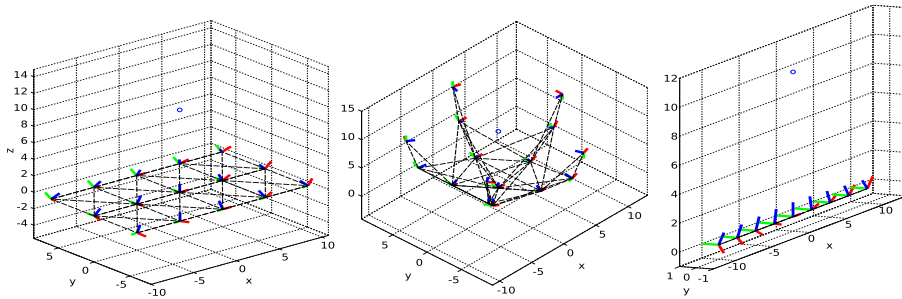
Fig. 3. The 3 generated graph topologies used for experimental evaluation: grid (left), hemisphere (center) and line (right).

on the x-axis, uniformly spaced around the origin and oriented toward c (Fig 3, right). A set of 30 edges links together adjacent cameras. For all the three topologies, camera intrinsic parameters were set with unitary focal length and principal point lying at the origin. Furthermore, we placed the first camera C_1 so that its frame coincides with the world reference frame. From this ground truth, many different perturbed instances of such topologies were generated by means of a normally-distributed additive angular error (with zero mean and standard deviation σ_r) to the rotation matrix of rigid motions associated to each edge.

In our first test we analyze the sensitivity of the proposed method wrt. payoff parameters σ_p and σ_s. In Figure 5 (left) we show the spatial distance, varying σ_p and σ_s, between a reconstructed 3D point and the generated ground truth for a grid topology instance with $\sigma_r = 9.1*10^{-3}$. As expected, there exists a large area around $\sigma_p = 0.08, \sigma_s = 0.12$ in which the interplay between the two parameters leads to satisfactory results. This also give us a clue that the skewness, even if is not a quantity directly measuring the final reconstruction error, can help the effectiveness of the payoff function.

In the second experiment (Fig. 4) we compared the reconstruction accuracy of our approach against "dual-quaternions" [12] and "SBA" [13]. The first method exploits properties of dual-quaternions to diffuse the camera network error and creates a new set of coherent motions to a common reference frame. The latter is commonly used in structure-from-motion applications to simultaneously optimize extrinsic camera parameters and the reconstructed 3D points given their observations on each image plane. Reconstruction error is evaluated for grid, hemisphere and line topologies (top to bottom row) varying the network graph error σ_r (first column) and observation error σ_o (second column) by means of an additive zero-mean Gaussian error with variance σ_o^2. All the tests were performed by reconstructing a structure of 10 points. For [12], we considered the best triangulation in terms of re-projection error between all the graph paths with less than 3 vertices whereas, for [13], all the structure points were optimized at the same time. We can observe that SBA, even if it has the advantage of recomputing the camera poses while triangulating, mostly suffers from the observation errors. On the contrary, our proposed approach dynamically discards incoherent observations producing structures that are less noisy and more reliable. A
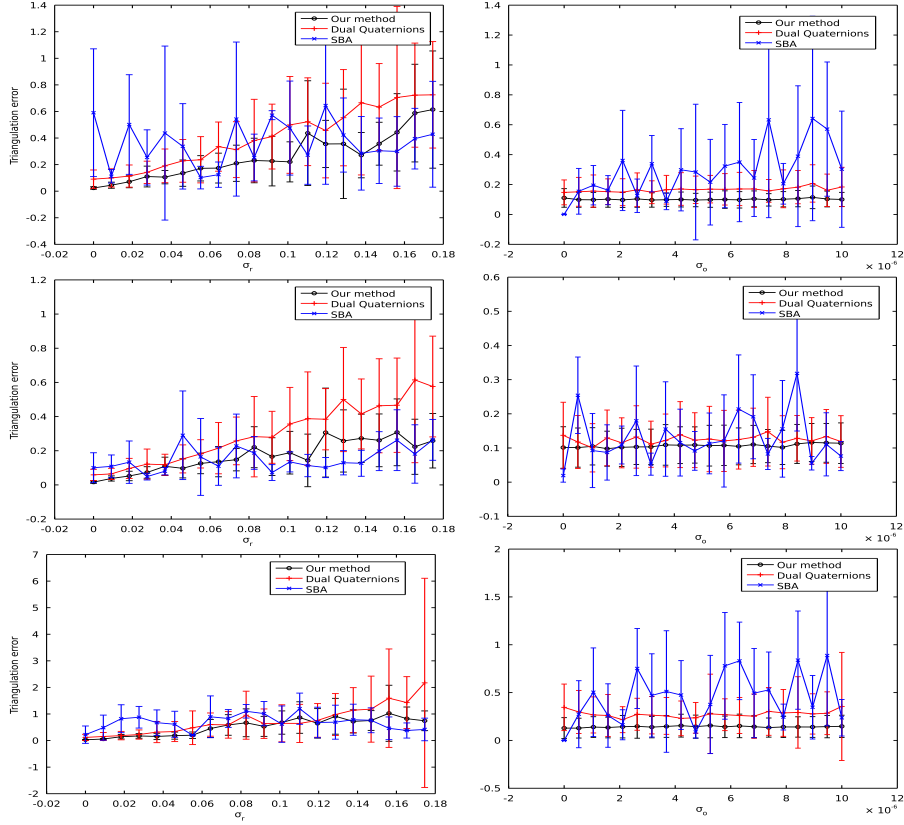
Fig. 4. Comparison between our method, dual-quaternions [12] and SBA [13]. Triangulation error is evaluated while perturbing the graph edges (left col.) and observations (right col.) for grid, hemisphere and line graph topologies (top to bottom).

similar behaviour can be observed for [12] varying the network graph error σ_r. Since [12] can only diffuse errors without discarding any edge, it suffers from large relative motion displacements that may happen during graph calibration. Overall, our method can deliver the best of the two worlds, being either able to smooth the errors by averaging the triangulation among many cameras while still being very effective when selecting a small set of reliable paths and observations. Finally, we tested the behaviour of our method in case of observation outliers. To this extent, we computed the number of false positives (wrong observations still involved in the point triangulation) and false negatives (number of good observations wrongly discarded) in the winning population for many different triangulation attempts varying the observations outlier distribution for the grid network topology. Specifically, we generated exactly one inlier observation (with a random uniform uncertainty of $\sigma_o = 10^{-3}$) and one outlier displaced from the ground truth observation by $K\sigma_o$. The more K is increased, the more the distance on the image plane between inlier and outlier observations increase. As
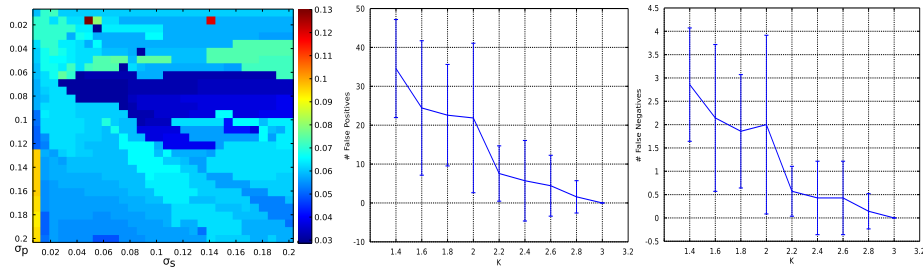
Fig. 5. Left: triangulation error varying σ_p and σ_s. Right: false positives and negatives wrt. inlier/outlier distance ratio K.

we can observe from Figure 5 (right), both the number of false positive and false negatives decrease proportionally with K. With just 3 times σ_o, the number of wrongly selected observations is almost 0. In a real-case scenario, we expect wrong observations being either close to the correct one, so that they should not hinder the triangulation, or more than 3 times the standard deviation apart and thus discarded from the winning population.

5 Conclusions and Future Work

In this paper we introduced a dynamic path selection method that can be used to perform robust 3D reconstructions when using multiple cameras characterized by inaccurate extrinsic calibrations. Reconstruction is performed point-wise by dynamically selecting the best possible set of camera poses and observations that maximize the consistency of each pairwise triangulation. The ability to locally exclude part of the graph or observations from the final triangulation makes our method particularly effective in all the scenarios when either a precise calibration or a reliable localization of the target cannot be provided. Finally, it should be noted that the presented study is preliminary and our future goal is to complete it with respect to both theoretical aspects and evaluation. From a theoretical standpoint, the current solution works well only when the cameras are well overlapped. Suitable techniques should be introduced to obtain a transitive closure throughout a long sequence of partially overlapped cameras. From a validation point of view, we are presenting an entirely synthetic test set. An evaluation using real-world data [24] could give a better insight about the practical merits and limitations of the method.

References

1. Lee, K.H., Chu, C.T., Lee, Y., Fang, Z., Hwang, J.N.: Consistent human tracking over self-organized and scalable multiple-camera networks. In: Distributed Embedded Smart Cameras. Springer (2014) 189–209
2. Raman, R., Sa, P.K., Majhi, B.: Occlusion prediction algorithms for multi-camera network. In: Distributed Smart Cameras (ICDSC), 2012 Sixth International Conference on, IEEE (2012) 1–6

3. Hoppe, C., Wendel, A., Zollmann, S., Pirker, K., Irschara, A., Bischof, H., Kluckner, S.: Photogrammetric camera network design for micro aerial vehicles. In: Computer vision winter workshop (CVWW). Volume 8. (2012) 1–3

4. Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S.M., Szeliski, R.: Building rome in a day. Communications of the ACM **54** (2011) 105–112

5. Ramagiri, S., Kavi, R., Kulathumani, V.: Real-time multi-view human action recognition using a wireless camera network. In: Distributed Smart Cameras (ICDSC), IEEE (2011) 1–6

6. Wang, X.: Intelligent multi-camera video surveillance: A review. Pattern recognition letters **34** (2013) 3–19

7. Morye, A., Ding, C., Song, B., Roy-Chowdhury, A., Farrell, J.: Optimized imaging and target tracking within a distributed camera network. In: American Control Conference (ACC), 2011, IEEE (2011) 474–480

8. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multi-view stereopsis. IEEE Trans. on Pattern Analysis and Machine Intelligence **32** (2010) 1362–1376

9. Alsadik, B., Gerke, M., Vosselman, G.: Optimal camera network design for 3d modeling of cultural heritage. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **3** (2012)

10. Zhang, Z.: A flexible new technique for camera calibration. Pattern Analysis and Machine Intelligence **22** (2000) 1330–1334

11. Heikkila, J., Silvén, O.: A four-step camera calibration procedure with implicit image correction. In: CVPR, IEEE (1997) 1106–1112

12. Torsello, A., Rodola, E., Albarelli, A.: Multiview registration via graph diffusion of dual quaternions. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE (2011) 2441–2448

13. Lourakis, M.I., Argyros, A.A.: Sba: A software package for generic sparse bundle adjustment. ACM Transactions on Mathematical Software (TOMS) **36** (2009) 2

14. Svoboda, T., Martinec, D., Pajdla, T.: A convenient multicamera self-calibration for virtual environments. Teleoperators and virt. environments **14** (2005) 407–422

15. Barreto, J., Daniilidis, K.: Wide area multiple camera calibration and estimation of radial distortion. In: Proceedings of the 5th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Prague, Czech Republic. (2004)

16. Nash, J.: Non-cooperative games. Annals of mathematics (1951) 286–295

17. Weibull, J.: Evolutionary Game Theory. MIT Press (1995)

18. Taylor, P., Jonker, L.: Evolutionarily stable strategies and game dynamics. Mathematical Biosciences **40** (1978) 145–156

19. Albarelli, A., Rota Bulò, S., Torsello, A., Pelillo, M.: Matching as a non-cooperative game. In: Proc. ICCV. (2009) 1319–1326

20. Ibragimov, B., Likar, B., Pernus, F., Vrtovec, T.: A game-theoretic framework for landmark-based image segmentation. Medical Imaging **31** (2012) 1761–1776

21. Banerjee, J., Klink, C., Peters, E.D., Niessen, W.J., Moelker, A.: 4d liver ultrasound registration. In: Biomedical Image Registration. (2014) 194–202

22. Albarelli, A., Rodolá, E., Bergamasco, F., Torsello, A.: A non-cooperative game for 3d object recognition in cluttered scenes. In: 2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 3DIMPVT 2011. (2011) 252–259

23. Rodolà, E., Bronstein, A., Albarelli, A., Bergamasco, F., Torsello, A.: A game-theoretic approach to deformable shape matching. In: CVPR. (2012) 182–189

24. Joo, H., Park, H.S., Sheikh, Y.: Map visibility estimation for large-scale dynamic 3d reconstruction. (2014)