Computer Vision and Image Understanding 121 (2014) 108–118

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Automatic detection of calibration grids in time-of-flight images ☆

Miles Hansard [a,b], Radu Horaud [a,*], Michel Amat [a], Georgios Evangelidis [a]

[a] INRIA Grenoble Rhône-Alpes, 38330 Montbonnot Saint-Martin, France
[b] School of Electronic Engineering and Computer Science, Queen Mary, University of London, Mile End Road, United Kingdom

ARTICLE INFO

Article history:
Received 25 February 2013
Accepted 23 January 2014
Available online 31 January 2014

Keywords:
Range imaging
Time-of-flight sensors
Camera calibration
Hough transform

ABSTRACT

It is convenient to calibrate time-of-flight cameras by established methods, using images of a chequerboard pattern. The low resolution of the amplitude image, however, makes it difficult to detect the board reliably. Heuristic detection methods, based on connected image-components, perform very poorly on this data. An alternative, geometrically-principled method is introduced here, based on the Hough transform. The projection of a chequerboard is represented by two pencils of lines, which are identified as oriented clusters in the gradient-data of the image. A projective Hough transform is applied to each of the two clusters, in axis-aligned coordinates. The range of each transform is properly bounded, because the corresponding gradient vectors are approximately parallel. Each of the two transforms contains a series of collinear peaks; one for every line in the given pencil. This pattern is easily detected, by sweeping a dual line through the transform. The proposed Hough-based method is compared to the standard OpenCV detection routine, by application to several hundred time-of-flight images. It is shown that the new method detects significantly more calibration boards, over a greater variety of poses, without any overall loss of accuracy. This conclusion is based on an analysis of both geometric and photometric error.

1. Introduction

Time-of-flight (TOF) cameras produce a *depth image*, each pixel of which encodes the distance to the corresponding point in the scene. These devices emit pulsed infrared illumination, and infer distances from the time taken for light to reflect back to the camera. The TOF sensor can therefore be modelled, geometrically, as a pinhole device. Furthermore, knowledge of the TOF camera-parameters can be used to map raw depth-readings (i.e. distances along lines of sight) into Euclidean scene-coordinates. The calibration thereby enables these devices to be used as stand-alone 3-D sensors, or to be combined with ordinary colour cameras, for complete 3-D modelling and rendering [1–7].

TOF cameras can, in principle, be calibrated with any existing camera calibration method. For example, if a known chequerboard pattern is detected in a sufficient variety of poses, then the internal and external camera parameters can be estimated by standard routines [8–10]. It is possible to find the chequerboard vertices, in ordinary images, by first detecting image-corners [11], and subsequently imposing global constraints on their arrangement [12–14]. This approach, however, is not reliable for low-resolution images (e.g. in the range 100–500 px^2) because the local image-structure is disrupted by sampling artefacts, as shown in Fig. 2. Furthermore, these artefacts become worse as the board is viewed in distant and slanted positions, which are essential for high quality calibration [15,16]. *The central motivation of this work is to detect a greater number and variety of calibration board-poses, in TOF images, without increasing the geometric error of the vertices*. The geometric error can be conveniently defined with respect to the known geometry of the board, as will be shown in Section 3.

TOF sensors provide low-resolution depth and amplitude images. This is because relatively large detector-elements are required in order to allow accumulation of electrons, which increases the signal-to-noise ratio and yields accurate depth-estimates [17] but, in turn, limits the spatial resolution of the devices. This explains the poor performance of heuristic detection-methods, when applied to TOF camera calibration. For example, the amplitude signal from a typical TOF camera [18,19] resembles an ordinary greyscale image, but is of very low spatial resolution (e.g. 176×144 for the SR4000 camera, or 160×120 for the PMD PhotonICs on-chip sensor), as well as being noisy. A 712×496 CMOS colour-depth sensor is currently being developed, but the resolution of the TOF image delivered by this sensor is only 356×248 pixels [20]. A 340×96 pixels TOF camera has also been developed, for driving

☆ This paper has been recommended for acceptance by Antonis Argyros.

* Corresponding author. Fax: +33 476 61 54 54.
E-mail address: Radu.Horaud@inria.fr (R. Horaud).

http://dx.doi.org/10.1016/j.cviu.2014.01.007
1077-3142/

applications [21].[1] Lindner et al. [16] used the calibration module included in the OpenCV library [14] to estimate the parameters of a 200×200 PMD depth-camera and noticed a high dependency between the intrinsic and extrinsic parameters. To overcome these issues, a calibration method that combines a TOF camera with colour cameras was proposed [2,16]. While this method yields very accurate parameters, it requires a multiple-sensor setup composed of both TOF and standard cameras.

Calibration grids are essentially composed of two pencils of lines, therefore chequerboard detection should explicitly take this structure into account. For instance, the method in [22,23] starts by extracting points of interest, followed by eliminating those points that do not have a local chequerboard pattern, and finally by grouping together points lying along lines. This method puts a lot of emphasis on interest points, which are difficult to detect in low-resolution images, and does not take full advantage of the global structure of the calibration grid.

Two families of mutually orthogonal lines may also be detected by finding a dominant pair of vanishing-points. In [24] it is proposed to represent image lines on the Gaussian sphere (a unit sphere around the optical centre of the camera). Under perspective projection, an image line projects onto a great circle on the Gaussian sphere, and a pencil of lines corresponds to a family great circles that intersect at antipodal points (see for example Fig. 1 in [25]). Therefore, a vanishing point may be found by detecting the intersections, provided that the camera's internal parameters are known. Vanishing point detection was implemented using a quantized Gaussian sphere and a hierarchical (scale-space) Hough method, e.g. [26]. In general, Gaussian sphere-based methods require the detection of edges or of straight lines which are then projected as point sets (circles) on an azimuth-elevation grid, which may also produce spurious vanishing points [25].

More recently, vanishing-point detection was addressed as a clustering problem in the parameter space, using maximum likelihood and the EM algorithm [27], which requires suitable initialization. Alternatively, parameter-space clustering can be implemented using minimal sets [28] and random sampling. A method that combines [28] with an EM algorithm was recently proposed to find the three most orthogonal pencils of lines in indoor and outdoor scenes [29]. We tested this method using the software provided by the author[2] but found that the algorithm was not able to reliably extract and label edges from the low-resolution TOF amplitude images. We conclude that vanishing point methods, e.g., [29,30] fail to extract pencils of lines because they require accurate edge detection that is difficult to accomplish in low resolution, noisy images.

The method described in this paper is also based on the Hough transform [31], but it effectively fits a specific model to the chequerboard pattern, e.g., Fig. 1. This process is much less sensitive to the resolution of the data, for two reasons. Firstly, information is integrated across the source image, because each vertex is obtained from the intersection of two fitted lines. Secondly, the structure of a straight edge is inherently simpler than that of a corner feature. However, for this approach to be viable, it is assumed that any lens distortion has been pre-calibrated, so that the images of the pattern contain straight lines. This is not a serious restriction, because it is relatively easy to find enough boards (by any heuristic method) from which to obtain adequate estimates of the internal and lens parameters. Indeed there exist lens-calibration methods that require only a single image [32–34]. The harder problems of reconstruction and relative orientation can then be addressed after adding the newly detected boards, ending with a bundle-adjustment that also refines the initial internal parameters. Furthermore, the TOF devices used here have fixed lenses, which are sealed inside the camera body. This means that the internal and lens-distortion parameters from previous calibrations can be re-used.

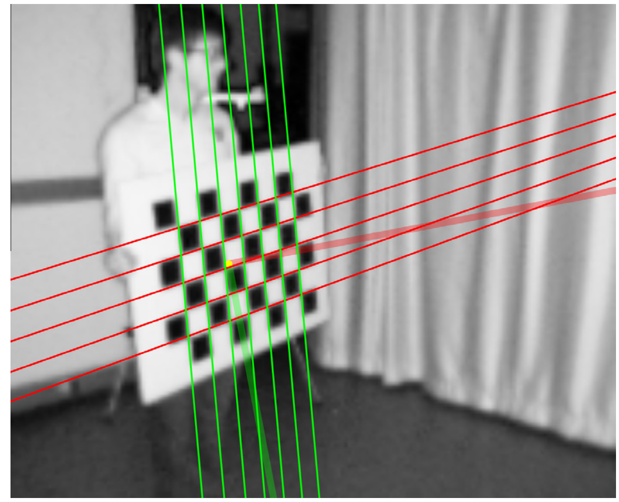

Fig. 1. An example of a TOF amplitude image taken in the infrared range in a normal neon-lit room. The original 176×144 image has been magnified and smoothed for display. The OpenCV software is unable to detect the vertices of the chequerboard in this case and in many other images as the board is viewed in distant and slanted positions. The red and green lines are the initial pencils detected by the proposed Hough-based method (prior to the refinement described in Section 3). The thick lines are the local Hough coordinate-systems, which are automatically established, as described in Section 2.3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Another Hough-method for chequerboard detection has been presented by de la Escalera and Armingol [35]. Their algorithm involves a *polar* Hough transform of all high-gradient points in the image. This results in an array that contains a peak for each line in the pattern. It is not, however, straightforward to extract these peaks, because their location depends strongly on the unknown orientation of the image-lines. Hence all local maxima are detected by morphological operations, and a second Hough transform is applied to the resulting data in [35]. The true peaks will form two collinear sets in the first transform (cf. Section 2.4), and so the final task is to detect two peaks in the second Hough transform. This iteration makes it hard to determine an appropriate sampling scheme, and also increases the computation and storage time of the procedure [36].

The method described in this paper is quite different. It makes use of the gradient *orientation* as well as magnitude at each point, in order to establish an axis-aligned coordinate system for each image of the pattern. Separate Hough transforms are then performed in the x and y directions of the local coordinate system. By construction, the slope-coordinate of any line is close to zero in the corresponding *Cartesian* Hough transform. This means that, on average, the peaks occur along a fixed axis of each transform, and can be detected by a simple sweep-line procedure. Furthermore, the known $\ell \times m$ structure of the grid makes it easy to identify the optimal sweep-line in each transform. Finally, the two optimal sweep-lines map directly back to pencils of ℓ and m lines in the original image, owing to the Cartesian nature of the transform. The principle of the method is shown in Fig. 3.

It should be noted that the method presented here was designed specifically for use with TOF cameras. For this reason, the *range*, as well as intensity data are used to help segment the image in Section 2.1. However, this step could easily be replaced with an appropriate background subtraction procedure [14], in which case

[1] Neither of these two cameras are commercially available at the time of writing.
[2] http://www-etud.iro.umontreal.ca/tardifj/fichiers/VPdetection-05-09-2010.tar.gz.