Neurocomputing 167 (2015) 44–51

Contents lists available at ScienceDirect

Neurocomputing

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

Local iterative DLT soft-computing vs. interval-valued stereo calibration and triangulation with uncertainty bounding in 3D reconstruction

José Otero*, Luciano Sánchez

Computer Science Department, Oviedo University, 33204 Gijón, Spain

ARTICLE INFO

Article history:
Received 20 March 2014
Received in revised form
17 October 2014
Accepted 9 November 2014
Available online 9 May 2015

Keywords:
Computer vision
Stereo
Camera calibration
Interval-valued
Possibility

ABSTRACT

The use of stereo vision for 3D data gathering is affected by constraints in the position of the cameras, the quality of the optical elements and the numerical algorithms for calibration and matching. Also, there is not a wide agreement on the best procedure for bounding the 3D errors within an uncertainty volume. In this work, this problem is solved by implementing the whole set of computations, including calibration and triangulation, with interval data. This is in contrast with previous works that rely on Direct Linear Transform (DLT) as a camera model. To keep better with real lens aberrations, a local iterative modification is proposed that provides an on-demand set of calibration parameters for each 3D point, comprising those nearest in 3D space. In this way, the estimated camera parameters are closely related with camera aberrations at the lens area through which that 3D point is imaged. To further reduce the triangulation uncertainty volume, a Soft Computing approach is proposed that represents each 3D point uncertainty as a cloud of crisp points compatible with interval-valued calibration data.

Real data from previous works in related research areas is used to judge whether the new approach improves the precision and accuracy of other crisp and interval-valued estimations without degrading precision, and it is concluded that the new technique is able to significantly improve the uncertainty volumes.

1. Introduction

There is a wide range of practical problems where a portable 3D data gathering system is needed. To name a few, there are field measurements where the equipment must be carried to site, or industrial measurements of objects whose position varies, for instance vehicles in MOT tests [1]. The major obstacles to non-laboratory 3D data gathering include cost, availability and the lack of suitable instrumentation [2]. According to [3], only optical scanners can recover curved surfaces, and these devices have a high cost. Alternative devices for 3D data capture may involve constraints on positioning of the necessary instruments that are not suitable in the field.

Because of this, techniques from photogrammetry and computer vision are seldom applied to such a class of problems, with a few exceptions. For example, in [4] stereo vision is used to reconstruct three-dimensional biological forms. The authors conclude that the precision of stereo vision methods improves photogrammetry by a large margin. Moreover, they assert that the lack of precision in some of their experiments was mainly related to two causes: (a) a less precise camera calibration and (b) the quality of the camera. To this it may be added that a rigorous procedure for estimating the accuracy of stereo vision-based measurements has still to be established.

The problems found by these authors are not exclusive to bioinformatics. Generally speaking, computer Vision Systems deal with measurement devices that, to a certain extent, are subjected to different kinds of errors. There are many works in this area that propose different models to understand how light refracts through camera lenses. The simplest is a plain projective model called pinhole [5]. The estimation of the parameters of this model or "calibration" consists in capturing images of an object with several landmarks at known positions. Let the image coordinates of a 3D point be $w=(x,y,z)$, and let the projection of this point at the image be $m=(u,v)$; if a model f with parameters $p_1 \dots p_n$ holds for this data, then $f(w,(p_1 \dots p_n))=m$. If the model parameters are unknown at least n equations are needed, obtained from n correspondences between 3D points and 2D points. The standard procedure uses more than n point correspondences, in order to partially cancel measurement errors, from which an over-constrained system of equations is obtained.

* Corresponding author.
E-mail addresses: jotero@uniovi.es (J. Otero), luciano@uniovi.es (L. Sánchez).

http://dx.doi.org/10.1016/j.neucom.2014.11.087
0925-2312/

Real lenses depart from the ideal pinhole camera model and thus, more elaborate models are needed in order to compute where a 3D point is projected on an image when optical aberrations are taken into account. Some good sources of information about available general purpose models are books such as [6,7]. Classical models are Tsai [8], Zhang [9] and (with several additional parameters that account for different kinds of lens imperfections) Direct Linear Transform [10]. More recently, new models have been developed, intended for lenses that are far apart from ideal ones, like fisheye lenses [11–13]. Finally, some papers are devoted to the so-called "generalized camera calibration" framework [14,15,12].

When a Computer Vision System comprises several cameras, each one calibrated using a suitable method, 3D point coordinates can be obtained from at least two 2D projections, one from each camera, as in stereo or trinocular systems [16]. This process is known as resection or triangulation. In some sense, the model of each camera must be inverted and the subset of 3D space that projects on each 2D point must be identified. Of course, from the geometry of the system, some additional constraints can be obtained that help to solve the problem of matching points between images (also called correspondence) [17], that is, finding pairs of points (one from each camera) that are a 2D projection from the same 3D point.

An important issue when dealing with a Computer Vision System is the assessment of the accuracy of the measurements provided. There are some sources of imprecision that lead to measurements with unknown error bounds. This work proposes an improvement of [18] that produces smaller uncertainty volume, gives information about the orientation of that volume and is able to handle real optics. Prior work and open problems are detailed in Section 2. In Section 3 the proposed approach is shown. Some numerical results are provided in Section 4 and in Section 5 conclusions and future work are mentioned.

2. Problem statement and related work

Digital cameras have several sources of error and uncertainty:

- Lens imperfections lead to image distortions, blurriness, chromatic aberration and other defects. Usually camera models account for some of these issues, but others are intrinsic to optics.
- Digital images are discrete by definition. Intensity values at each pixel and color channel are also sampled, usually along an evenly distributed grid in the x and y axes. It is commonly assumed that the center of each pixel is the 2D projection of a given 3D point, however, this is not entirely true.
- Camera sensor noise is noticed mostly when illumination conditions are not optimal. Due to its stochastic nature, it is not possible to fully remove noise from images.

There are several works in the literature that try to minimize the impact of these sources of error in computer vision systems. In recent years, a plethora of models that try to fit even the most extreme kinds of optics (like fisheye lenses) have emerged [11–13]. A totally different approach is to provide a generic framework for camera calibration decoupled from the nature of the lenses like in [14,15,19,12]. The effect of digital image quantization has been covered usually in connection with stereo, in works such as [20–23] or (to a lesser extent) [1]. The third source of error, image acquisition noise, has received some attention from the Computer Vision community in works like [24] or [25].

None of the previously mentioned works tackle the problem of *bounding* the errors that propagate from all the reported sources to the final goal of a Stereo Computer Vision System: the 3D position estimation of a visible point in the scene. A few examples that try to give some information about the tolerance of 3D measurements with stereoscopic systems are [26,27]. To our knowledge [18] is the only work where an interval-valued approach is proposed. Our work is based on this approach and will be thoroughly explained below.

2.1. Interval-valued estimation of the tolerance

In [18] both camera calibration and stereo triangulation (3D coordinates recovery from at least two pairs of matched 2D coordinates) are addressed as an interval-valued problem. The authors propose the use of a pinhole model in Eq. (1), where $w=(x,y,z)^T$ are 3D coordinates, $m=(u,v)^T$ 2D or image coordinates and $p_i, i=1\ldots3$ a 3×4 full-rank matrix, as shown in Eq. (2)

$$\begin{cases} u=\dfrac{p_1^T w+p_{1,4}}{p_3^T w+p_{3,4}} \\ v=\dfrac{p_2^T w+p_{2,4}}{p_3^T w+p_{3,4}} \end{cases} \tag{1}$$

$$P=\begin{pmatrix} p_1^T & p_{1,4} \\ p_2^T & p_{2,4} \\ p_3^T & p_{3,4} \end{pmatrix} \tag{2}$$

The values of this matrix are obtained using Eq. (3) where there are 11 parameters ($p_{3,4}=1$). In this equation one pair (u_i,v_i) (2D coordinates) is related with a triplet $w_i=(x,y,z)$ (3D coordinates), that is, for each 2D to 3D correspondence, two equations are obtained. From this it follows that at least six 2D to 3D correspondences are needed in order to obtain a overconstrained system of equations. Usually more than six correspondences will be used, in order to overcome measurement errors

$$\begin{pmatrix} w_i^T & 1 & 0 & 0 & -u_i w_i^T \\ 0 & 0 & w_i^T & 1 & -v_i w_i^T \end{pmatrix} \begin{pmatrix} p_1 \\ p_{1,4} \\ p_2 \\ p_{2,4} \\ p_3 \end{pmatrix} = \begin{pmatrix} u_i \\ v_i \end{pmatrix} \tag{3}$$

A stereoscopic system comprises two cameras, then the previous procedure, applied to both cameras, yields two different parameter matrices, P and P'. If two 2D coordinates (u^1,v^1) and (u^2,v^2), one from each image, are known to be the projection of the same (but with unknown coordinates) 3D point w, then replacing the known 2D coordinates and camera parameters for both images in Eq. (1) leads to the system of equations in (4) with only three unknowns, the 3D coordinates w, related with the known camera parameters and the 2D coordinates from both images. This system can be solved using any suitable overconstrained system solution method

$$\begin{pmatrix} (p_1^1-up_3^1)^T \\ (p_2^1-vp_3^1)^T \\ (p_1^2-u^2p_3^2)^T \\ (p_2^2-v^2p_3^2)^T \end{pmatrix} w = \begin{pmatrix} -p_{1,4}^1+u^1p_{3,4}^1 \\ -p_{2,4}^1+v^1p_{3,4}^1 \\ -p_{1,4}^2+u^2p_{3,4}^2 \\ -p_{1,4}^2+v^2p_{3,4}^2 \end{pmatrix} \tag{4}$$

In the same work, the authors consider that 2D pixel coordinates are discrete and thus that a rounding error of ±0.5 pixels may occur, that is, pixels are rectangles instead of crisp points. If this is translated to Eq. (3), it is obvious that an interval-valued system of linear equations is obtained and thus, P values will also be intervals. The authors state that the numerical solution estimation of interval-valued systems of equations using the approach in [28] is the tightest *interval-valued* one for this problem. The