



HAL
open science

Calibration of a non-single viewpoint system

Peter Sturm

► **To cite this version:**

Peter Sturm. Calibration of a non-single viewpoint system. Katsushi Ikeuchi. Computer Vision: A Reference Guide, Springer, pp.118–122, 2021, 978-0-387-30771-8 (Print) 978-0-387-31439-6 (Online). 10.1007/978-0-387-31439-6_161 . hal-00759938

HAL Id: hal-00759938

<https://inria.hal.science/hal-00759938v1>

Submitted on 6 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Calibration of a non-single viewpoint system

PETER STURM, INRIA, FRANCE

Related Concepts

- Center of Projection
- Intrinsic parameters
- Distortion
- Extrinsic parameters
- Camera calibration
- Camera model
- Camera parameters
- Projection

Definition

A non-single viewpoint system refers to a camera for which the light rays that enter the camera and contribute to the image produced by the camera, do not pass through a single point. The analogous definition holds for *models* for non-single viewpoint systems. Hence, a non-single viewpoint camera or model does not possess a single center of projection. Nevertheless, a non-single viewpoint model (NSVM), like any other camera model such as the pinhole model, enables to project points and other geometric primitives, into the image and to back-project image points or other image primitives, to 3D. Calibration of a non-single viewpoint model consists of a process that allows to compute the parameters of the model.

Background

There exist a large variety of camera technologies (“regular” cameras, catadioptric cameras, fish-eyes, etc.) and camera models designed for these technologies. Often, technologies are developed in order to accommodate a desired model, such as for example to provide a uniform spatial resolution.

Most cameras used in computer vision and other areas, can be well modeled by so-called single viewpoint, or central, camera models. These usually model the 3D-to-2D mapping carried out by a camera, *via* lines of sight (or, camera rays) that all pass through a single point (the center of projection or optical center), and a mapping from these lines of sight to the image points where they hit the image plane.

Some camera types, especially some cameras having a wide field of view, but not only these, cannot be modeled very well using a single viewpoint model. This may be the case because a camera was designed to possess lines of sight that do not pass through a single point. This is for example the case for catadioptric cameras where the mirror surface is of conical shape: even if the camera looking at the mirror is positioned on the mirror’s axis, the lines of sight of the system do not converge to a single point, rather there exists a *viewpoint locus*. Another

example is a single-lens stereo system consisting of a pinhole camera and two planar mirrors, such that the obtained images represent two perspective images acquired from two different effective viewpoints.

A camera may also be unintentionally of the non-single viewpoint type, for example catadioptric cameras that were designed to have a single viewpoint but that due to a bad alignment of the camera and the mirror of the system, lose the single viewpoint property. Another example are fish-eye cameras; fish-eye optics are complex and in principle, one can probably consider them as non-single viewpoint systems. However, in this and the previous example, it is not clear without further investigation of the actual system under consideration, if a single viewpoint model or an NSVM is better suited. This indeed depends on “how much” the system deviates from having a single viewpoint, how close the scene is in a typical application, how much image resolution is available, and so forth. This issue is further discussed in [1].

Finally, let us also note that it is possible to treat a set of multiple cameras that are rigidly attached to one another, as a single imaging system [2]. In case the cameras are sufficiently far from one another, the compound imaging system can be described as an NSVM. Particularly interesting is the case where there is no overlap in the fields of view of the individual cameras [3].

In the following, it is supposed that calibration is performed by acquiring one or several images of a calibration object, whose geometry is known and whose characteristic features (for simplicity, points shall be considered) can be extracted and identified in images.

Theory

There are different types of NSVM's. One usually distinguishes parametric from non-parametric such models. For example, for non-single viewpoint catadioptric systems, if the shape of the system's mirror is known or is known to belong to a parametric family of shapes, then the entire system can be described by few parameters: intrinsic parameters of the camera looking at the mirror, relative pose of mirror and camera, and possibly, shape parameters for the mirror. If such a parametric model is considered, calibration is, conceptually speaking, analogous to that of pinhole cameras. The main difference to calibration of pinhole cameras usually concerns the initialization process that allows to compute initial estimates of the camera parameters. Other than that, one may in general formulate calibration by a bundle adjustment type optimization of camera parameters, by minimizing for example the reprojection error, i.e. a measure related to the distance between the predicted projections of points of the calibration object and those extracted in the images. Examples of parametric NSVM's are the so-called two-plane and GLC models [4,5,6,7], where lines of sight are parameterized by linear or higher order transformations applied to points in two basis planes, other similar models where lines of sight are parameterized by linear transformations [8,9,10,11], models for pushbroom and X-slit cameras [12,13,14,15], and others [16].

A different concept consists in using non-parametric models to calibrate cameras. An example is the raxel model introduced by Grossberg and Nayar [17]. It

essentially associates, to each pixel, a ray in 3D supporting the line of sight, and possibly properties such as its own radiometric response function. Importantly, one may use such a model without making any assumption about a parametric relationship between the position of pixels and the position and orientation of the associated lines of sight. Rather, one may store the coordinates of the lines of sight of all pixels, in a look-up table. Simplified versions of this model (without considering optical properties for individual pixels) have been used in several similar calibration approaches, for example [18,19,20,17].

The principle of these approaches is thus to compute, for every camera pixel, a line of sight in 3D. To do so, at least two images of a calibration object are required. The simplest scenario considers the case where the calibration object is displaced by some known motion, between the image acquisitions. For each image, one has to estimate correspondences between camera pixels and points on the calibration object. One way of achieving such dense correspondences is to use structured light principles, for instance to use as calibration object a computer screen and to display a series of black-and-white patterns on it that encodes each pixel of the screen by a unique sequence of black and white intensities (e.g. Grey codes). Once correspondences of camera pixels and points of the calibration object (pixels of the computer screen in the above example) are known, the lines of sight can be computed by simply fitting straight 3D lines to the matched points on the calibration object. To do so, the latter must be expressed in the same 3D coordinate system, which is possible since it was assumed above that the motion of the calibration object between different acquisitions, is known. This approach was proposed independently by different researchers [18,19,17].

The above approach requires a minimum of two images, for different positions of the calibration object, and knowledge of the object's displacements. An extension to the case of unknown displacements was proposed in [20]. That approach requires at least three images; from matches of camera pixels and points on the calibration object, it first recovers the displacements of the object using an analysis of this scenario's multi-view geometry, and then computes lines of sights as above. Other approaches following this line are [21,22,23].

The above approaches compute, for each camera pixel, an individual line of sight. If one assumes that the relation between pixels and lines of sight is particular, for instance radially symmetric about an optical axis, then alternative solutions become possible. Such a possibility is to use a non-parametric representation of the distortion or undistortion function of a camera, i.e. a function that maps viewing angles (angles between lines of sight and the optical axis) to distances in the image, between image points and the principal point or a distortion center. This can be done for both, single viewpoint and non-single viewpoint models. In the former case, it is assumed that all lines of sight pass through a single center of projection whereas in the latter case, the model usually includes a mapping from viewing angles to the position of the intersection between lines of sight and the optical axis. Approaches of the latter type include [24,25]. Besides making and using the assumption that the camera is radially symmetric, these calibration approaches resemble those explained above.

Different other approaches exist, for example one based on the “surface model” [26], where the raxels of the model of Grossberg and Nayar are interpolated through the help of spline surfaces, or the method of [27] which estimates the refractive surface in front of a camera that makes the compound imaging system a non-central one. As for the case of multi-camera systems modeled as NSVM’s, mentioned under “Background”, dedicated calibration approaches have been developed especially for the case of cameras with non-overlapping fields of view, see for instance [28].

Application

All approaches described above, be they parametric or non-parametric, allow to perform 3D-to-2D projection and/or 2D-to-3D backprojection, the latter meaning the mapping from an image point to the associated line of sight. By definition, the parametric models give analytical expression to perform these operations. As for non-parametric ones, projection and back-projection usually implies some interpolation and possibly, a search. For instance, if a non-parametric model consists of a look-up table that gives, for each pixel, its line of sight: back-projection of an image point with non-integer coordinates requires interpolation whereas projection of a 3D point requires the search of the closest line(s) of sight in the look-up table and again an interpolation stage.

Other than these particular aspects, NSVM’s can be used for many structure-from-motion computations completely analogously to other camera models, in particular the pinhole model. Among the essential building blocks of structure-from-motion, there are pose estimation, motion estimation and 3D point triangulation for calibrated cameras. As for pose and motion estimation (and other tasks), one usually requires two types of methods in an application: so-called minimal methods, which perform the estimation task from the minimum required number of point matches and which can be efficiently embedded in robust estimation schemes such as RANSAC, and non-linear optimization methods that refine initial estimates obtained from minimal methods. Minimal methods for pose [29,30,31] and motion estimation [32,33,34] are formulated analogously to those for the pinhole model, although their algebraic and algorithmic complexity is generally higher and the minimal configurations different (such as the required number of point matches or views).

Nevertheless, all that is essentially required by these methods from the NSVM is to compute lines of sight of interest points that are extracted and matched to another image (for motion estimation) or to a reference object (for pose estimation). As for the non-linear optimization stage, the minimization of the reprojection errors requires 3D-to-2D projections to be carried out, which, as explained above, may require search and interpolation, in which case the computation of the cost function’s derivatives may have to rely on numerical differentiation. Other than that, there is no major conceptual difference compared to pose/motion estimation with pinhole cameras.

Another essential structure-from-motion task is 3D point triangulation. Here again, suboptimal methods work with lines of sight computed by the camera model for interest points in the images and optimal methods perform the non-

linear optimization of reprojection errors, where the same considerations hold as above for pose and motion estimation.

Recommended Readings

- [1] Sturm, P., Ramalingam, S., Tardif, J.P., Gasparini, S., Barreto, J. (2011). Camera models and fundamental concepts used in geometric computer vision. *Foundations and Trends in Computer Graphics and Vision* **6**(1-2) 1–183
- [2] Pless, R. (2003). Using many cameras as one. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Madison, USA. Volume 2.* 587–593
- [3] Caspi, Y., Irani, M. (2002). Aligning non-overlapping sequences. *International Journal of Computer Vision* **48**(1) 39–51
- [4] Chen, N.Y. (1979). Visually estimating workpiece pose in a robot hand using the feature points method. PhD thesis, University of Rhode Island, Kingston
- [5] Chen, N.Y., Birk, J., Kelley, R. (1980). Estimating workpiece pose using the feature points method. *IEEE Transactions on Automatic Control* **25**(6) 1027–1041
- [6] Martins, H., Birk, J., Kelley, R. (1981). Camera models based on data from two calibration planes. *Computer Graphics and Image Processing* **17** 173–180
- [7] Yu, J., McMillan, L. (2004). General linear cameras. *Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic. Volume 2.* 14–27
- [8] Pajdla, T. (2002). Stereo with oblique cameras. *International Journal of Computer Vision* **47**(1-3) 161–170
- [9] Ponce, J. (2009). What is a camera? *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA*
- [10] Seitz, S., Kim, J. (2002). The space of all stereo images. *International Journal of Computer Vision* **48**(1) 21–38
- [11] Batog, G., Goaoc, X., Ponce, J. (2010). Admissible linear map models of linear cameras. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA*
- [12] Gupta, R., Hartley, R. (1997). Linear pushbroom cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(9) 963–975
- [13] Pajdla, T. (2002). Geometry of two-slit camera. Technical Report CTU-CMP-2002-02, Center for Machine Perception, Czech Technical University, Prague
- [14] Zomet, A., Feldman, D., Peleg, S., Weinshall, D. (2003). Mosaicing new views: The crossed-slit projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(6) 741–754
- [15] Feldman, D., Pajdla, T., Weinshall, D. (2003). On the epipolar geometry of the crossed-slits projection. *Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France.* 988–995

- [16] Gennery, D. (2006). Generalized camera calibration including fish-eye lenses. *International Journal of Computer Vision* **68**(3) 239–266
- [17] Grossberg, M., Nayar, S. (2005). The raxel imaging model and ray-based calibration. *International Journal of Computer Vision* **61**(2) 119–137
- [18] Gremban, K., Thorpe, C., Kanade, T. (1988). Geometric camera calibration using systems of linear equations. *Proceedings of the IEEE International Conference on Robotics and Automation*, Philadelphia, Pennsylvania, USA. 562–567
- [19] Champlébourg, G., Lavallée, S., Sautot, P., Cinquin, P. (1992). Accurate calibration of cameras and range imaging sensors: the NPBS method. *Proceedings of the IEEE International Conference on Robotics and Automation*, Nice, France. 1552–1558
- [20] Sturm, P., Ramalingam, S. (2004). A generic concept for camera calibration. *Proceedings of the 8th European Conference on Computer Vision*, Prague, Czech Republic. 1–13
- [21] Ramalingam, S., Sturm, P., Lodha, S. (2005). Towards complete generic camera calibration. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, USA. Volume 1. 1093–1098
- [22] Dunne, A., Mallon, J., Whelan, P. (2010). Efficient generic calibration method for general cameras with single centre of projection. *Computer Vision and Image Understanding* **114**(2) 220–233
- [23] Nishimura, M., Nobuhara, S., Matsuyama, T., Shimizu, S., Fujii, K. (2015). A linear generalized camera calibration from three intersecting reference planes. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile. 2354–2362
- [24] Tardif, J.P., Sturm, P., Trudeau, M., Roy, S. (2009). Calibration of cameras with radially symmetric distortion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**(9) 1552–1566
- [25] Ying, X., Hu, Z. (2004). Distortion correction of fisheye lenses using a non-parametric imaging model. *Proceedings of the Asian Conference on Computer Vision*, Jeju Island, Korea. 527–532
- [26] Rosebrock, D. (2016). “The Surface Model” – An Uncertain Continuous Representation of the Generic Camera Model and its Calibration. PhD thesis, Technische Universität Braunschweig, Germany
- [27] Szabolcs, P., Csanád, S., Lehel, C. (2019). Distortion estimation through explicit modeling of the refractive surface. *Proceedings of the 28th International Conference on Artificial Neural Networks*, Munich, Germany. 17–28
- [28] Lébraly, P., Royer, E., Ait-Aider, O., Dhome, M. (2010). Calibration of non-overlapping cameras – application to vision-based robotics. *Proceedings of the British Machine Vision Conference*, Aberystwyth, UK
- [29] Chen, C.S., Chang, W.Y. (2004). On pose recovery for generalized visual sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(7) 848–861
- [30] Ramalingam, S., Lodha, S., Sturm, P. (2004). A generic structure-from-motion algorithm for cross-camera scenarios. *Proceedings of the 5th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras*, Prague, Czech Republic. 175–186

- [31] Nistér, D., Stewénius, H. (2007). A minimal solution to the generalised 3-point pose problem. *Journal of Mathematical Imaging and Vision* **27**(1) 67–79
- [32] Stewénius, H., Åström, K. (2004). Structure and motion problems for multiple rigidly moving cameras. *Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic. Volume 3.* 252–263
- [33] Stewénius, H., Nistér, D., Oskarsson, M., Åström, K. (2005). Solutions to minimal generalized relative pose problems. *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*
- [34] Ventura, J., Arth, C., Lepetit, V. (2015). An efficient minimal solution for multi-camera motion. *Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile.* 747–755