ROSE 2007 – IEEE International Workshop on
Robotic and Sensors Environments
Ottawa – Canada, 12-13 October 2007

Flexible Multi-Camera Network Calibration for Human Gesture Monitoring

Silvain Bériault[1], Pierre Payeur[1], Gilles Comeau[2]
[1]School of Information Technology and Engineering
[2]Department of Music
University of Ottawa
Ottawa, Ontario, Canada, K1N 6N5
[sberi081, ppayeur]@site.uottawa.ca
gcomeau@uottawa.ca

Abstract *– This paper presents the design and implementation of a flexible and easy-to-use multi-camera acquisition setup for markerless human gesture monitoring in unconstrained environments. A robust 2-stage framework is proposed to achieve full calibration of a variable number of synchronized cameras separated by long baselines. In the first stage, the intrinsic parameters are computed for each camera independently. In the second stage, the cameras are registered based on their relative positioning by waving a red light emitting device to produce a set of feature points. Matches are regrouped by camera pair such that pair-wise stereo relations can be found for as many pairs as possible before being scaled to create a consistent weighted camera graph which is used to link all cameras. Experimental results demonstrate the accuracy of the calibration that is achieved and the suitability of the proposed approach for almost any multi-camera configurations. An application is presented for volumetric reconstruction of human beings to validate the implementation.*

Keywords *– Camera Calibration, Multi-Camera Networks, Bundle Adjustment, Motion Capture, 3D Reconstruction.*

I. INTRODUCTION

With rising interest in multi-camera applications for 3D computer vision, the purpose of registering, with high precision, multiple viewpoints becomes more and more critical. This process is commonly referred to as multi-camera calibration and is a major issue for many 3D computer vision algorithms, such as shape-from-silhouette volumetric reconstruction, which is often used in applications dedicated to 3D human motion capture. Such applications typically require a complex multi-camera setup composed of at least 5 to 10 cameras. Furthermore, the cameras must be distributed in the working volume to surround the human subject such that motion happening frontward, backward and sideward can be captured. As a result, cameras are separated by large baselines and point in very different orientations.

With such complex camera positioning, accurate and easy camera registration becomes a challenging task. Complex calibration rigs are often proposed [1, 2, 3] but several important drawbacks can be raised. Caillette [1] proposed to use a single 2D calibration rig to perform the calibration of a multi-camera system for human gesture monitoring. While this calibration approach is simple, it is prone to inaccuracies because all calibration points are co-planar. It also requires all cameras to see the full calibration rig, which imposes major constraints on camera positioning. Moreover, during calibration, the working volume must be completely empty as the calibration rig will occupy most of the floor. In order to overcome the coplanarity constraint, Rander [3] suggested to use calibration bars mounted on tripods. The bars are translated vertically and horizontally in the working volume to obtain full coverage. This approach is however very cumbersome as it requires numerous manipulations. It also imposes an empty working volume. Drouin *et al.* [2] suggest a pair-wise camera calibration using multiple views of a 2D calibration rig. This approach also eliminates the coplanarity limitation but lacks flexibility for cases where cameras are orthogonal or are separated by large baselines.

Because of important problems with classical calibration, new approaches were recently investigated [4, 5, 6]. The strategy in all of these approaches is to first find a coarse estimate of the cameras registration which is then refined through an iterative method. To obtain the initial estimate, all of these methods use a single visible feature point as the calibration target. The latter is being waved in the workspace to create a virtual 3D calibration object. Pair-wise relationships between cameras are estimated and then relations linking all cameras to a reference camera are found. Chen *et al.* [4] apply an extended Kalman filter iteratively to perform the final optimization. However, more accurate results are obtained by Ihrke *et al.* [5] and by Svoboda *et al.* [6] who use a bundle adjustment [7] as the final optimization step.

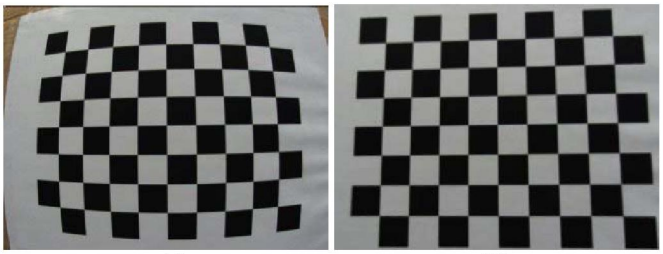
This paper presents a full multi-camera calibration procedure, which is also based on a coarse to fine parameter estimation using bundle adjustment optimization. The framework aims at providing a scalable and easy-to use procedure that can be performed by non-experts, and that imposes minimal constraint on the camera positioning. Accurate final calibration must also be achieved even when cameras suffer from high lens distortion. The following sections discuss these aspects. In section II, an overview of the complete camera calibration scheme is provided. Section III details the estimation of the extrinsic camera parameters. In section IV, the performance of the calibration procedure is analyzed and an example of 3D reconstruction in human motion capture is presented.

II. CAMERA CALIBRATION SCHEME

As for the majority of existing multi-camera calibration methods, the proposed approach is executed in two stages. Initially, the intrinsic parameters and lens distortion are estimated for each camera separately. This operation is performed only once for each camera and does not typically need to be repeated frequently assuming the use of fixed focal length lenses. In the second stage, all cameras are positioned in their final configuration around the working environment and are registered together. This is referred to as extrinsic camera calibration. These parameters remain constant until a change in the position or orientation of one or more cameras occur.

A. Intrinsic Camera Calibration

The problem of finding the intrinsic camera parameters and the distortion coefficients has been widely covered in the literature [8, 9, 10]. For our application, we used an 8x10 checkerboard calibration rig with 2cm x 2cm cell dimensions, as shown in Fig. 1. This type of pattern was selected because it is well integrated with the classical calibration schemes of Tsai [9] and Zhang [10] and there are implementations available in computer vision libraries such as OpenCV [11]. To facilitate the intrinsic calibration, we developed an application which captures frames of the checkerboard from streaming video. Our experimentation revealed that 20 views are sufficient to obtain a good estimate of the intrinsic parameters and of the distortion coefficients. Proper distortion correction is necessary to prevent destructive effect in the camera registration and 3D reconstruction, especially when dealing with wide angle lenses, such as demonstrated in Fig 1. Therefore both radial and tangential distortions are modeled.



(a) (b)

Fig. 1.Distortion compensation of a wide angle lens.
a) A checkerboard pattern in the distorted image plane.
b) The same checkerboard pattern in the undistorted image plane.

B. Extrinsic Camera Calibration

The estimation of extrinsic parameters is not as straightforward given the large baselines and wide angular variation between the cameras. The literature has raised several issues regarding the extrinsic calibration of a multi-camera system. Our proposed approach takes advantage of the bundle adjustment technique [7] which is well known to output very accurate calibration provided a good initial estimate of each camera's position and orientation. However, finding an initial estimate which is reliable is mandatory to ensure proper bundle adjustment convergence. Therefore, the core of the proposed procedure is to find such an initial estimate robustly while respecting the requirements imposed by the application of markerless human motion capture in unconstrained environments. These constraints summarize as: *i)* ease-of-use: the proposed method should be fully automated and should require minimal human manipulation, especially regarding manual measurements in the working volume; *ii)* straightforward: the procedure should be easy to repeat since extrinsic parameters need to be recomputed each time there is a change in the camera positioning; *iii)* scalability: the method must be scalable in terms of the number of cameras in the network as well as the dimensions of the working volume; and *iv)* free camera positioning: no restriction should be imposed on the camera positioning apart from the requirement of a sufficient overlap to achieve matching between viewpoints, as required for 3D reconstruction of the subject under observation.

III. EXTRINSIC CALIBRATION FRAMEWORK

The proposed framework counts on seven major steps to achieve complete and accurate estimation of the extrinsic parameters. First, matching feature points are collected across the entire camera network. Using those matches, a pair-wise fundamental matrix is computed for as many pairs of cameras as possible. Each fundamental matrix is then decomposed to extract a stereo rotation matrix and a unit translation vector. Translation magnitude is then found for each pair of cameras to define a consistent camera network. The next step consists of unifying the cameras extrinsic parameters from a weighted graph of cameras. The former are then optimized using a bundle adjustment procedure. Finally, the camera network is rescaled to absolute dimensions.

A. Acquiring Matches

From a classical point of view, a match is defined as a correspondence between a measured point in the 3D working volume and the pixel coordinates of that point projected on an image plane. However, performing accurate 3D measurements manually or constructing complex calibration rigs is cumbersome. The concept of a virtual calibration object offers numerous benefits [4, 5, 6]. Several calibration points can be acquired by waving a small visible marker, over the full working volume. Here a light emitting device (LED) mounted at the extremity of a stick is waved randomly in space to generate a clear blob when the room lighting is turned off. Since the 3D position of the target is not known in absolute coordinates, image correspondences between the cameras are being recorded instead. Therefore, according to this new definition, a match occurs when a point is seen simultaneously by two or more cameras. Following this

procedure, with full coverage of the working volume, over a thousand of matches can be obtained in less than 3 minutes of video recording. The procedure however requires cameras to be synchronized, as will be described in section IV. Another requirement is that the lighting level must be controlled during the calibration procedure. An alternative calibration target to resolve this issue is proposed by Ihrke *et al.* [5] for outdoor applications.

B. Pair-wise Fundamental Matrix Computation

Once a sufficient number of matches are found, they are re-organized by pair of cameras and the corresponding fundamental matrix is computed. To ensure a reliable estimate, the fundamental matrix is estimated only if at least 30 matches are found between a given pair of cameras. A RANSAC implementation [11, 12] is used to eliminate outliers which are also removed from the global database of matches for later steps. Unlike Svoboda *et al.* [6], matches are recorded in the undistorted image plane, therefore a stricter outlier rejection threshold can be used in the RANSAC analysis. Thus, any match is identified as an outlier if it has a distance from point to epipolar line above 2 pixels. It potentially allows a better estimate of the fundamental matrix since inaccurate matches are rejected from the computation.

When computing the fundamental matrix from a set of image matches, difficulties may occur from the use of degenerative configurations [13]. However those situations are easy to avoid or almost unlikely to occur [6]. Indeed, in a human motion capture system, cameras need to be separated by large baselines, discarding the problem of cameras sharing the same optical center. Furthermore, the fact that the full working volume is covered in a random manner with the LED calibration stick significantly reduces the risk to find all points over a co-planar or a ruled-quadratic distribution [13].

C. Fundamental Matrix Decomposition

Each fundamental matrix needs to be decomposed into a rotation matrix and a unit translation vector. This is done using the method proposed by Hartley and Zisserman [13]. An essential matrix, E_{12}, is first extracted from the fundamental matrix, F_{12}, using the intrinsic matrices, K_1, K_2, for both cameras of the pair such that $E_{12} = K_2{}^T F_{12} K_1{}^T$. An important constraint about the essential matrix is that it must have two non-zero singular values which are equal. Therefore, E_{12} is refined by computing the average of the first two singular values and setting the third singular value to zero [14]. E_{12} can then be decomposed into a rotation matrix and a unit translation vector. Four mathematical solutions are found: two possible rotation matrices with an offset of 180° about the baseline and two possible translation vectors (positive or negative).

To determine which solution is correct, we triangulate [15] one match and we verify which solution provides a triangulated 3D point which is in front of both cameras (positive z-axis). Ihrke *et al.* [5] reported that two solutions may remain valid if two cameras are close to a 180° rotation. The solution which yields to the smallest triangulation error is therefore retained. When using mid-point triangulation [15], the triangulation error is the half-distance of the segment of intersection between the two 3D rays.

D. Solving Pair-wise Scale Factors

In the previous step, stereo relation for each pair was found up to a pair-related scale factor. During the present phase, the magnitude of all translation vectors is estimated such that the entire camera structure becomes consistent up to a unique, global, scale factor. To do so, the method introduced by *Chen et al.* [4], which attempts to scale the links incrementally, is used. The basis of this technique is to incrementally scale new links based on a scaled link as shown by the camera triplet of Fig. 2. The translation T_{12} is fully scaled but T_{13} and T_{23} are yet to be scaled. To scale T_{13} and T_{23}, the camera positions, P_{cam2} and P_{cam3}, are computed with respect to the camera 1 reference frame. P_{cam2} is obtained directly since it is equal to T_{12}. P_{cam3} is obtained by computing the vector intersection of T_{13} and T_{23}. Since T_{23} is expressed with respect to the camera 2 reference frame, it first needs to be transformed to be expressed with respect to the camera 1 frame using: $\hat{T}_{23} = R_{12} T_{23}$. With P_{cam3} known, the scaling factor of T_{13} is the Euclidian distance between P_{cam3} and (0,0,0) and the scale factor of T_{23} is the Euclidian distance between P_{cam3} and P_{cam2}. Once T_{13} and T_{23} are scaled, they can be used to find scale factors of other links. As more links get scaled, additional links can be created using intermediate transformations, which can also be used to find scale factors of unprocessed links. This procedure is repeated until all links are scaled or until a path linking all cameras to the reference camera is found. Obviously, to start this algorithm, one link needs to be scaled to an arbitrary value (ie. a scale factor of 1). Hence, the structure is fully determined up to a global scale factor.

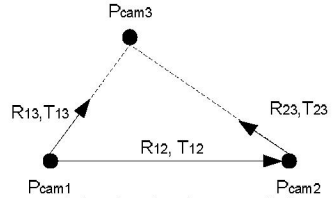


Fig. 2. A camera triplet showing how two links are intersected and scaled from a link with a known scale factor.

However, in the original approach of Chen *et al.*, links are scaled in an arbitrary order which yields to a less accurate initial estimate of the camera structure especially as the number of camera increases or when a link with large inaccuracies is used in the computation of other links. To

overcome this drawback, we propose the use of a weighted camera graph such that the cameras are scaled in a specific order to improve the accuracy of the initial estimate. Weighted graph linking will be detailed in the following sections.

E. Global Unification

The purpose of the global unification step is to find a rotation matrix and a translation vector that link all cameras in the network to the reference camera. These rotation and translation parameters correspond to the respective extrinsic camera parameters. A weighted graph of cameras is built using all scaled links obtained from the previous step. Then the shortest path [16] linking each camera to the reference camera is computed.

F. Bundle Adjustment Optimization

Upon successful global unification, a coarse estimate of the extrinsic parameters for all cameras is obtained. To reach a more precise calibration, the extrinsic parameters are optimized using a sparse bundle adjustment implementation [17, 18]. The same database of 2D image matches collected in the first step is reused as an input to the bundle adjustment. Virtual 3D points are computed using direct linear triangulation [13, 15] on these matches. To help improve the speed of convergence as well as the accuracy of the extrinsic parameters, an analytical Jacobian is provided to the bundle adjustment rather than a numerical approximation from finite differentiation of the point projections [18].

G. Link Cost Assignment to Improve Initial Estimate

In order to guaranty convergence as well as improving the final output of the bundle adjustment, it is important to provide an initial estimate as close as possible to the actual extrinsic camera parameters. Special care needs to be given in the order in which the camera links are being scaled to prevent inaccuracies of one link to be carried over other links. This problem is commonly referred to as error accumulation. A set of five rules is proposed to improve the overall accuracy of such initial estimate: *i*) The startup link should be any link that connects to the reference camera. The graph will be optimized such that lower weighted links will be located near the reference camera. *ii)* A weight of 1 is assigned to the startup link. When links are scaled by triangulation, they will take the weight of the base link + 1. *iii)* Links found by intermediate transformations of multiple scaled links will be weighted as the sum of weights of all intermediate links. *iv)* A queue is used to regulate the order in which the links are evaluated. At the beginning of each iteration, all links that have already been scaled, including links that can be solved by intermediate transformations, are added to an ordered queue where lower weights get dequeued first. When a link is dequeued, an attempt is made to scale as many pairs as possible using this sole link, testing all possible cameras as the third camera of a triplet. Then a next link is dequeued until the queue is empty. A new iteration begins until no new link can or needs to be scaled. *v)* Unscaled links are added incrementally in the camera graph, such that poor quality links are not part of the network if they are not required by the global unification. The quality of an unscaled link is computed as a mixture of the average and the maximal reprojection error of all matches for this pair. This quality factor gets computed after the fundamental matrix decomposition step. The steps of solving pair-wise scale factors and global unification are performed iteratively starting only with high quality links and introducing lower quality links later in the process only if they are absolutely needed to solve the global unification.

H. Rescaling Network to Absolute Units

Recalling that the first link was set to the arbitrary scale factor of 1.0, the extrinsic parameters are estimated up to a single (global) scale factor. While, in terms of accuracy, there are absolutely no benefits to use an absolute scale factor for the camera structure, it may still be desired, in some applications, to use a camera network which is calibrated with meaningful units such as metric units. It would, in particular, allow 3D data to be reconstructed in units that are meaningful to a human operator. A few strategies are proposed to determine the absolute scale factor.

An approximate solution consists of manually measuring the baseline between the reference camera and any other camera in the network. The absolute scale factor becomes the ratio between the measured baseline and the baseline extracted from the camera model. This approach is easy to implement but is not accurate since the position of the camera's optical center is not precisely known and manipulations are necessary.

Another approach builds upon the identification of two features in the working volume seen by two or more cameras. The absolute scale factor becomes the ratio between the absolute (metric) distance and the distance calculated from triangulation of the two features. This approach leads to a more accurate absolute scale factor but requires the capability of identifying and reliably matching the features in multiple camera views. However, this issue can be addressed by using a dual point calibration target with known spacing, and where each marker can be distinguished in the images using, for example, two different colors. The later approach can also be repeated multiple times to obtain more samples of the absolute scale factor across all views and use the average of those samples as the actual real scale factor.

IV. EXPERIMENTAL RESULTS

An experimental evaluation of the proposed calibration framework has been conducted on a complete hardware implementation that has been designed for markerless motion

capture in piano playing performance evaluation. Calibration results that were obtained are analyzed here and an early example of a 3D reconstruction of a piano player achieved using the obtained calibration is presented to demonstrate the validity of the proposed approach.

A. Multi-Camera Network

The system used to monitor human gesture is composed of 8 Flea2™ Firewire 1394b cameras. The cameras are mounted on a structure surrounding a 2.5m x 2.5m x 2.5m workspace which allows cameras to be positioned in various configurations. Video frames are acquired using 3 Pentium IV 3.40 GHz processors running Windows XP. Frames are compressed and recorded to disk in real-time using xvid video compression. Video frames are synchronized using image timestamps across multiple PC using PointGrey Multi-Sync™ software. In our current setup, at most three cameras can be connected per computer such that all cameras can operate at a speed of 30 fps, with a maximal resolution of 640x480. To demonstrate the robustness of the proposed calibration method, lenses from different manufacturers and of different focal length were combined. In particular, five cameras have highly distorting wide angle lenses (f=3.5 mm). The three other cameras have lenses with a focal length of 6 mm.

The intrinsic calibration part is performed under standard lighting conditions. During the extrinsic camera calibration, the lights in the acquisition room are turned off to improve visibility of the LED calibration stick. The level of exposure of the camera is also reduced as much as possible in order to prevent motion blur. It also helps to compensate if the room is not completely dark.

B. Quantitative Analysis

The average reprojection error is used as a reference to evaluate the accuracy of the calibration. A new set of virtual 3D points is calculated by performing linear triangulation of new image matches. Virtual 3D points are reprojected on every image plane. Euclidian distances between reprojected points and the originally measured pixel positions are computed and averaged. The selection of a new database of points ensures that the extrinsic parameters were not overfitted with respect to the original database of matches used for calibration parameters estimation [4].

Table 1 shows the results of multiple calibration experiments using networks composed of 3 to 8 cameras. The first two columns present a comparison of the average reprojection error before and after the bundle adjustment stage. In the last column, a new set of virtual 3D points were used to verify cases of data overfitting. From these results, we can notice that the average reprojection error, before the bundle adjustment, is under 4 pixels for all configurations containing 3 to 8 cameras. However, these results tend to fluctuate considerably between experiments. After the bundle adjustment, the average reprojection error is more stable and is reduced to ½ pixel. Moreover this calibration procedure clearly does not suffer from data overfitting since the average reprojection error is not degraded when using a new set of 3D points. In Table 2, detailed results per camera are given after bundle adjustment. Cameras 2, 3, 4, 6 and 7 are equipped with highly distorting lenses. Nevertheless, all cameras have similar error statistics regardless of the level of lens distortion. Therefore, the proposed extrinsic calibration method is barely affected by lens distortion provided that adequate compensation is obtained at the intrinsic calibration stage, as proposed.

Table 1. Calibration statistics for a network of 3 to 8 cameras.

Number of cameras	Average pixel reprojection error		
	before bundle adjustment	after bundle adjustment	with new set of 3D points
3	1.8221	0.2899	0.3128
4	3.9793	0.3537	0.3547
5	1.6730	0.3660	0.3355
6	3.7134	0.3895	0.3934
7	3.9671	0.4152	0.4090
8	3.9367	0.4174	0.4126

Table 2. Detailed calibration statistics for an 8-camera network.

Camera id	Pixel reprojection error	
	average	std deviation
0	0.3817	0.2189
1	0.2867	0.1707
2	0.4709	0.2677
3	0.5287	0.3180
4	0.4903	0.2804
5	0.4681	0.2540
6	0.3474	0.1991
7	0.4284	0.2164
All Cameras	0.4126	0.2374

C. Qualitative Analysis

The proposed calibration method meets the predefined requirements from section II. This method is easy-to-use as it requires very few human manipulations. Indeed, waving a LED calibration stick randomly in the working volume is simpler and faster than having to perform precise measurements in 3D space. Excellent coverage of the working volume is easily achieved with no coplanarity concerns, as shown in Fig. 3, where black dots correspond to calibration points. The compactness of the calibration target allows the calibration to be performed in a non-empty working volume (Fig. 4a). The calibration results also remain very accurate with the increase in the number of cameras, which makes the proposed method scalable. This calibration procedure may be applied to small or large working volumes with very few modifications to the calibration target. Finally, the procedure imposes very minimal constraints to the camera positioning, as long as enough camera overlap is provided. Fig. 3 shows a complex calibrated camera configuration where some camera pairs have very large baselines and others have much smaller ones (ie. the two vertical top-view cameras).

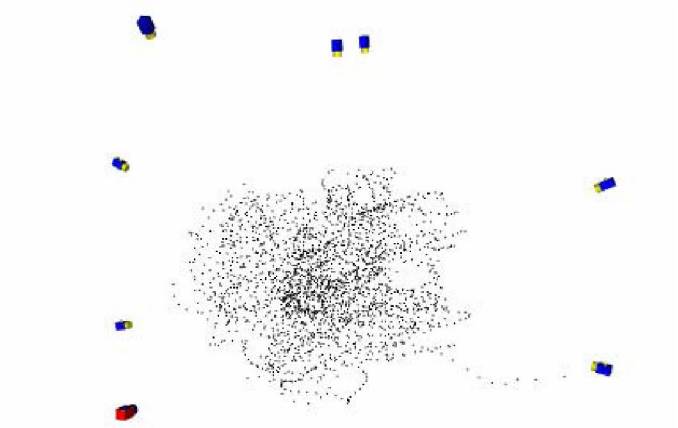
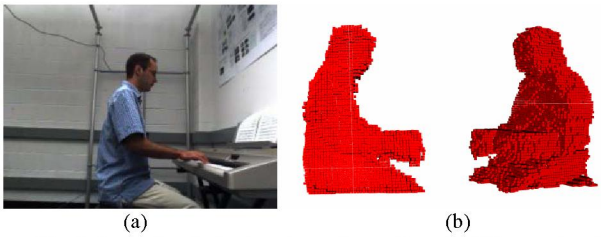


Fig. 3. A model of the full calibration showing the position and orientation of all cameras, with the reference camera in red, and all 3D calibration points.

D. Applications

To demonstrate how the proposed calibration scheme can be used in real 3D computer vision applications, Fig. 4 shows an example of a 7-view shape-from-silhouette reconstruction. A Mixture of Gaussian implementation [19] was used to segment the human silhouette from the background in every view. Fig. 4a shows an undistorted frame from one of the 7 views. Fig. 4b illustrates two views of the reconstructed subject using a 64^3 voxels model. The accuracy of the camera calibration and the complementarity between views allow the successful reconstruction of the human body in such a complex position where many cases of self-occlusion occur.



(a) (b)

Fig. 4. Example of a 7-view shape-from-silhouette 3D reconstruction.
(a) A frame of video from a lateral camera view.
(b)The reconstructed 3D model at a 64^3 voxel resolution.

V. CONCLUSION

A flexible procedure to achieve full calibration of a multi-camera network has been presented. Experimental results demonstrated that the use of a weighted graph analysis is very robust and provides accurate calibration estimates before and after the bundle adjustment. Overall, a reprojection accuracy of up to ½ pixel is achieved for networks containing as much as 8 cameras. This approach also meets all pre-established requirements of scalability and ease-of-use required by non-expert users. Future work will include gathering image matches using self-calibration, by using scene features instead of artificially created (target-based) features.

ACKNOWLEDGMENTS

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

[1] F. Caillette, *Real-Time Markerless 3-D Human Body Tracking*, Ph.D Thesis, University of Manchester, 2006.

[2] S. Drouin, R. Poulin, P. Hébert and M. Parizeau, "Monitoring Flexible Calibration of a Wide Area System of Synchronized Cameras", In *Proc. of the 16th International Conference on Vision Interface*, pp. 49-56, Halifax, June 2003.

[3] P. Rander, *A Multi-Camera Method for 3D Digitization of Dynamic, Real-World Events*, PhD Thesis, Robotics Institute, Carnegie Mellon University, May 1998.

[4] X. Chen, J. Davis, P. Slusallek, "Wide Area Camera Calibration Using Virtual Calibration Objects", In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 520-527, 2000.

[5] I. Ihrke, L. Ahrenberg, M. Magnor, "External camera calibration for synchronized multi-video systems", In *Proc. of the International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG'04)*, Plzen, pp. 537-544, February 2004.

[6] T. Svoboda, D. Martinec, T. Pajdla, "A convenient Multi-Camera Self-Calibration for Virtual Environments", In *Presence: Teleoperators and Virtual Environments*, vol. 14, no. 4, August 2005.

[7] B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon, "Bundle Adjustment: A Modern Synthesis", Vision Algorithms: Theory and Practice, *LNCS*, vol. 1883, pp. 298-372, Springer-Verlag, 2000.

[8] J. Heikkilä, "Geometric camera calibration using circular control points", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1066-1077, October 2000.

[9] R.Y.Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323-344, August 1987.

[10] Z. Zhang, "A Flexible New Technique for Camera Calibration", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330-1334, November 2000.

[11] Open Computer Vision Library, release 1.0, November 2006, *http://sourceforge.net/projects/opencvlibrary*.

[12] M. Fischler, R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography", *Comunications of the ACM*, vol. 24, no. 6, June 1981.

[13] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, UK, 2000.

[14] O. Faugeras, Q-T. Luong, *The Geometry of Multiple Images*, The MIT Press, 2001.

[15] R.I. Hartley, P.Sturm, "Triangulation", *Computer Vision and Image Understanding*, vol. 68, no. 2, pp.146-157, November 1997.

[16] S. Skiena, M. Revilla, *Programming Challenges: The Programming Constest Training Manual*, Springer-Verlag, New York 2003.

[17] M. Lourakis, A. Argyros, "A Generic Sparse Bundle Adjustment C/C++ Package Based on the Levenberg Marquardt Algorithm", version 1.3, *http://www.ics.forth.gr/~lourakis/sba*, 2006.

[18] M. Lourakis, A. Argyros, "The Design and Implementation of a Generic Sparse Bundle Adjustment Software Package Based on the Levenberg-Marquardt Algorithm", Institute of Computer Science, Forth, Technical Report 340, August 2004.

[19] M. Côté, P. Payeur, G. Comeau, "Comparative Study of Adaptive Image Segmentation Techniques for Gesture Analysis in Unconstrained Environments", In *Proc. of the IEEE International Workshop on Imaging Systems and Techniques*, pp. 28-33, Minori, Italy, 2006.