

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221703096>

Reconstruction error in a motion capture system

Article · March 2012

Source: arXiv

CITATIONS

3

READS

394

2 authors:



[Andrea Masiero](#)

University of Florence

107 PUBLICATIONS 818 CITATIONS

SEE PROFILE



[Angelo Cenedese](#)

University of Padova

202 PUBLICATIONS 1,943 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Modeling and Analysis of Brain Network from a Control Theory Point of View [View project](#)

Reconstruction error in a motion capture system

Andrea Masiero and Angelo Cenedese

Abstract—Marker-based motion capture (MoCap) systems can be composed by several dozens of cameras with the purpose of reconstructing the trajectories of hundreds of targets. With a large amount of cameras it becomes interesting to determine the optimal reconstruction strategy. For such aim it is of fundamental importance to understand the information provided by different camera measurements and how they are combined, i.e. how the reconstruction error changes by considering different cameras. In this work, first, an approximation of the reconstruction error variance is derived. The results obtained in some simulations suggest that the proposed strategy allows to obtain a good approximation of the real error variance with significant reduction of the computational time.

I. INTRODUCTION

Nowadays marker-based motion capture (MoCap) systems can be composed by several dozens of cameras with the purpose of reconstructing the trajectories of hundreds of targets. However, as costs of modern microprocessor and camera hardware decrease, it becomes economically viable to consider MoCap systems made of large camera networks of several hundreds of cameras, meeting the growing request for higher precision reconstruction of larger scenarios. This requirement, in terms of both minimizing the single target estimation error and of increasing the quality in the scene description, translates into scaling up with both the number of markers and the number of cameras.

The MoCap task can typically be divided into two steps: Reconstructing the 3D target positions by means of the measurements at time t , and merging such reconstructions with the dynamic evolutions of previously detected targets (data association and tracking). This paper focuses on the first step. If the system is composed by a limited number of cameras and targets, the classical reconstruction algorithm based on geometric triangulation [7], [8], [16] can be implemented in a centralized fashion on a single machine to track the targets in real time. On the other hand, when considering the envisaged large system scenarios, it becomes difficult to simultaneously take into account the data provided by all the cameras. So, first, only portions of the system are considered simultaneously, and then the 3D reconstruction is achieved by progressively merging data from different parts of the system. In this framework it is important how the information

is elaborated and merged by different cameras, i.e. some pairs of cameras will allow a better reconstruction¹ with respect to others.

This work deals with the problem of determining the information provided by different cameras about a target, and, consequently, what are the cameras that allow the optimal reconstruction of the investigated targets.

Even if the problem here is formulated in the MoCap framework, actually it is closely related also to other areas in computer vision: In the structure from motion framework [13], a quality measure among tensors is derived in [14] for reconstruction based on a hierarchical computational structure of trifocal tensors. Furthermore, in the multi-view stereo context, [6] and [5] used suitable “affinity” functions to properly select a set of “optimal” views.

II. RECONSTRUCTION ERROR STATISTICAL DESCRIPTION

Because of noise and discrete measurements the ray associated by a camera to a target’s position actually represents only an estimation of the “mean” direction along which the target is positioned. The uncertainty on the 3D position provided by such mean direction grows as the distance between the viewing camera and the target increases (and it is parallel to the sensor plane, which practically is approximatively orthogonal to the ray direction).

The reconstruction error using two cameras changes depending on cameras’ positions and orientations: Cameras at orthogonal orientations typically provides reconstructions with the smallest estimation errors, and conversely for cameras viewing along the same direction. However, because of the different views and of occlusions (both due to other objects in the scene and to the targets’ object themselves), cameras at very different positions and orientations usually retrieve measurements of only few common targets.

In this section, the uncertainties on single and multi-camera reconstructions are presented in detail.

A. Single camera measurements

Consider a target i placed at $\bar{\phi}_i = (\bar{x}_i, \bar{y}_i, \bar{z}_i)^\top$ in the 3D space. The noise of the target on the j -th camera measurement $\xi_{ij} = (u_{ij}, v_{ij})^\top$ is assumed to be additive and Gaussian [7], [8]:

$$\xi_{ij} = \bar{\xi}_{ij} + e_{ij} \quad , \quad (1)$$

where $\bar{\xi}_{ij} = (\bar{u}_{ij}, \bar{v}_{ij})^\top$ is the measurement without noise and $e_{ij} \sim \mathcal{N}(0, \Sigma_{e_{ij}})$ is the measurement noise. Hereafter

A.Masiero A.Cenedese are with the Dipartimento di Ingegneria dell’Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy masiero@dei.unipd.it, angelo.cenedese@unipd.it.

The research leading to these results has received funding from the European Community’s Seventh Framework Programme under agreement n. FP7-ICT-223866 FeedNetBack and n. 257462 HYCON2 Network of excellence. This activity contributes to the seed project R3D of the Department of Information Engineering - University of Padova.

¹The concept of reconstruction quality considered here wants to take into account of several factors, among them: The number of reconstructed targets, the reconstruction accuracy, and the required computational time.

the noise variance matrix $\Sigma_{e_{ij}}$ is modeled as $\text{diag}(\sigma_{e_{ij}}^2, \sigma_{e_{ij}}^2)$, where $\sigma_{e_{ij}}$ is the standard deviation. Note that $\sigma_{e_{ij}}$ typically depends on camera and target reciprocal positions (and on camera orientation). In addition, the value of $\sigma_{e_{ij}}$ depends also on the image analysis algorithm used for detecting it. Since the complete coverage of this topic is out of the scope of this work, hereafter the value of $\sigma_{e_{ij}}$ is taken as known.

Each measurement from camera j is a point on its sensor that corresponds to a ray passing through such point and camera's optical center, as shown in Fig. 1.

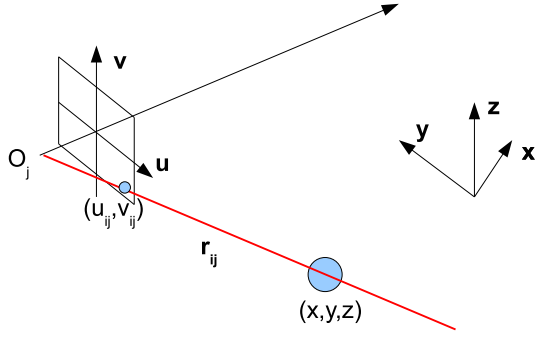


Fig. 1. Red line: ray associated to measurement (u_{ij}, v_{ij}) of target i in camera j .

If different information are not available (e.g. the size of the target), target's 3D position cannot be reconstructed using a single measurement. However, it is possible to reconstruct by means of geometric triangulation using at least two measurements, as shown in Fig. 2.

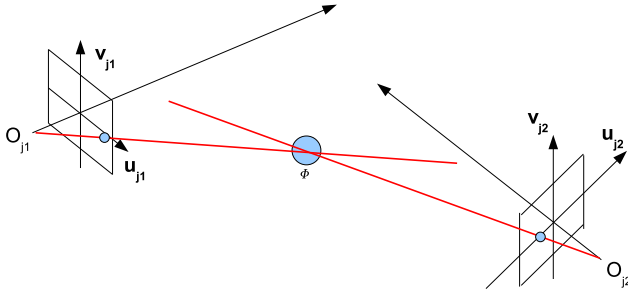


Fig. 2. Triangulation between two cameras. Crossing point between rays determined by different cameras allows to obtain target's 3D position.

Let the plane \mathcal{P}_{ij} be parallel to the image plane I_j of camera j and passing through the target i . Because of the measurement noise e_{ij} the ray associated to target i by camera j will intersect with plane \mathcal{P}_{ij} on a point $\phi_i \neq \bar{\phi}_i$. Let $e'_{ij} = \phi_i - \bar{\phi}_i$: e'_{ij} is obtained by propagation of error e_{ij} according to:

$$f e'_{ij} = f' e_{ij}, \quad (2)$$

where f is the camera focal length and f' is the distance from I_j to \mathcal{P}_{ij} . Actually, the above equation holds for all planes \mathcal{P}_j (on the front side of camera j) parallel to I_j (see Fig. 3).

While the measurement error propagates on the \mathcal{P}_j plane as described, the measurement does not provide any infor-

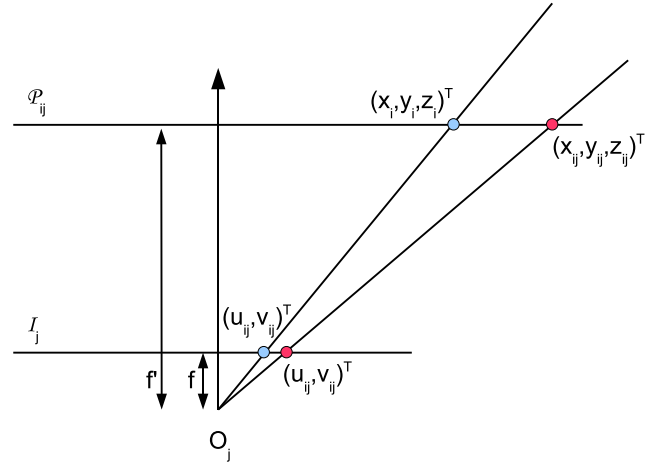


Fig. 3. Propagation of camera measurement error.

mation about the target position along the line starting from the optical center O_j and passing through ξ_{ij} .

Exploiting different measurements of the same target allows to obtain a good estimation of the real distance f' , therefore by combining several camera measurements the information provided by camera j about target i can be modeled as:

$$\phi_{ij} \sim \mathcal{N}(\bar{\phi}_i, \Sigma_{ij}), \quad (3)$$

where

$$\Sigma_{ij} = M \psi_{ij} \psi_{ij}^T + \sigma_{e_{ij}}^2 \left(\frac{f'}{f} \right)^2 \Psi_{ij} \Psi_{ij}^T, \quad (4)$$

and M is a number much larger than the maximum room's side size multiplied by m , ψ_{ij} is the unit vector along the direction from O_j to $\bar{\phi}_i$, and Ψ_{ij} is an orthonormal basis of the plane \mathcal{P}_{ij} parallel to I_j and passing through the origin. Since M is very large, the first term in (4) expresses the practical absence of information provided by camera j about target i along the ψ_{ij} direction, i.e. along the direction of the line from O_j to the point. Instead, the second term corresponds to the variance of the measurement error propagated using (2) to the plane \mathcal{P}_{ij} . We stress the fact that the approximation of the reconstruction error variance (4) is good in nonsingular conditions, i.e. when the target position can be adequately reconstructed (which is a typical operating condition when using a large number of cameras). An experimental proof of the goodness of the approximation obtained by (4) in the framework of multiple-cameras reconstruction is given by the simulations in the Subsec. II-B.

B. Multiple camera reconstruction

The approximation of equation (3) is particularly useful when combining measurements from different cameras.

Without loss of generality, consider the reconstruction of the position of target i from the measurements of cameras $j = 1, 2, \dots, m_i$. When at least two non aligned measurements are available, the position of the target can be estimated by means of geometric triangulation. Then, f' in (2) is approximatively known, and (3) is a good

approximation of the information provided by each camera j among those available for the reconstruction of target i . Thus, from (3) the uncertainty on the reconstructed position can be approximated as follows (minimum variance estimation, [10]):

$$\Sigma_i = \left(\sum_{j=1}^{m_i} \Sigma_{ij}^{-1} \right)^{-1}, \quad (5)$$

and the overall standard deviation of the reconstruction error can be estimated as $\sqrt{\text{trace}(\Sigma_i)}$.

For comparison, a direct evaluation of the reconstruction error variance can be obtained as the sample reconstruction variance in a Monte Carlo (MC) simulation:

$$\hat{\Sigma}_i = \frac{1}{N-1} \sum_{k=1}^N e_{i,k} e_{i,k}^\top, \quad (6)$$

where N is a large integer number, and $e_{i,k}$ is the reconstruction error (difference between true and reconstructed position) in the MC iteration k . Fig. 4 shows the percent error between the sample reconstruction standard deviation (6) and that computed from approximation (5) varying the number of cameras m considered for reconstruction from 2 to 256 (cameras are equally spaced along a circle of 10 m radius). The reported values are the mean of the results obtained on 1000 randomly sampled points (all positioned in the volume delimited by the cameras) for each choice of m . At each iteration, the m cameras used for the reconstruction are randomly selected among the 256 available.

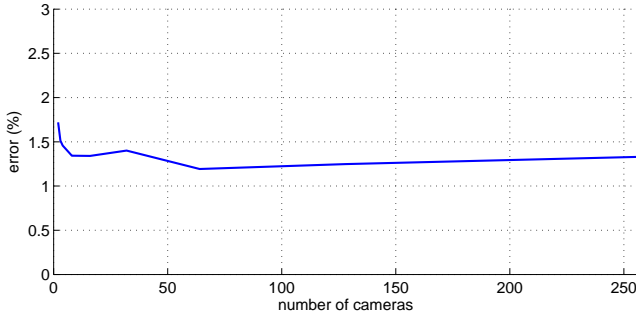


Fig. 4. Percent difference between the standard deviation computed from samples (6) and the approximated theoretical one (5).

In Fig. 5, it is highlighted how the reconstruction error (using two cameras) depends on the angle between the cameras: The closer the angle is to $\pi/2$ the better the estimated position results. In this example, 16 cameras are positioned (equally spaced) on a circle of 10 m radius. The reconstruction error is computed for the point in the center of the cameras' circle. As shown in Fig. 5, the 1σ -level curve computed from (5) is practically overlapped to the 1σ -level curve estimated by sample data.

The performance evaluation of a MoCap system typically requires to compute the reconstruction error on a (quite large) representative number of points (voxels). Since the MC variance estimation can be quite time demanding, it is

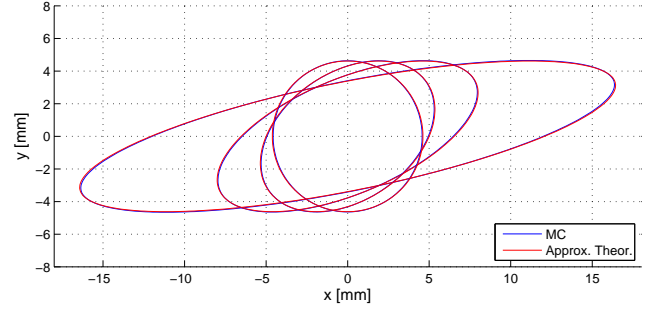


Fig. 5. Comparison of 1σ -level curve of the reconstruction error: Variance obtained by MC simulation (blue) and theoretical approximation (red). The error is evaluated for different angles between the two cameras: $\pi/8$ (which is represented by the external ellipse), $\pi/4$, $3\pi/8$, $\pi/2$ (small circle inside the other curves).

worth to consider (5) that allows to compute in closed form good approximations (as in Figs. 4-5), at a computational cost largely lower than using the MC method.

III. CONCLUSIONS

In this work, an approximation of the reconstruction error variance has been derived for marker-based motion capture system. Such approximation can be useful in deriving the optimal strategy for pairing cameras to reduce the reconstruction computational time in a distributed approach.

REFERENCES

- [1] H. Aghajan and A. Cavallaro. *Multi-Camera Networks, Principles and Applications*. Academic Press, 2009.
- [2] E. Borovikov and L. Davis. A distributed system for real-time volume reconstruction. pages 183–189, In *CAMP*, 2000.
- [3] J. Falcou, J. S  rot, T. Chateau, and F. Jurie. A parallel implementation of a 3d reconstruction algorithm for real-time vision. pages 663–670, In *ParCo*, 2005.
- [4] J.S. Franco, C. M  nier, E. Boyer, and B. Raffin. A distributed approach for real-time 3d modeling. In *CVPRW*, 2004.
- [5] Y. Furukawa, B. Curless, S.M. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *CVPR*, 2010.
- [6] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S.M. Seitz. Multi-view stereo for community photo collections. In *ICCV*, 2007.
- [7] R.I. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.
- [8] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge Univ. Press, 2003.
- [9] Milan Jovanovic, Andreas Klausner, Markus Quaritsch, Bernhard Rinner, and Allan Tengg. Smart cameras for embedded vision. *Telematik*, 12(3):14–19, 2006.
- [10] T. Kailath, A.H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice-Hall, 2000.
- [11] Andreas Klausner, Allan Tengg, and Bernhard Rinner. Distributed multi-level data fusion for networked embedded systems. *J. on Selected Topics in Signal Processing*, 2(3):538–555, 2008.
- [12] Shubao Liu, Kongbin Kang, Jean-Philippe Tarel, and David B. Cooper. Distributed volumetric scene geometry reconstruction with a network of distributed smart cameras. pages 2334–2341, In *CVPR*, 2009.
- [13] Y. Ma, S. Soatto, J. Kosecka, and S.S. Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer, 2004.
- [14] D. Nister. Reconstruction from uncalibrated sequences with a hierarchy of trifocal tensors. pages 649–663, In *ECCV*, 2000.
- [15] Bernhard Rinner and Wayne Wolf. An introduction to distributed smart cameras. *Proc. of the IEEE*, 96(10):1565–1575, 2008.
- [16] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. *Bundle Adjustment - A modern Synthesis, Springer Lecture Notes on Computer Science*. Springer Verlag, 2000.