シーンモデルを利用した移動カメラキャリブレーション方式

メヒルダド パナヒプル テヘラニ† 石川 彰夫† 酒澤 茂之† 小池 淳†
†映像通信グループ、KDDI 研究所 〒105-0123 埼玉県ふじみ野市大原 2-1-15
E-mail: †{te-mehrdad, ao-ishikawa, sakazawa, koike}@kddilabs.jp

あらまし ユーザにより生成されたシーンのモデルを利用した自動カメラキャリブレーションアルゴリズムを提案する。新しいアプローチとして、フレームごとに撮影画像と生成モデルの間で SIFT 等の方式によって特徴点を検出し、マッチングを行う。マッチング自体は、ANN 探索手法、SIFT マッチング法が適用可能である。また、その結果得られた特徴点ペアに対して、幾何的に対応しないものを排除する手法を導入し、ヒストグラム分布による絞り込み、RANSAC/LMedS 反復法と基礎行列算出、逐次精密化のアルゴリズムを採用した。実験の結果、オクルージョンや劣悪な照明環境においてもロバストに、幾何的に対応する特徴点ペアのみが単一フレームから検出できることを確認した。これにより、移動カメラのキャリブレーションを高精度に実現可能となる。

キーワード カメラキャリブレーション、シーンモデル利用、特徴点抽出、非幾何対応特徴点排除
.

Moving Camera Calibration Using Generated Model of a Scene

Mehrdad Panahpour Tehrani† Akio Ishikawa† Shigeyuki Sakazawa† Atsushi Koike†
†Visual Communication Laboratory, KDDI R&D Laboratories Inc.
2-1-15 Ohara, Fujimino, Saitama, 356-8502 Japan
E-mail: †{te-mehrdad, ao-ishikawa, sakazawa, koike}@kddilabs.jp

Abstract We propose an automatic camera calibration algorithm using generated model of a scene by user. In our approach, for each frame, the algorithm detects feature points in image and the model using SIFT or any other feature detectors. Matching can be performed using ANN search method, SIFT matching algorithm or any other matching algorithm. Non-geometrically matched features points are suppressed by using "histogram based suppression", "RANSAC/LMedS iteration and fundamental matrix calculation" and "iterative refinement". The detected feature points are used for camera calibration. Experimental results show that the proposed method can robustly detect geometrically matched feature points in the case of occlusion and bad lightening condition; therefore it can be used for precise moving camera calibration.

Keyword User generated model of a scene, Feature points detection, Suppression of non-geometrically matched feature points, Camera Calibration.

1. Introduction

Camera calibration is a necessary step in 3D representation applications such as FTV [1]. In order to extract camera parameters from 2D images much works have been done. Generally camera calibration methods are performed by observing a calibration object whose geometry is known with a very good precision. Camera parameters are extracted using this information through a minimization process. We can classify those approaches into four categories as follows.

3D based calibration: In this method [2], a 3D object (with two or three planes) is used for camera calibration, whose its geometry in 3D space is accurately known. This approach can perform calibration very efficiently.

2D based calibration: In Zhang method [3], a 2D object (planar pattern) with known 2D geometry requires to be captured at a few different orientations. Method in [3] does not need the plane motion information, whereas in Tsai's method [4], the knowledge of the plane motion is necessary for camera calibration.

1D based calibration: Zhang [5] also proposed a method that can calibrate camera using 1D

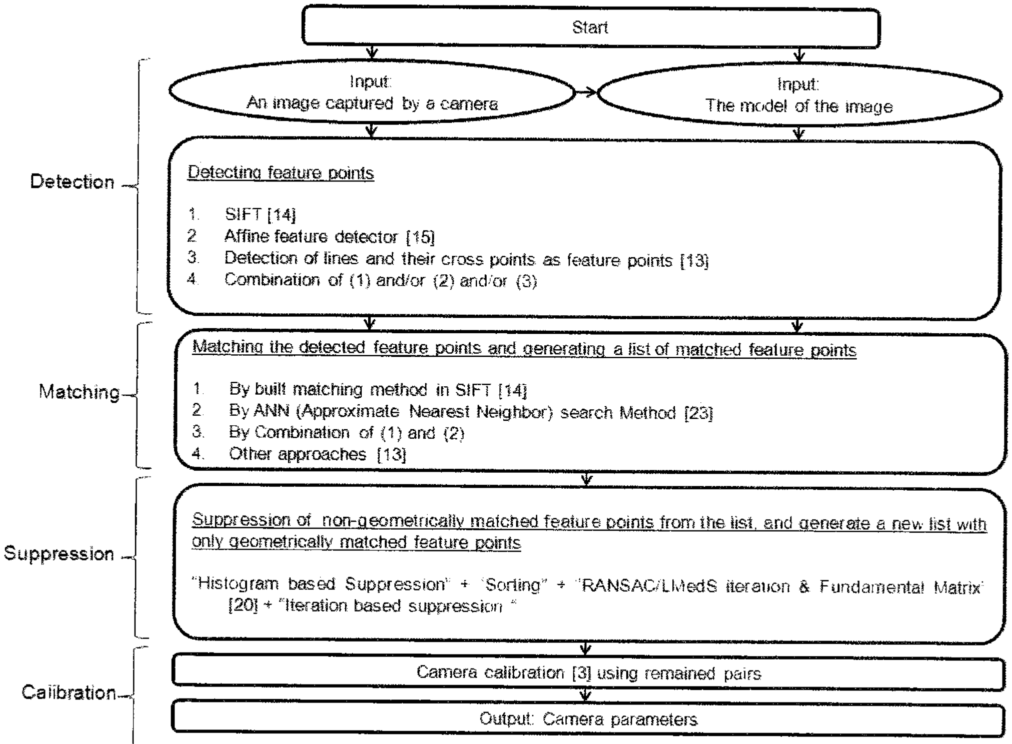

Fig. 1. Flowchart of the proposed camera calibration method.

object (points aligned on a line). Camera calibration is possible by moving 1D object, when one point is fixed. The solution is available if at least six or more observations of such a 1D object are made. This technique is practical especially when we are calibrating multiple cameras and the calibration objects are required to be visible simultaneously.

0D based calibration: Techniques in this category can perform self-calibration. They can be considered as 0D approach because only captured images are required. Methods in [6], [7] use captured images from static scene using moving camera, including rotations, translations, and changes of focal length. If images are taken by the same camera with fixed internal parameters, correspondences between three images are sufficient to recover camera parameters. So far, several approaches tried to address this problem for specific situation such as calibration of cameras that capture sport events [8-13]; having the dimension of the sport court as a model in advance. Methods in [8-12] extract camera parameters partially for specific sport court (i.e. Tennis, or Soccer). Method in [13] can compute the camera parameters for an eight-parameter perspective model, in any sport court.

3D, 2D or 1D object based camera calibration methods are suitable for still cameras. Camera calibration can be done only using fixed pattern (e.g. chess board, stick, and etc).

However, 0D object based calibration scheme have capability to be used for moving camera calibration. Although no calibration objects are necessary, a large number of parameters need to be estimated, resulting in a much harder mathematical problem. These methods can handle arbitrary camera motion. Methods used for sport court have a limited application, and even weak camera calibration cannot be performed in some cases

In this research, we address the problem of moving camera calibration for wider range of applications (i.e. not limited to sport court). Our approach is to assume that for any captured scene, a user generated computer graphic scene with precise geometry is available. Using the generated model and each captured frame by the moving camera, geometrically matched feature points are detected through several processes, and they are used to extract camera parameters. The proposed calibration method can be categorized to either self-calibration scheme, or a new category, namely, camera calibration using arbitrary calibration pattern.

2. Algorithm

Fig. 1 shows the algorithm of the proposed camera calibration. We assume that for each frame of a captured image (hereafter is referred to as "image") a human generated model of the place of capturing (hereafter is referred to as "model") is available. Having the image and model, camera calibration for the given frame can be performed.

In summary, our proposed algorithm for camera calibration, firstly finds feature points in each image, and model. We use the SIFT (Scale Invariant Feature Transform) feature point detector [14]. A typical image contains several thousand feature points. Other feature detectors could also potentially be used, such as affine detectors [15]. In addition to the feature points locations themselves, SIFT provides a local descriptor for each feature point. Next, for each pair of image and model, we match feature points descriptors between the pair, using the ANN approximate nearest neighbors search method [16] or built in matching algorithm in SIFT or their combination. To suppress non-geometrically matched feature points we propose to use "histogram based suppression", and/or "RANSAC (RANdom SAmple Consensus)[17]/LMedS (Least Median Square)[18] iteration and fundamental matrix calculation using eight points algorithm [19]"[20] and/or "iterative refinement process". Following four steps explain the detail of the proposed camera calibration.

Step1. Detection: For given image and model, any of feature detectors can be used as follows. To detect the feature points in model and image, SIFT [14] or any other feature detector such as affine detectors [15] can be used. To detect more similar feature points in model and image, we propose to perform color correction [21] on model when image is reference. Note that feature points detection using SIFT generates a list of feature points locations, orientations, scale, and description data.

If the structure of the scene is simple with lines, we propose to use method in [13] that detects lines. This method applies color filtering to detect line (i.e. lines are with specific color), removes texture area using structure matrix [22], generates lines using Hough transform, rejects line by scoring them according to their distance to the detected line, and calculates the cross points of the remained lines, as feature points. Note that if we use cross points of lines in the scene as feature points, we do not need to perform color correction on the model and image, since the model lines and cross points are available in advance. To increase the detected feature points we can combine the result of all detection method for better performance in following steps.

Step2. Matching: Given the detected feature points for image and model in step 1, any matching methods can be used as following. Matched feature points can be detected by using built in matching algorithm in SIFT [14] (i.e. BBN (Best Bin Fit) algorithm [23]), or ANN (Approximate Nearest Neighbor) search method [16] that use the descriptors that are available from SIFT detector. To increase the detected feature points we can combine the result of all detection method for better performance in following steps.

Step3. Suppression: Given detected matched feature points between model and image from step 2, we propose following method to suppress non-geometrically matched feature points.

Histogram Based Suppression: First, histogram of lines connecting matched feature points, provided in step 4, is generated, and mean (μ) and standard deviation (σ) of the histogram is calculated. Lines which their angles are in the range of ($m\pm\lambda.s$) are kept and the rest are suppressed.

Sorting: Second, the feature points in (image/model), they are sorted based on their location in the image, from up-left corner to the down-right side of the image, and their matched feature points are attached from model as sorted pairs. Note we found out that sorting the list of matched feature points enhances the performance of the further suppression processes.

Suppression Using RANSAC/LMedS iteration and Fundamental Matrix Calculation: Third, RANSAC or LMedS iteration with fundamental matrix calculation using eight points algorithm

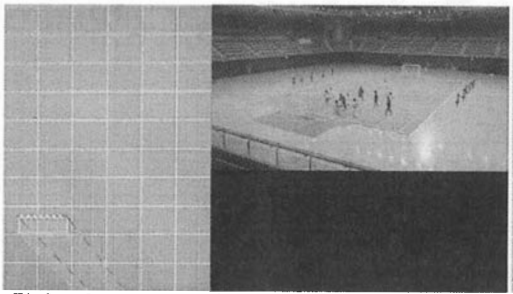

Fig2. Input capture image and generated model after color correction [20].

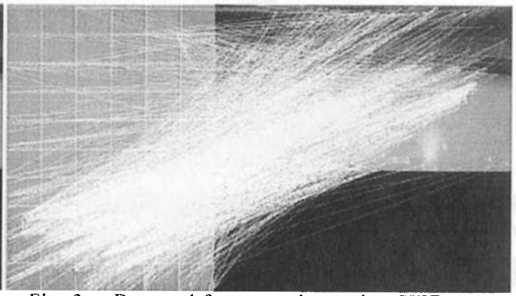

Fig. 3: Detected feature points using SIFT and matched ANN

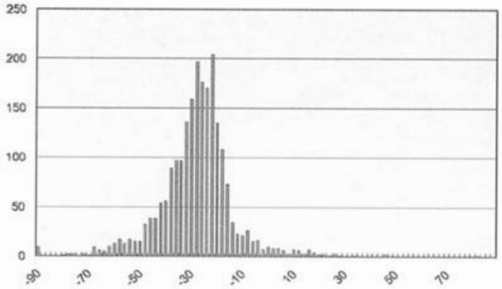

Fig. 4. Histogram of matched feature point

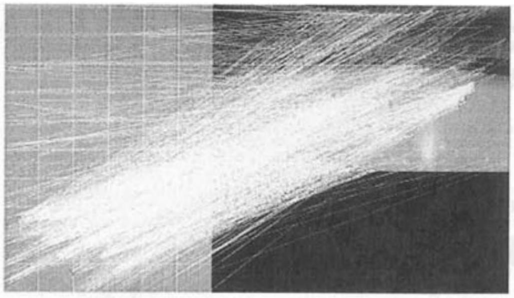

Fig. 5. Histogram based suppression

[19] is performed to suppress the non-geometrically matched feature points. In this step most of non-geometrically matched feature points are suppressed [20].

Refinement Using Iteration Based Suppression: Forth, the remained and sorted matched feature points are divided into eight or four groups. Sets of eight/four pairs, with members from different groups are generated. For each set, fundamental matrix [19]/homography matrix from image to model is calculated, and feature points in image are projected into model coordinate using the calculated matrix. A constraint in which "if more than th% (th<100) of feature points are projected nearly on the same location (location difference in pixel value is less than "d")" is applied. If the constraint is satisfied, pairs in the set are flagged as "OK", whereas not satisfied constraint, as "suspended". For each pair of "suspended" in each group, a new set is generated with "OK" pairs in other groups, and checked with the constraint to either flag them as "OK", or suppressed. This process is done for all "suspended" pairs, until "OK" pairs are left. The same procedure explained in this stage is iteratively done by changing the thresholds (i.e. increasing "th", and/or decreasing "d") for better projection result.

Note that the according to the required accuracy or complexity, any combination of the abovementioned stages can be used for suppression.

Step4. Calibration: Given the geometrically matched feature points by step 3, camera calibration can be done by a nonlinear minimization of the distance between the feature points in the model and projected feature points from image to model as done in [3].

3. Experiment

To evaluate our method, we have generated a model by precisely generating top view image of "*futsal*" court from side captured image when there is no player in the court. Therefore, the projected image can simply be used as a realistic model for further captured scene that players are in the scene. Fig. 2 shows the input image and the generated model for "*futsal*" sequence.

Following the proposed algorithm, as shown in Fig. 3, in this experiment feature points are detected using SIFT, and matching is performed using ANN search method. Then, using "histogram based suppression", part of non-geometrically matched feature points are suppressed,

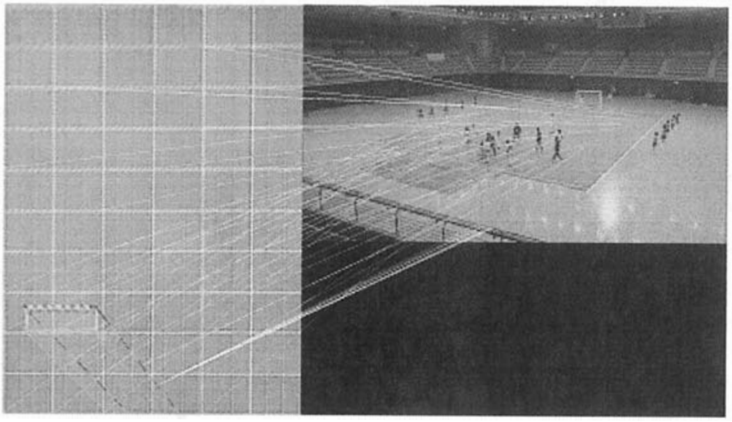

Fig. 6. Suppression using RANSAC/LMedS iteration and fundamental matrix calculation.

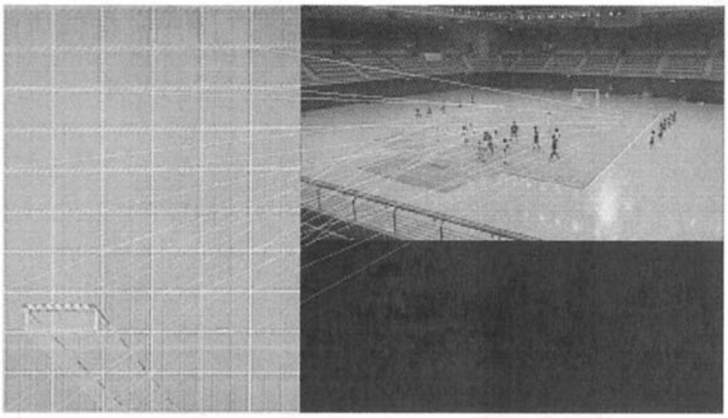

Fig. 7 Suppression using RANSAC/LMedS iteration and fundamental matrix calculation.

as shown in Fig. 4, and Fig. 5, following by "sorting" the pairs. "Histogram based suppression" and "Sorting" enhance the performance of "RANSC/LMedS + Fundamental Matrix" suppression. Result of the later process is shown in Fig. 6. As it can be seen there are some non-geometrically matched points left that impair the result of camera calibration. To suppress the remained non-geometrically matched feature points, "Iterative refinement" is applied as shown in Fig. 7. Remained matched feature points are all geometrically matched feature points. Experimental result shows that the proposed method can detect sufficient geometrically matched feature points; therefore it can be used for precise camera calibration being robust to occlusion, and bad lightening condition of moving cameras.

4. Conclusions

The proposed camera calibration method can be used for both still and moving camera when a user generated model of the scene is available. This method is independent of the scene and any calibration pattern, robust to occlusions, and bad lighting conditions. This calibration method can be categorized to either self-calibration scheme, or camera calibration using arbitrary calibration pattern.

In our future research, we will focus on enhancing the calibration accuracy and accelerating the algorithm for realtime operation.

References

[1] P. Na Bangchang, M. Panahpour Tehrani, T. Fujii, M. Tanimoto, "Realtime System of Free Viewpoint Television", The journal of the institute of Image information and Television Engineers (ITE), vol. 59, no. 8, pp. 63-70, Aug. 2005.

[2] O. Faugeras, "Three-Dimensional Computer Vision: a Geometric Viewpoint," MIT Press, 1993.

[3] Z. Zhang, "A flexible new technique for camera calibration", Technical report MSR-TR-98-71, Microsoft research, Microsoft corporation, WA, USA, 1998.

[4] 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, 3-4, pp. 323–344, Aug. 1987.

[5] Z. Zhang, "Camera calibration with one-dimensional objects", Technical report MSR-TR-2001-120, Microsoft research, Microsoft corporation, WA, USA, Aug 2002.

[6] R. I. Hartley. An algorithm for self calibration from several views. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 908–912, Seattle, WA, Jun. 1994.

[7] Q.-T. Luong and O.D. Faugeras. "Self-calibration of a moving camera from point correspondences and fundamental matrices," The International Journal of Computer Vision, 22-3, pp. 261–289, 1997.

[8] G. Sudhir, J. C. M. Lee, and A. K. Jain, "Automatic classification of tennis video for high-level content-based retrieval," in IEEE International Workshop on Content Based Access of Image and Video Databases, in conjunction with ICCV'98, pp. 81-90, 1998.

[9] C. Calvo, A. Micarelli, and E. Sangineto, "Automatic annotation of tennis video sequences," in DAGM-Symposium, pp. 540-547, Springer, 2002.

[10] A. Ekin and A. M. Tekalp, "Automatic soccer video analysis and summarization," in SPIE Storage and Retrieval for Media Databases IV, pp. 339-350, Jan. 2003.

[11] A. Yamada, Y. Shirai, and J. Miura, "Tracking players and a ball in video image sequence and estimating camera parameters for 3D interpretation of soccer games," in Proc. 16th Int. Conf. on Pattern Recognition, pp. 303-306,Aug. 2002.

[12] H. Kim and K. Hong, "Robust image mosaicing of soccer videos using self-calibration and line tracking," Pattern Analysis & Applications 4(1), pp. 9-19, 2001.

[13] D. Farin, Susanne K., P.H.N. de With, and Wolf Effelsberg, "Robust camera calibration for sport videos using court models," Proc. SPIE Storage and Retrieval Methods and Applications for Multimedia 2004, vol.5307, pp. 80–91, Jan. 2004.

[14] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International. Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[15] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas., F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," Int. J. of Computer Vision 65, 1-2, pp. 43-72, 2005.

[16] A. Arya, D. Mount, N.S. Netanyahu, R. Silverman, and A.Y. Wu, "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," J. of the ACM 45, 6, pp. 891-923, 1998.

[17] M.A. Fischler., R.C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Comm. of the ACM, vol 24, pp. 381-395, 1981.

[18] Peter Meer, "Robust Techniques for Computer Vision", Chap 4 of the book Emerging Topics in Computer Vision.

[19] R.I. Hartley, and A. Zisserman, "Multiple View Geometry," Cambridge University Press, Cambridge, UK, 2004.

[20] N. Snavely, S.M. Seitz, and R. Szeliski, "Photo tourism: Exploring photo collections in 3D," ACM Transactions on Graphics (SIGGRAPH Proceedings), 25-3, pp. 835-846, 2006.

[21] M. Panahpour Tehrani, A. Ishikawa, S. Sakazawa, and A. Koike, "The Optimal Color Correction of Multicamera Systems" IEICE Technical Report, IE2007-94, pp. 23-28, Nov. 2007.

[22] B. Jahne, "Digital Image Processing," Springer Verlag, 2002.

[23] Beis, J. and Lowe, D.G. 1997. Shape indexing using approximate nearest-neighbour search in high dimensional spaces. In Conference on Computer Vision and Pattern Recognition, Puerto Rico, pp. 1000-1006.