Automated thermal 3 D reconstruction based on a robot equipped with uncalibrated infrared stereovision cameras

Thierry Sentenac, Florian Bugarin, Benoit Ducarouge, Michel Devy

Automated thermal 3*D* reconstruction based on a robot equipped with uncalibrated infrared stereovision cameras

T. Sentenac[a,b,*], F. Bugarin[b], B. Ducarouge[a], M. Devy[a]

[a] *LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France*
[b] *Institut Clément Ader (ICA), Université de Toulouse, CNRS, Mines Albi, UPS, INSA, ISAE-SUPAERO, Campus Jarlard, F-81013 Albi CT Cedex 09, France*

A B S T R A C T

In many industrial sectors, Non Destructive Testing (NDT) methods are used for the thermomechanical analysis of parts in assemblies of engines or reactors or for the control of metal forming processes. This article suggests an automated multi-view approach for the thermal 3*D* reconstruction required in order to compute 3*D* surface temperature models. This approach is based only on infrared cameras mounted on a Cartesian robot.

The low resolution of these cameras associated to a lack of texture to infrared images require to use a global approach based first on an uncalibrated rectification and then on the simultaneous execution, in a single step, of the dense 3*D* reconstruction and of an extended self-calibration.

The uncalibrated rectification is based on an optimization process under constraints which calculates the homographies without prior calculation of the Fundamental Matrix and which minimizes the projective deformations between the initial images and the rectified ones.

The extended self-calibration estimates both the parameters of virtual cameras that could provide the rectified images directly, and the parameters of the robot. It is based on two criteria evaluated according to the noise level of the infrared images. This global approach is validated through the reconstruction of a hot object against a reference reconstruction acquired by a 3*D* scanner.

Keywords:
3*D* reconstruction
Hand-eye calibration
Rectification
Vision
Robotic
Infrared cameras
Uncalibrated cameras

1. Introduction

This article addresses the problem of a fully automated 3*D* thermal reconstruction [1–3] from sensors embedded on a robotic system. Such a method can be suitable for performing diagnostics on mechanical assemblies, such as nuclear reactors [4], or for improving energy efficiency in building construction [5] or for monitoring forming processes [6]. The first sub-problem that arises is to define the system architecture using heterogeneous sensors. The second sub-problem is selecting the dense 3*D* reconstruction methods with overlaid thermal data. The third sub-problem is to make the inspection task totally automatic with a self-extended calibration (i.e. a calibration without a specific target, covering all the geometric parameters of the robot-sensor system, including the intrinsic and extrinsic sensor parameters and the robot parameters).

The most conventional architecture is based on a 3*D* laser scanner and infrared cameras mounted on a robot [7–9]. To overcome the significant cost of the 3*D* laser scanner, several articles [5,10,11] have suggested an architecture using only cameras. Inexpensive and readily available digital visible cameras (CCD camera, color camera, Kinect, etc.) give images processed by a 3*D* modeler, while infrared cameras provide the thermal data mapped on the 3*D* model. One successful system is the HeatWave system [3,12], i.e. a hand-operated device consisting of rigidly attached infrared and color cameras. These multi-sensory architectures face the difficulty of fusing 3*D* data provided by a 3*D* modeler and temperature data acquired by infrared cameras into a common coordinate frame. A joint geometric calibration of heterogeneous sensors [13] must be performed, which requires finding a pattern that is completely visible by both the 3*D* sensors and the IR cameras. This could be a tricky task, because these heterogeneous sensors have different spectral sensitivities, spatial resolutions and fields of view. The ideal architecture is then only based on infrared cameras for a direct 3*D* thermal reconstruction. Assuming that thermal methods already described in [14–16] are not in the scope of this paper, the challenge is then to provide a dense image-based 3*D* reconstruction [17,18] with infrared cameras.

Several image-based 3*D* reconstruction algorithms have been proposed using visible cameras. The first step, image registration, requires the detection and matching of features between images. Many feature detectors (e.g., Harris, SIFT, SURF, FAST, ORB, …) automatically and

* Corresponding author at: CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France.
E-mail addresses: sentenac@laas.fr (T. Sentenac), florian.bugarin@univ-tlse3.fr (F. Bugarin), benoit.ducarouge@laas.fr (B. Ducarouge), michel.devy@laas.fr (M. Devy).

correctly extract and match interest points on infrared images. Then two classes of methods can simultaneously build a sparse 3*D* model and the camera trajectory using only these matched interest points. The Robotics community has developed several Vision-based Simultaneous Localization And Mapping (VSLAM) techniques [19,20], taking advantage of other proprioceptive data acquired from the robot (odometry, IMU…), but assuming generally that the intrinsic camera parameters are known. The Vision community has proposed Structure from Motion (SfM) [21] approaches (Bundler, OpenMVG…) in order to recover from an image sequence both the 3*D* environment structure and the camera Motion; extrinsic and intrinsic camera parameters can be estimated simultaneously when computing the 3*D* point positions. The recovered parameters should be consistent with the reprojection error (i.e., the sum of distances between the projections of each set of 3*D* corresponding feature points and its corresponding image features). This minimization problem can be formulated as a non-linear least squares problem and solved from a Bundle Adjustment (BA) algorithm [22]. Exploiting VSLAM or SfM methods, an accurate and dense 3*D* model could be incrementally and gradually built and refined with, typically, a sequence of one thousand images, either from an Iterative Closest Point (ICP) algorithm (stereovision) or by a Multi-View Stereo (MVS) [23] technique (monocular vision).

The paper proposes an automated thermal 3*D* reconstruction based on an architecture composed of a Cartesian robot equipped only with uncalibrated infrared cameras. The architecture requires a coupled method that deals simultaneously with a multi-view 3*D* thermal reconstruction and a self-extended geometric calibration. An infrared stereo vision rig provides a compensation to the lower spatial resolution of infrared images. It also improves the number of reconstructed points and thus the density of the 3*D* model. Moreover, it gives an initial guess for the 3*D* position of every point. For the first step of the method, a reasonable amount of stable and tractable matched points is obtained through a specific method for infrared images based on the phase congruency model [24,25] which is combined with more classical feature detectors. With few and low-textured infrared images, the result would be limited to a sparse 3*D* reconstruction. Next, the simultaneous reconstruction and self-calibration with uncalibrated infrared cameras is solved by the minimization of a cost function which integrates all the geometrical calibration parameters for both the cameras and the robot. Estimations variables are: four intrinsic parameters for each camera, six for the relative position and orientation between the two cameras and six for the rotational and the translational components of the Euclidean transformation. This latter transformation is named hand-eye, between hand (robot gripper) and eye (camera). The total number of parameters is twenty if it is assumed that geometrical distortions due to the lens are corrected beforehand. This assumption is a good trade-off between the accuracy and the computation time. Indeed, it decreases the accuracy, but it avoids additional degrees of freedom and high non-linearities in the geometric model which are consuming in computation times. For the targeted application in a robotic context, a real-time processing of the calibration, compatible with the speed of the robot, is preferred even if the accuracy is not optimal. Finally, these parameters have to be estimated on-line from features extracted and matched from images acquired on the unknown object from two or more positions of the robot. These positions are known accurately in the robot reference frame from its forward kinematics.

A projective rectification method is first introduced for improving the matching problem between pixels on the left and right images, limiting the search space from the whole image to only a line. Another objective is to reduce the number of parameters of the cost function. The projective rectification method, applied to uncalibrated stereovision infrared cameras, requires a specific algorithm. Two homographies, applied to rectify the left and right images, have to be estimated in a single step to cope with the problem of noisy and low-textured infrared images. A new cost function is proposed which is minimized under geometric constraints by a non-linear optimization process. These constraints are defined to keep the structure and the skewness of the images, which are essential to preserving their geometries.

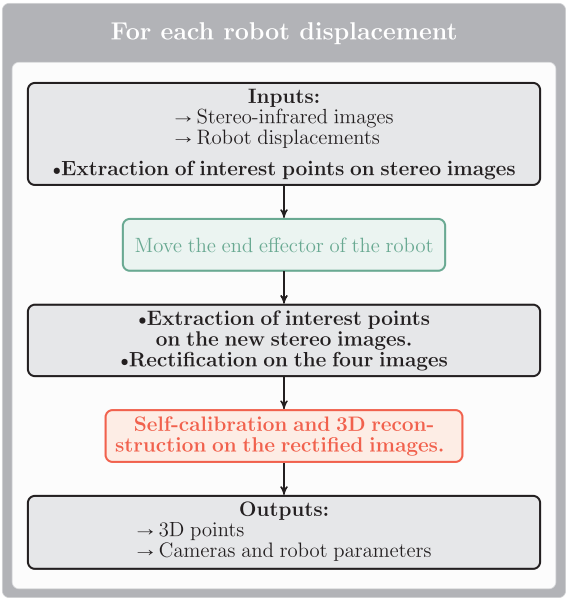


Fig. 1. Flowchart of the suggested method.

The set of points matched from rectified images is then used to perform simultaneously the extended self-calibration and the multi-view infrared 3*D* reconstruction. It is based on the minimization of two different functions depending on the observed noise level in infrared images. The first objective function is based on the minimization of the reprojection errors. When the noise level increases, a second objective function is expressed in the projective space which takes advantage of the epipolar constraint between two images. This second objective function is expressed using the intrinsic camera parameters and the essential matrix between two robot positions which depends itself on the hand-eye parameters and the robot motion. Fig. 1 summarizes the suggested coupled method which merges an extended self-calibration and a multi-view infrared 3*D* reconstruction applied on rectified images.

The paper is organized as follows. Section 2 briefly describes existing literature on rectification and places our suggested method with uncalibrated infrared images in this context. The method is then described and the results are compared to those in the literature. Section 3 outlines the formulation of the multi-view infrared 3*D* reconstruction simultaneous to the extended self-calibration. The method is evaluated on synthetic data. Finally, Section 4 summarizes the results of the fully automated 3*D* reconstruction performed from multiple views acquired by an uncalibrated infrared stereo rig mounted on a Cartesian robot. The whole approach is evaluated comparing the 3*D* model of a reconstructed object with a reference *CAD* model.

2. Suggested rectification method of uncalibrated cameras

The suggested rectification method, applied to uncalibrated stereovision cameras, takes the advantage of calculating the homographies, projective transformations applied to rectify images, without a previous calibration and in a single step to cope with the problem of noisy and low-textured infrared images. These homographies are then calculated with only one non-linear optimization process under geometric constraints.

The section begins with a short description of the background for calculating the homographies and works related to the rectification problem. The suggested projective rectification method is then

introduced and the non-linear optimization process under geometric constraints is detailed. These constraints are defined to keep the structure and the skewness of the images, which are essential to preserving their geometries. This property is essential for the self-calibration introduced in the next section. Finally, the method is evaluated by comparison with conventional methods that only work on visible images to prove its effectiveness even under these conditions.

2.1. Background and works related to the rectification problem

The rectification process reduces the two-dimensional matching problem on stereo images to a one-dimension matching problem. Using epipolar geometry, it consists of aligning the epipolar lines to make them parallel to the horizontal axis of the image. For uncalibrated cameras, knowledge of the epipolar geometry is packaged in the Fundamental matrix. The problem is then the computation of two projective transformations (homographies) from the Fundamental matrix to align the epipolar lines parallel to the horizontal image axis. The following paragraphs detail the computation of homographies and discuss the approach used to rectify uncalibrated images.

2.1.1. Background to epipolar geometry

Epipolar geometry defines the geometry between a pair of images $\mathscr{I}$ and $\mathscr{I}'$ from two stereoscopic cameras or two different locations of a mobile camera. Let Q, [1] a 3D point, be simultaneously seen by two pinhole cameras in 3D space. Let $\mathbf{c}$ and $\mathbf{c}'$ the optical centers of these two cameras. Let $\mathbf{q}$ and (resp. $\mathbf{q}'$) the projections of Q through $\mathbf{c}$ and $\mathbf{c}'$ in images $\mathscr{I}$ (resp. and $\mathscr{I}'$). Q, $\mathbf{c}$ and $\mathbf{c}'$ define an epipolar plan in 3D space, denoted by $\mathscr{P}$. The left epipolar line $\mathbf{l}_{\mathbf{q}'}$ (resp. right epipolar line $\mathbf{l}'_{\mathbf{q}}$) in $\mathscr{I}$ (resp. $\mathscr{I}'$) is defined by the intersection of $\mathscr{P}$ and $\mathscr{I}$ (resp. $\mathscr{P}$ and $\mathscr{I}'$). By geometric construction, $\mathbf{q}$ has to be on the right epipolar line $\mathbf{l}'_{\mathbf{q}}$ (and resp. $\mathbf{q}'$ on the left epipolar line $\mathbf{l}_{\mathbf{q}'}$). This constraint is the epipolar constraint: for a given point $\mathbf{q} \in \mathscr{I}$, its corresponding point $\mathbf{q}' \in \mathscr{I}'$ lies on its epipolar line $\mathbf{l}'_{\mathbf{q}}$, i.e. $\mathbf{q}'^{\top}\mathbf{l}'_{\mathbf{q}} = 0$. Similarly $\mathbf{q}$ lies on the epipolar line $\mathbf{l}_{\mathbf{q}'}$, i.e. $\mathbf{q}^{\top}\mathbf{l}_{\mathbf{q}'} = 0$. Because the relationships between retinal coordinates of corresponding points $(\mathbf{q}, \mathbf{q}')$ and their epipolar lines $(\mathbf{l}'_{\mathbf{q}}, \mathbf{l}_{\mathbf{q}'})$ are projective linear, the epipolar constraint can be rewritten as follows:

$$\mathbf{q}'^{\top}\,\mathbf{F}\;\mathbf{q} = 0. \tag{1}$$

where $\mathbf{F}$ is Fundamental Matrix ($\mathbf{F} \in \mathscr{M}_3(\mathbb{R})$). It encapsulates the projective motion between two uncalibrated perspective cameras. The epipoles $\mathbf{e}$ and $\mathbf{e}'$ are points which satisfy the following equation:

$$\mathbf{F}\mathbf{e} = \mathbf{F}^{\top}\mathbf{e}' = 0_{\mathbb{P}^2}. \tag{2}$$

The epipole $\mathbf{e}$ (resp. $\mathbf{e}'$) is the intersection of all the epipolar lines included in $\mathscr{I}$ (resp. $\mathscr{I}'$). Moreover, Eq. (2) implies that the rank of $\mathbf{F}$ is lower or equal to two. $\mathbf{F}$ is then defined up to a scale factor. It theoretically depends upon seven independent parameters. More details about epipolar geometry and the fundamental matrix are provided in the book [21].

2.1.2. Computation of homographies

Using epipolar geometry, the rectification provides corresponding epipolar lines which are collinear and parallel with the x-axis. The rectified fundamental matrix $\mathbf{F}_0$ is then expressed as follows:

$$\mathbf{F}_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}.$$

This process is accomplished by applying two homographies $\mathbf{H}$, $\mathbf{H}'$ on $\mathscr{I}$ and $\mathscr{I}'$. These homographies map the epipoles $\mathbf{e}$ and $\mathbf{e}'$ to points at infinity. Hence, $\mathbf{H}$ and $\mathbf{H}'$ transform the matched points $(\mathbf{q}, \mathbf{q}') \in \mathscr{I} \times \mathscr{I}'$ to $(\widetilde{\mathbf{q}}, \widetilde{\mathbf{q}}') \in \widetilde{\mathscr{I}} \times \widetilde{\mathscr{I}'}$ as:

$$\widetilde{\mathbf{q}} = \mathbf{H}\mathbf{q}; \quad \widetilde{\mathbf{q}}' = \mathbf{H}'\mathbf{q}' \tag{3}$$

It follows from Eq. (1) that:

$$\widetilde{\mathbf{q}}'^{\top}\mathbf{F}_0\widetilde{\mathbf{q}} = \mathbf{q}'^{\top}\underbrace{\mathbf{H}'^{\top}\mathbf{F}_0\mathbf{H}}_{=\mathbf{F}}\mathbf{q} = 0. \tag{4}$$

Finally, $\mathbf{H}$ and $\mathbf{H}'$ are compatible homographies only if they satisfy the following equation:

$$\mathbf{F} = \mathbf{H}'^{\top}\mathbf{F}_0\mathbf{H}. \tag{5}$$

The practical computation of $\mathbf{H}$ and $\mathbf{H}'$ is then achieved by solving (5). However, due to the first row of $\mathbf{F}_0$, the pair of homographies is not unique. This remaining degree of freedom can introduce an undesirable distortion to the rectified images. Methods suggested in the literature to reduce distortions can be divided into two classes. The first class assumes that $\mathbf{F}$ is fixed and the second class $\mathbf{F}$ is implicitly recomputed. The first class is divided into two steps. A preliminary accurate estimation of $\mathbf{F}$ and a calculation of $(\mathbf{H}, \mathbf{H}')$ are first introduced from Eq. (5). The distortions are then corrected by applying symmetrical [26,27] or independent [28–32] matrix on $(\mathbf{H}, \mathbf{H}')$. The second class proceeds only in a single step. The fundamental matrix $\mathbf{F}$ is directly recomputed by replacing $\mathbf{F}$ by $\widetilde{\mathbf{F}} = \mathbf{H}'^{\top}\mathbf{F}_0\mathbf{H}$ and by solving the following minimization problem:

$$\min_{\mathbf{H},\mathbf{H}'} \sum_{k=1}^{N} d(\mathbf{q}'_k, \widetilde{\mathbf{F}}\mathbf{q}_k) + d(\widetilde{\mathbf{F}}^{\top}\mathbf{q}'_k, \mathbf{q}_k). \tag{6}$$

where $(\mathbf{q}, \mathbf{q}')_{k=1\ldots N}$ is a set of N matched points between the images $\mathscr{I}$ and $\mathscr{I}'$. $d(\mathbf{q}'_k, \widetilde{\mathbf{F}}\mathbf{q}_k)$ (resp. $d(\widetilde{\mathbf{F}}^{\top}\mathbf{q}'_k, \mathbf{q}_k)$) is the distance from the point $\mathbf{q}'_k$ (resp. $\mathbf{q}_k$) to the epipolar line $\widetilde{\mathbf{F}}\mathbf{q}_k$ (resp. $\widetilde{\mathbf{F}}^{\top}\mathbf{q}'_k$). However, this minimization alone is not enough to overcome the distortions problem which can be fixed thanks to the specific parametrization of the homographies $\mathbf{H}$ and $\mathbf{H}'$ [33,34]. These parametrizations go hand in hand with assumptions on the parameters of $\mathbf{F}$. For instance, in articles [33,34], the authors assume that the principal point is centered and the aspect ratio is equal to one (so that the skew is equal to zero).

2.2. Formulation of the uncalibrated rectification

The suggested formulation of the uncalibrated rectification is consistent with the second class of rectification methods, which works in only one step to take as fully as possible into account the low number of points matched in infrared images (images with little texture and a weak spatial resolution). Iteration of many steps could propagate and amplify the uncertainty due to low numbers of matched points. The suggested method then consists in computing the Fundamental matrix $\mathbf{F}$ and the rectification homographies $\mathbf{H}$ and $\mathbf{H}'$ in a single step without prior assumptions on the geometry and on the parametrization of $\mathbf{H}$ and $\mathbf{H}'$. The contribution is then to achieve the calculation of homographies by applying the non-linear objective function of Eq. (6) under constraints to minimize the geometrical distortions induced by $\mathbf{H}$ and $\mathbf{H}'$. Introduced by [31] and commonly used in the literature [33,34], criteria based on the aspect ratio and the orthogonality of the images are applied after the rectification process to control loss or pixels creation. The main idea of the suggested computation of the homographies is to include these classical criteria as constraints of minimization process in order to keep the structure and the skewness of the image. The constraint space is then based on the aspect and size ratio and the orthogonality of the image which are invariant to affine transformations [27].

2.2.1. Definition of the non-linear objective function

The distance d from a point $\mathbf{q}'$ to its corresponding epipolar line $\widetilde{\mathbf{F}}\mathbf{q}$ of Eq. (6) is defined in $\mathbb{R}^2$ as follows:

[1] Note that to improve the readability of Sections 2 and 3, sans-serif font upper-case is used for 3D points (e.g. Q) while bold lower-case is used to denote 2D points (e.g. $\mathbf{q}$).

$$d(\mathbf{q}', \widetilde{\mathbf{F}}\mathbf{q}) = \frac{|\mathbf{q}'^{\top}\widetilde{\mathbf{F}}\mathbf{q}|}{\|\pi(\widetilde{\mathbf{F}}\mathbf{q})\|_2^2} \tag{7}$$

with π: $(x_1, x_2, x_3) \rightarrow (x_1, x_2)$ the canonical projection. Conversely, the distance for a point $\mathbf{q}$ to its corresponding epipolar line $\mathbf{F}\mathbf{q}'$ can also be defined.

Eq. (6) is then rewritten symmetrically on both images as follows:

$$\min_{P \in C} \sum_{k=1}^{N} \frac{|\mathbf{q}_k'^{\top}\widetilde{\mathbf{F}}\mathbf{q}_k|^2}{\|\pi(\widetilde{\mathbf{F}}\mathbf{q}_k)\|_2^2 + \|\pi(\widetilde{\mathbf{F}}^{\top}\mathbf{q}_k')\|_2^2} \tag{8}$$

where $P = (\mathbf{H}, \mathbf{H}') \in \mathbb{R}^9 \times \mathbb{R}^9 \simeq \mathbb{R}^{18}$ is the vector of the nine parameters of $\mathbf{H}$ and the nine parameters of $\mathbf{H}'$. This vector is minimized on a constraint space $C \subset \mathbb{R}^{18}$ to find a pair of homographies reducing the geometric distortions. The notation of only one constraint space, denoted by C, actually unifies and merges heterogeneous space constraints (i.e. with different units) applied either on $\mathscr{I}$ or on $\mathscr{I}'$. C is then the intersection of several constraint spaces which are detailed in the next subsection.

2.2.2. Space constraints definition

The need to minimize distortion requires keeping the image structure and the skewness. Hence, each pixel of the original image should map to a single pixel in the rectified images (pixel creation and loss should be minimal). In the article [29] the creation or the loss of pixels is modeled by the change in local area of a patch around the point before and after the rectification. This change can be determined by the numerical properties of the Jacobian homography. These numerical properties can be controlled by matrix operators based on the determinant [30] or on singular values [31]. Thus, minimizing these operators involves reducing the distortion in the whole image and their implementation then becomes very time-consuming in terms of computation. Our approach is to define a set of numerical stable constraints C which only enforces the physical properties of the image. The image distortions are then defined as modifications of the image structure: aspect ratio, size (width and height) and orthogonality of the image.

Aspect ratio criterion. The constraints of the image structure can be first quantified by the invariance of the ratio of image diagonals computed from the aspect ratio between original and rectified images. Ideally, the aspect ratio is equal to 1 (see Fig. 2(a)). The four image corners are sufficient to compute this criterion. Let $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4\}$ and $\{\mathbf{p}_1', \mathbf{p}_2', \mathbf{p}_3', \mathbf{p}_4'\}$ be the corners of the images $\mathscr{I}$ and $\mathscr{I}'$. In the rectified space, let $\{\widetilde{\mathbf{p}}_1, \widetilde{\mathbf{p}}_2, \widetilde{\mathbf{p}}_3, \widetilde{\mathbf{p}}_4\}$ and $\{\widetilde{\mathbf{p}}_1', \widetilde{\mathbf{p}}_2', \widetilde{\mathbf{p}}_3', \widetilde{\mathbf{p}}_4'\}$ be the corners of rectified images. The *aspect ratio* E_a and the *aspect ratio constraint* C_a are then defined by:

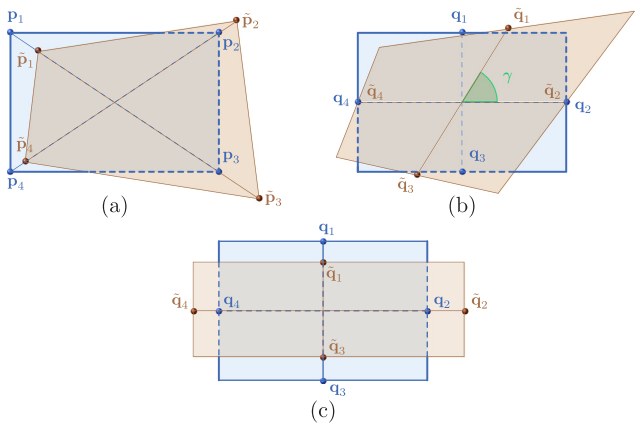


Fig. 2. Criteria summary. Image (a) shows deformations due to an aspect ratio different to 1. Image (b) displays deformations due to an orthogonality criterion different to 90°. Image (c) displays deformations due to size ratio different to 1.

$$\begin{aligned} E_a: \mathbb{R}^{18} &\rightarrow \mathbb{R}^2 \\ P &\rightarrow \left(\frac{\|\widetilde{\mathbf{p}}_1 - \widetilde{\mathbf{p}}_3\|_2}{\|\widetilde{\mathbf{p}}_2 - \widetilde{\mathbf{p}}_4\|_2}, \frac{\|\widetilde{\mathbf{p}}_1' - \widetilde{\mathbf{p}}_3'\|_2}{\|\widetilde{\mathbf{p}}_2' - \widetilde{\mathbf{p}}_4'\|_2} \right) \end{aligned} \tag{9}$$

$$C_a = \{P \in \mathbb{R}^{18} \mid E_a(P) \in [1-\epsilon, 1+\epsilon]^2\} \tag{10}$$

where ϵ is the tolerant variation of the aspect ratio.

Orthogonality criterion. The angles invariance between original and rectified images can be modeled by the orthogonality which is ideally equal to 90° (see Fig. 2(b)). Article [27] considers that applying constraints on the four side mid-points preserves the orthogonality of the image. With the same previous notations, *orthogonality criterion* E_o and the *orthogonality constraint* C_o are then defined by:

$$\begin{aligned} E_o: \mathbb{R}^{18} &\rightarrow \mathbb{R}^2 \\ P &\rightarrow (\text{acos}(\gamma), \text{acos}(\gamma')) \end{aligned} \tag{11}$$

$$C_o = \{P \in \mathbb{R}^{18} \mid E_o(P) \in [90°-\theta, 90° + \theta]^2\} \tag{12}$$

where γ, γ' are the angle defined by:

$$\gamma = \frac{\|\widetilde{\mathbf{q}}_1 - \widetilde{\mathbf{q}}_3\|^2 + \|\widetilde{\mathbf{q}}_2 - \widetilde{\mathbf{q}}_4\|^2 - \|(\widetilde{\mathbf{q}}_1 - \widetilde{\mathbf{q}}_3) - (\widetilde{\mathbf{q}}_2 - \widetilde{\mathbf{q}}_4)\|^2}{2\|\widetilde{\mathbf{q}}_1 - \widetilde{\mathbf{q}}_3\|\|\widetilde{\mathbf{q}}_2 - \widetilde{\mathbf{q}}_4\|} \tag{13}$$

$$\gamma' = \frac{\|\widetilde{\mathbf{q}}_1' - \widetilde{\mathbf{q}}_3'\|^2 + \|\widetilde{\mathbf{q}}_2' - \widetilde{\mathbf{q}}_4'\|^2 - \|(\widetilde{\mathbf{q}}_1' - \widetilde{\mathbf{q}}_3') - (\widetilde{\mathbf{q}}_2' - \widetilde{\mathbf{q}}_4')\|^2}{2\|\widetilde{\mathbf{q}}_1' - \widetilde{\mathbf{q}}_3'\|\|\widetilde{\mathbf{q}}_2' - \widetilde{\mathbf{q}}_4'\|} \tag{14}$$

and θ is the tolerant variation of the orthogonality.

Size ratio criteria. The invariance of the width and the height of images between original and rectified images can be only computed by a set of points defined by the side mid-points of the horizontal side for image $\mathscr{I}$ (resp. $\mathscr{I}'$), $\mathbf{q}_1$, $\mathbf{q}_3$ (resp. $\mathbf{q}_1'$, $\mathbf{q}_3'$) and vertical image side, $\mathbf{q}_2$, $\mathbf{q}_4$ (resp. $\mathbf{q}_2'$, $\mathbf{q}_4'$). In the rectified images, the corresponding points are defined as: $\{\widetilde{\mathbf{q}}_1, \widetilde{\mathbf{q}}_2, \widetilde{\mathbf{q}}_3, \widetilde{\mathbf{q}}_4\}$ and $\{\widetilde{\mathbf{q}}_1', \widetilde{\mathbf{q}}_2', \widetilde{\mathbf{q}}_3', \widetilde{\mathbf{q}}_4'\}$. The *size ratio* E_s and the *size ratio constraint* C_s are then defined by:

$$E_s: \mathbb{R}^{18} \rightarrow \mathbb{R}^4 \tag{15}$$

$$P \rightarrow \left(\frac{\|\mathbf{q}_1 - \mathbf{q}_3\|_2}{\|\widetilde{\mathbf{q}}_1 - \widetilde{\mathbf{q}}_3\|_2}, \frac{\|\mathbf{q}_1' - \mathbf{q}_3'\|_2}{\|\widetilde{\mathbf{q}}_1' - \widetilde{\mathbf{q}}_3'\|_2}, \frac{\|\mathbf{q}_2 - \mathbf{q}_4\|_2}{\|\widetilde{\mathbf{q}}_2 - \widetilde{\mathbf{q}}_4\|_2}, \frac{\|\mathbf{q}_2' - \mathbf{q}_4'\|_2}{\|\widetilde{\mathbf{q}}_2' - \widetilde{\mathbf{q}}_4'\|_2} \right)$$

$$C_s = \{P \in \mathbb{R}^{18} \mid E_s(P) \in [1-\delta, 1+\delta]^4\} \tag{16}$$

where δ the tolerant variation of the image size. Ideally, $E_s(P)$ is equal to 1 (see Fig. 2(c)).

Finally, the constraints space C is the intersection of the previous constraints, i.e. $C = C_a \cap C_s \cap C_o$. Although C merges heterogeneous constraints, the optimization method used to solve Eq. (8) considers each constraints separately and then deals with constraints of different units. The minimization of Eq. (8) is performed by a local optimization method which is the Sequential Quadratic Programming algorithm (S.Q.P.) [35–37]. A good initial estimate P_0 is then necessary to guarantee the convergence to a right solution. The choice of the constraints space C affects the ease of finding this initial estimate. Therefore, the nine initial parameters of the homographies have to satisfy the constraints C. The identity homographies are the simplest initial parameter vector that satisfies the constraints. Indeed, the image without bias (distortion) that satisfies the constraints is the original image. Moreover, the identity has the property (that we verified experimentally on all our image tests) of being situated in a basin of attraction of a minimum of physically practicable premises.

2.3. Validation experiment

The suggested rectification method was evaluated on images from the Mallon's test set [31] which are taken by the same camera under a fixed lens configuration. Each set (Arch, Boxes and Drive) consists of two RGG images with 640 × 480 pixels resolution. The lens distortion has been removed and the ground truth is not available. Results carried out by the suggested rectification method were compared to Monasse's method [33], which also works in a single step. This article [33] also

Table 1
Comparison of the suggested rectification method with Monasse's method of the article [33].

Sample	Method	Orthogonality E_o		Aspect Ratio E_a		Rectification error E_r	
		$\mathbf{H}'$ (°)	$\mathbf{H}$ (°)	$\mathbf{H}'$ (no unit)	$\mathbf{H}$ (no unit)	*mean* (pix.)	σ (pix.)
Boxes	Proposed	89.44	90.66	0.9897	1.0105	0.11	0.06
	Monasse	89.60	89.63	0.9884	0.9892	0.1293	0.0887
Arch	Proposed	89.67	89.88	0.9955	0.9983	0.17	0.16
	Monasse	89.80	90.05	0.9942	1.0014	0.2520	0.2349
Drive	Proposed	90.61	89.32	1.0128	0.9876	0.55	0.44
	Monasse	89.95	90.00	0.9977	1.0001	0.7139	0.8253

compares the results of other conventional methods implemented by Loop and Zhang [27], Hartley [29] and Mallon [31].

2.3.1. Evaluation criteria

The rectification performance is concerned with quantifying the rectification error E_r. For each point of one image, it is the distance between its corresponding point and epipolar line in the other image. It is represented by the mean (*mean*) and the standard deviation (σ). Then, the other criteria are the aspect ratio E_a and the orthogonality E_o. Ideally, the aspect ratio E_a must be 1 and the orthogonality E_o must be 90°.

2.3.2. Results

Table 1 gives the relative performance on the three set of images (Arch, Boxes and Drive) from the Mallon's dataset and the comparison with Monasse's method.

The values of the orthogonality E_o (respectively the aspect ratio E_a) criterion are similar for both methods and close to the reference value of 90° (respectively of 1). The image structure and the skewness are well preserved and remain invariant. The rectified images can be applied in a geometric calibration process. The suggested method provides a smaller rectification error E_r (mean and standard deviation) than the reference method (notably for the Drive set). The method is based on no prior assumptions on the parametrization of the homographies $\mathbf{H}$ and $\mathbf{H}'$. No geometry is assumed and all parameters of homographies can be estimated with all possible values. However, at the same time the estimation is constrained to preserve the structure of the image and the generality of the solution. Finally, it should be emphasized that the identity provides a good initial estimate and convergence to a relevant solution.

3. Suggested extended self stereovision hand-eye calibration

The extended self hand-eye calibration with one uncalibrated camera is described with ten parameters: four intrinsic parameters of the camera (focal length, optical center), and six ones for the rotational and the translational component of the Euclidean transformation between hand and eye. With a stereovision rig, ten parameters must be added: four intrinsic parameters for the second camera, and six for the relative position and orientation between the two cameras. The previous rectification step decreases the number of parameters to be estimated, from twenty to thirteen (six intrinsic parameters for cameras, six parameters for the rigid hand-eye transformation and one for the rigid rectified stereovision configuration). It is assumed that image distortions are neglected. The estimation problem of the high number of parameters of the extended self stereovision hand-eye calibration is highly relevant in the presence of noise in the matching points between each camera and each displacement.

A first paragraph deals with the background and related works of hand-eye calibration. The second paragraph introduces the suggested extended self stereovision hand-eye calibration which is a simultaneous calibration of hand-eye, camera-intrinsic and stereo parameters. In the third paragraph, the problem is formulated as two non-linear optimization processes in relation to the noise on the matching points. In the last paragraph, experimental results on synthetic data show the influence of the noise on the accuracy of the 3*D* reconstruction.

3.1. Background and works related to hand-eye calibration

Hand-eye calibration is the computation of the unknown transformation, named $\mathbf{X}$ between the camera frame and the frame of the hand of the robot. The hand-eye problem is solved knowing the displacement of the camera, named $\mathbf{A}$, in a reference frame and the displacement of the hand of the robot, named $\mathbf{B}$, between two positions i and j.

Let $\mathscr{R}_{\mathbf{c}}^{j}$ (resp. $\mathscr{R}_{\mathbf{c}'}^{j}$) be the coordinates system associated to $\mathbf{c}$ (resp. $\mathbf{c}'$) at the j^{th} position and $\mathscr{R}_w$, the world coordinates system associated to the initial position of the robot. The hand-eye calibration then consists in estimating the parameters of the homogeneous Euclidean transformation, $\mathbf{X}$, between the robot (hand) and the camera (eye) coordinate frame. The main assumption is that the single camera is rigidly coupled to the robot. As shown in Fig. 3, the robot performs discrete displacements from the position i to the position j, which are encoded through the euclidean transformation, $\mathbf{B}_{ij}$. The displacements are also measured by the camera $\mathbf{c}$ as a transformation $\mathbf{A}_{ij}$ in its own coordinate frame. The parameters of hand-eye transformation, $\mathbf{X}$, are then calculated according to the graph of coordinate frames of Fig. 3 and by solving the following equation:

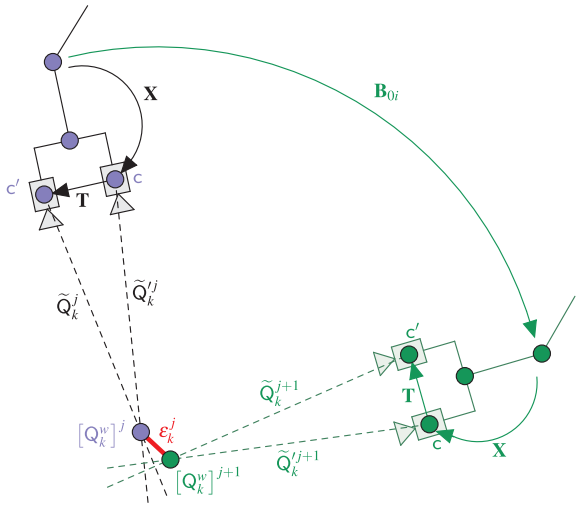


Fig. 3. Graph of transformations. Transformation matrix $\mathbf{B}_{ij}$ displaces the coordinates system of the robot, transformation matrix $\mathbf{A}_{ij}$ displaces the coordinates of the camera $\mathbf{c}$, hand-eye transformation $\mathbf{X}$ displaces the coordinates system of the robot (hand) on coordinates of the camera (eye).

$$\mathbf{A}_{ij} = \mathbf{X}\mathbf{B}_{ij}\mathbf{X}^{-1} \tag{17}$$

By referring rigid transformations by 4×4 homogeneous matrices composed of a rotation and a translation $(\mathbf{R}, \mathbf{t})$, Eq. (17) is rewritten as follows:

$$\mathbf{R}_{\mathbf{A}_{ij}} = \mathbf{R}_{\mathbf{X}}\mathbf{R}_{\mathbf{B}_{ij}}\mathbf{R}_{\mathbf{X}}^{\top} \tag{18}$$

$$\mathbf{t}_{\mathbf{A}_{ij}} = \mathbf{R}_{\mathbf{X}}\mathbf{t}_{\mathbf{B}_{ij}} - \Big(\underbrace{\mathbf{R}_{\mathbf{X}}\mathbf{R}_{\mathbf{B}_{ij}}\mathbf{R}_{\mathbf{X}}^{\top}}_{\mathbf{R}_{\mathbf{A}_{ij}}} - \mathbf{I}\Big)\,\mathbf{t}_{\mathbf{X}} \tag{19}$$

Several methods exist to solve the hand-eye calibration problem with different optimization processes. A general taxonomy of the classical approaches is provided in article [38] and in the focus of our contribution, they can be split into two classes: non-unified and unified methods. Using the decomposition (18), the non-unified methods separately estimate the translation $\mathbf{t}_{\mathbf{X}}$ and the rotation $\mathbf{R}_{\mathbf{X}}$. Several parametrizations can be used to estimate the rotation: the angle-axis parametrization of the group of rotations $SO(3)$ [39,40], quaternions [41] or dual quaternions [42]. Moreover, with noisy inputs, according to investigations carried out in the article [43], unified methods with a simultaneous estimation of both rotation and translation are highly relevant to balance the estimation of the rotation part of $\mathbf{X}$. It has also been shown in the article [42], that camera intrinsic parameters are not independent from hand-eye parameters. The extended unified methods thus include both hand-eye and camera parameters [44] (e.g. camera pose, intrinsic parameters and lens distortion [45,46]) which are simultaneously estimated. The optimization process for solving Eq. (18) is linear for unified methods and non-linear for non-unified methods. According to prior knowledge of the external camera calibration, the conventional non-linear optimization process is based on the norm two: $\|\mathbf{A}_{ij}\mathbf{X} - \mathbf{X}\mathbf{B}_{ij}\|_2$. The most recent non-linear optimization processes use branch-and-bound [47–49], Second Order Cone Programming [50] or polynomial optimization [51] techniques. For extended methods, all parameters are mainly simultaneously estimated with the Structure-from-Motion (SfM) technique. In the article [52] camera pose parameters are recovered up to a scale factor and in articles [42,53,54] they are explicitly included. Branch-and-bound algorithms have recently been introduced to estimate camera poses with a L_∞-norm formulation [48] and with a hypothesis that assumes all the translations are equal to zero. For extended methods with several parameters [45], the bundle adjustment technique is the most promising approach to retrieve all parameters simultaneously.

3.2. Extended self stereovision hand-eye calibration suggested for uncalibrated infrared cameras rig

For uncalibrated stereovision infrared cameras, the extended self stereovision hand-eye calibration suggested thus relies on a generalized extended unified method. Indeed, the suggested method performs simultaneously, and in a single step, the extended self-calibration and the multi-view infrared $3D$ reconstruction. The first paragraph recalled the parameters of the calibration after the previous rectification step, which are estimated using a bundle adjustment technique. The second paragraph introduces two objective functions to solve the double problem of calibration and reconstruction. The first one aims at minimizing the reprojection errors (in pixels). The second one is expressed in the projective space, from the epipolar constraints between two images acquired by the same camera from two robot positions. These two functions make explicit all the parameters of the transformations (intrinsic camera parameters, hand-eye parameters and robot motion parameters). Finally, the third paragraph validates the extended self stereovision hand-eye calibration using simulated data generated with different noise levels. The last paragraph concludes with the best objective function according to the noise level.

3.2.1. Parameters of extended self stereovision hand-eye calibration

As illustrated in Fig. 3, the rigid body transformation between $\mathscr{R}_{\mathbf{c}}^{j}$ and $\mathscr{R}_{\mathbf{c}'}^{j}$ is denoted by $\mathbf{T} = [\mathbf{R}\ \mathbf{t}]$. Given a $3D$ point $\mathsf{Q}^w \in \mathscr{R}_w$, its projections $\mathbf{q}^j$ and $\mathbf{q}'^j$ in $\mathscr{I}^j$ and $\mathscr{I}'^j$ after a displacement $\mathbf{B}_{0j}$ of the robot from the initial position are:

$$\mathbf{q}^j = \mathbf{K}\mathbf{X}\mathbf{B}_{0j}\mathbf{X}^{-1}\mathsf{Q}^w \tag{20}$$

$$\mathbf{q}'^j = \mathbf{K}'\mathbf{T}\mathbf{X}\mathbf{B}_{0j}\mathbf{X}^{-1}\mathsf{Q}^w \tag{21}$$

where the matrix $\mathbf{T}$ is reduced to a single translation over the horizontal axis for a rectified stereovision rig: $\mathbf{T} = [\mathbf{I}_{3\times 3}\ \mathbf{t}]$ with $\mathbf{t} = (t\ 0\ 0)^{\top}$. t is the baseline of the rectified stereovision rig. $\mathbf{K}$ and $\mathbf{K}'$ are the matrices of intrinsic parameters which are given by:

$$\mathbf{K} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}; \quad \mathbf{K}' = \begin{pmatrix} \alpha'_u & 0 & u'_0 \\ 0 & \alpha'_v & v'_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{22}$$

where u_0, v_0 (resp. u'_0, v'_0) represent the centers of images $\mathscr{I}$ (resp. $\mathscr{I}'$), α_u, α_v (resp. α'_u, α'_v) represent the number of pixels per millimeter for $\mathscr{I}$ (resp. $\mathscr{I}'$). For an uncalibrated rectified stereovision rig, the two rectified images are coplanar, abscissa are collinear and the epipolar lines are aligned between the two images. Consequently, $v_0 = v'_0$ and $\alpha_v = \alpha'_v$.

Finally, projection functions are fully described with only thirteen parameters: six intrinsic parameters (α_u, α'_u, α_v, u_0, u'_0, v_0), the baseline t and six for the hand-eye transformation (three for the rotation and three for the translation).

3.2.2. Discussion on the definition of the objective functions

Different approaches can be exploited to design the objective functions.

The first approach is inspired by the estimation pose problem in which the $3D$ location of observed points is estimated in terms of reprojection errors. The approach operates the geometry of the stereo image sequences. The objective function minimizes the reprojection errors (see articles [55,56]) between the measured points by the camera and points from the projection model (see Eq. (20)) in a recursive bundle adjustment. However, the objective function is symmetric to calculate the two distances in each image and the number of parameters involved is high. Instead of minimizing the reprojection error, it is more convenient to calculate the reconstruction error of $3D$ points built from the stereo image sequence between two robot positions. Each $3D$ point is expressed in the robot coordinate frame versus the intrinsic, the stereo and hand-eye parameters and the displacement of the robot. The estimation of intrinsic, stereo and hand-eye parameters is then performed in a single step starting from approximate initial guesses.

The second approach is based on the epipolar constraint geometry of the stereo image pairs and the stereo images sequences. It has some theoretical drawbacks compared to the bundle adjustment, in that it does not provide as much information and cannot achieve the same level of accuracy. However, the first benefit is a smaller parameter space. Moreover, methods based on epipolar geometry do not involve the $3D$ position of the observed object point, i.e. the epipolar constraint decouples the extrinsic camera parameters from the $3D$ structure of the observed object. Eq. (2) only depends on the $2D$ coordinates image. Eq. (20) additionally requires the depth of each observed point. If the matching point is not accurate, the error is not propagated in the $3D$ space. The epipolar constraint is then calculated between two images acquired by the same camera from two robot positions. This constraint is expressed using the Fundamental matrix, i.e. using the intrinsic camera parameters and the essential matrix which itself depends on the hand-eye parameters and the robot displacement. This second suggested objective function is thus expressed using the intrinsic camera parameters and the essential matrix between two robot positions. All the parameters are also estimated in a single step.

Based on these different approaches, our approach combines both

the geometry of the stereo image sequences and the epipolar constraint geometry of the stereo image pairs with two objective functions being derived in relation to the noise level of the images. Given noisy image points, the minimization of these objective functions provides intrinsic, stereo and hand-eye parameters and starts from approximate initial guesses. The main contribution of the suggested extended self stereo-vision hand-eye calibration is the proposal of two objective functions which depend on the noise level on the infrared images.

3.3. Definition of the two objective functions

3.3.1. Objective function from the geometry of the stereo image sequences

Given a set of 3D points $(\mathrm{Q}_k^w)_{k=1..n}$ in $\mathscr{R}_w$. It is assuming that all the points are seen by a stereo rig displaced using m rigid transformation $(\mathbf{B}_{0j})_{j=1..m}$. For an easy reading of this section, we denote $\mathbf{B}_j = \mathbf{B}_{0j},\ \forall\, j = 1..m$. Consider a static 3$D$ point $\mathrm{Q}_k^w \in \mathscr{R}_w$ from the initial position of the robot. It can be expressed in $\mathscr{R}_{\mathbf{c}}^j$ using the following equation:

$$\mathrm{Q}_k^j = \mathbf{X}\mathbf{B}_j\mathrm{Q}_k^w. \tag{23}$$

However, the localization of Q_k^w remains the same for two successive displacements:

$$\mathbf{B}_{j+1}^{-1}\mathbf{X}^{-1}\mathrm{Q}_k^{j+1} = \mathbf{B}_j^{-1}\mathbf{X}^{-1}\mathrm{Q}_k^j, \tag{24}$$

The aim is then the minimization of the 3D locations (in mm) for the successive displacements:

$$\min_{\mathbf{X}} \sum_{j=1}^{m-1}\sum_{k=1}^{n} \|\mathbf{B}_{j+1}^{-1}\mathbf{X}^{-1}\mathrm{Q}_k^{j+1} - \mathbf{B}_j^{-1}\mathbf{X}^{-1}\mathrm{Q}_k^j\|^2, \tag{25}$$

with $\mathbf{X}$ the vector of the hand-eye transformation. Recall that all the $\mathbf{B}_j$ matrices are known.

However, the problem expressed in equation (25) is highly over-determined. This equation can be rewritten by incorporating the intrinsic parameters of the two cameras, $\mathbf{K}$ and $\mathbf{K}'$, and the transformation between both cameras, $\mathbf{T}$.

The first step is the introduction of equality between $(\mathbf{q}_k^j, \mathbf{q}_k'^j)$, the projections of Q_k^w in the images taken at j^{th} displacement, which is given by the following equations:

$$\mathbf{q}_k^j = \mathbf{K}\mathbf{X}\mathbf{B}_j\mathrm{Q}_k^w \tag{26}$$

$$\mathbf{q}_k'^j = \mathbf{K}'\mathbf{T}\mathbf{X}\mathbf{B}_j\mathrm{Q}_k^w. \tag{27}$$

In the next step, Eq. (25) is rewritten so as to minimize the reprojection error (in pixels) as follows:

$$\min_{\mathbf{X},\mathbf{K},\mathbf{K}',\mathbf{T}} \sum_{j=1}^{m}\sum_{k=1}^{n} \|\mathbf{q}_k^j - \mathbf{K}\mathbf{X}\mathbf{B}_j\mathrm{Q}_k^w\|^2 + \|\mathbf{q}_k'^j - \mathbf{K}'\mathbf{T}\mathbf{X}\mathbf{B}_j\mathrm{Q}_k^w\|^2. \tag{28}$$

Note that solving Eq. (28) means knowing precisely the 3D locations of the set $(\mathrm{Q}_k^w)_k$. However, without a calibration object and without knowing the 3D points, the third step is the estimation of these 3D locations jointly to $\mathbf{X}$, $\mathbf{K}$, $\mathbf{K}'$ and $\mathbf{T}$ as follows:

$$\min_{\mathbf{X},\mathbf{K},\mathbf{K}',\mathbf{T},\mathrm{Q}_k^w} \sum_{j=1}^{m}\sum_{k=1}^{n} \|\mathbf{q}_k^j - \mathbf{K}\mathbf{X}\mathbf{B}_j\mathrm{Q}_k^w\|^2 + \|\mathbf{q}_k'^j - \mathbf{K}'\mathbf{T}\mathbf{X}\mathbf{B}_j\mathrm{Q}_k^w\|^2. \tag{29}$$

This formulation, known as bundle-adjustment, requires a initial estimate of $(\mathrm{Q}_k^w)_{k=1..n}$. The fourth step is to recover these 3D locations (unknown in our context) from the matched pixels by inverting Eq. (26). A 3D location Q_k^w is then seen as $[\mathrm{Q}_k^w]^j$, the intersection of two rays $\widetilde{\mathrm{Q}}_k^j$ and $\widetilde{\mathrm{Q}}_k'^j$ given by:

$$\widetilde{\mathrm{Q}}_k^j = \mathbf{R}_{\mathbf{B}_j}^\top\mathbf{R}_{\mathbf{X}}^\top\mathbf{K}^{-1}\mathbf{q}_k^j - \mathbf{R}_{\mathbf{B}_j}^\top\mathbf{R}_{\mathbf{X}}^\top\mathbf{t}_{\mathbf{X}} - \mathbf{R}_{\mathbf{B}_j}^\top\mathbf{t}_{\mathbf{B}_j} \tag{30}$$

$$\widetilde{\mathrm{Q}}_k'^j = \mathbf{R}_{\mathbf{B}_j}^\top\mathbf{R}_{\mathbf{X}}^\top\mathbf{R}_{\mathbf{T}}^\top\mathbf{K}'^{-1}\mathbf{q}_k'^j - \mathbf{R}_{\mathbf{B}_j}^\top\mathbf{R}_{\mathbf{X}}^\top\mathbf{R}_{\mathbf{T}}^\top\mathbf{t}_{\mathbf{T}} - \mathbf{R}_{\mathbf{B}_j}^\top\mathbf{R}_{\mathbf{X}}^\top\mathbf{t}_{\mathbf{X}} - \mathbf{R}_{\mathbf{B}_j}^\top\mathbf{t}_{\mathbf{B}_j}, \tag{31}$$

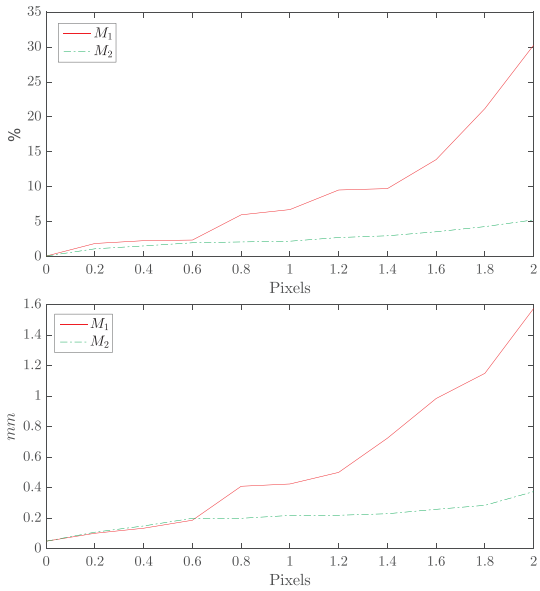


Fig. 4. Definition of the error ε_k^j (in red) minimized in Eq. (32). This error is defined by the distance between $[\mathrm{Q}_k^w]^j$ (the intersection of the rays – in black – $\widetilde{\mathrm{Q}}_k^j$ and $\widetilde{\mathrm{Q}}_k'^j$) and $[\mathrm{Q}_k^w]^{j+1}$ (the intersection of the rays – in green – $\widetilde{\mathrm{Q}}_k^{j+1}$ and $\widetilde{\mathrm{Q}}_k'^{j+1}$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Assuming that the localization of Q_k^w remains the same for all positions of the rig, the errors ε_k^j of two successive triangulations (see Fig. 4) are minimized in the following formulation:

$$\min_{\mathbf{X},\mathbf{K},\mathbf{K}',\mathbf{T}} \sum_{k=1}^{n} \varepsilon_k^j. \tag{32}$$

with $\varepsilon_k^j = \|[\mathrm{Q}_k^w]^j - [\mathrm{Q}_k^w]^{j+1}\|^2$ (see Fig. 4).

The optimization method, noted (M_1), provides an estimated vector p of thirteen parameters (6 intrinsic parameters ($\mathbf{K}$, $\mathbf{K}'$), the baseline ($\mathbf{T}$) and 6 parameters of the hand-eye transformation ($\mathbf{X}$)).

3.3.2. Objective function from the epipolar constraint

These parameters can be also estimated in the second optimization problem by using the epipolar geometry. As explained in Section 2, $\mathbf{q}_k^j$ lies on the epipolar line $\mathbf{l}'_{\mathbf{q}_k^{j+1}}$. The fundamental matrix $\mathbf{F}^j$ associated to camera $\mathbf{c}$ and camera $\mathbf{c}$ moved by applying the displacement $\mathbf{A}_{j,j+1}$ is estimated by solving:

$$\min_{\mathbf{F}^j} \sum_{k=1}^{n} \frac{|\mathbf{q}_k^{j+1\top}\mathbf{F}^j\mathbf{q}_k^j|^2}{\|\pi(\mathbf{F}^j\mathbf{q}_k^j)\|^2 + \|\pi(\mathbf{F}^{j\top}\mathbf{q}_k^{j+1})\|^2} \tag{33}$$

In the second step, $\mathbf{F}^j$ is decomposed as follows:

$$\mathbf{F}^j = \mathbf{K}^{-\top}\mathbf{E}^j\mathbf{K}^{-1} \tag{34}$$

where $\mathbf{E}^j$ is the transformation between the retinal plane of $\mathbf{c}$ and the retinal plane of $\mathbf{c}$ moved by applying the displacement $\mathbf{A}_j$. Then, the cost function of Eq. (33) can be rewritten as follows:

$$\sum_{j=1}^{m-1}\sum_{k=1}^{n} \frac{|\mathbf{q}_k^{j+1\top}\mathbf{K}^{-\top}\mathbf{E}^j\mathbf{K}^{-1}\mathbf{q}_k^j|^2}{\|\pi(\mathbf{K}^{-\top}\mathbf{E}^j\mathbf{K}^{-1}\mathbf{q}_k^j)\|^2 + \|\pi(\mathbf{K}^{-\top}\mathbf{E}^{j\top}\mathbf{K}^{-1}\mathbf{q}_k^{j+1})\|^2} \tag{35}$$

Note that the rigid transformation $[\mathbf{R}_{\mathbf{E}j}\ \mathbf{t}_{\mathbf{E}j}]$ extracted from the essential matrix $\mathbf{E}_j$ is strictly equal to $\mathbf{A}_{j,j+1}$:

$$[\mathbf{R}_{\mathbf{E}j}\ \mathbf{t}_{\mathbf{E}j}] = \mathbf{A}_{j,j+1} = \mathbf{X}\mathbf{B}_{j,j+1}\mathbf{X}^{-1} \tag{36}$$

Then the rotation $\mathbf{R}_{\mathbf{E}j}$ and the translation $\mathbf{t}_{\mathbf{E}j}$ can be deduced from Eq. (18). The same idea can be applied to the fundamental matrix $\mathbf{F}'_j$

associated to camera $\mathbf{c}'$ and camera $\mathbf{c}'$ moved by applying the displacement $\mathbf{A}_{j,j+1}\mathbf{T}$:

$$\mathbf{F}'^{j} = \mathbf{K}'^{-\top}\mathbf{E}'^{j}\mathbf{K}'^{-1} \tag{37}$$

$$[\mathbf{R}_{E'j}\,\mathbf{t}_{E'j}] = \mathbf{T}\mathbf{A}_{j,j+1}\mathbf{T}^{-1} \tag{38}$$

$$= \mathbf{T}\mathbf{X}\mathbf{B}_{j,j+1}\mathbf{X}^{-1}\mathbf{T}^{-1}. \tag{39}$$

Finally, the following optimization problem is stated:

$$\min_{\mathbf{X},\mathbf{K},\mathbf{K}',\mathbf{T}} \sum_{j=1}^{m-1}\sum_{k=1}^{n} \mathscr{E}_j^k + \mathscr{E}_j'^k. \tag{40}$$

with:

$$\mathscr{E}_j^k = \frac{|\mathbf{q}_k^{j+1\top}\mathbf{K}^{-\top}\mathbf{E}^j\mathbf{K}^{-1}\mathbf{q}_k^j|^2}{\|\pi(\mathbf{K}^{-\top}\mathbf{E}^j\mathbf{K}^{-1}\mathbf{q}_k^j)\|^2 + \|\pi(\mathbf{K}^{-\top}\mathbf{E}^{j\top}\mathbf{K}^{-1}\mathbf{q}_k^{j+1})\|^2}$$

$$\mathscr{E}_j'^k = \frac{|\mathbf{q}_k'^{j+1\top}\mathbf{K}'^{-\top}\mathbf{E}'^{j}\mathbf{K}'^{-1}\mathbf{q}_k'^{j}|^2}{\|\pi(\mathbf{K}'^{-\top}\mathbf{E}'^{j}\mathbf{K}'^{-1}\mathbf{q}_k'^{j})\|^2 + \|\pi(\mathbf{K}'^{-\top}\mathbf{E}'^{j\top}\mathbf{K}'^{-1}\mathbf{q}_k'^{j+1})\|^2}$$

The method, noted (M_2), minimizes Eq. (40) and exhibits sufficient conditions to estimate the thirteen parameters of $\mathbf{X}$, $\mathbf{K}$, $\mathbf{K}'$, and $\mathbf{T}$ (six for $\mathbf{K}$, $\mathbf{K}'$, one for the baseline $\mathbf{T}$ and six for the hand-eye transformation $\mathbf{X}$).

3.4. Validation of extended self stereovision hand-eye calibration

The experiments were carried out on synthetic images to evaluate and compare the accuracy and the robustness of the methods ($M1$) and ($M2$) against noisy images.

3.4.1. Generation of noisy synthetic images

For six given displacements ($\mathbf{B}_{0j}$), $j = 1, ..., 6$ (given by Table 2) between two successives images from the stereorig, for two given matrices $\mathbf{K}$ and $\mathbf{K}'$ of internal parameters (defined by Eq. (22) and Table 3) and for an hand-eye transformation $\mathbf{X}$ (given by Table 3), a set of $3D$ points $(\mathbf{Q}_i)_i$, $i = 1, ..., 100$ is projected into the two images as $(\mathbf{q}_i, \mathbf{q}'_i)_{i=1,...,100}$ using Eq. (20). At each coordinates pixel, a centered Gaussian noise with zero mean and standard deviation $\sigma_{\mathbf{q}}$ from 0 to 2 pixels is added. In order to have statistical evidence, the results are averaged over 100 trials. The averaged resulting noisy points $(\overline{\mathbf{q}}_i, \overline{\mathbf{q}}'_i)_{i=1,...,100}$ are used to estimate the performance of the two proposed methods.

3.4.2. Performances of methods M_1 and M_2

The minimization of Eqs. (32) and (40) was performed by a BFGS algorithm [57] with the initial guess vector composed of the initial intrinsic and baseline parameters and the initial parameters of the hand-eye transformation which are tabulated in Table 3.

The criteria for assessing the performance of the two methods are the relative error $\epsilon_x^r = \|\hat{x}-x\|/\|x\|$ and the absolute error $\epsilon_x^a = \|\hat{x}-x\|$ between true vector x and its estimation $\hat{x}$. These criteria are calculated after each simulation run. The quantities $\bar{\epsilon}_a^r$ or $\bar{\epsilon}_x^a$ denote the mean value over all simulation runs.

Table 2
Six displacements of the end-effector transformation, $\mathbf{B}_{0j} = [\mathbf{R}_{\mathbf{B}_{0j}}\,\mathbf{t}_{\mathbf{B}_{0j}}]$, used for the generation of synthetic images.

Displ.	rotation (°)			translation (mm)		
j	$\mathbf{r}_{\mathbf{B}_{0j_x}}$	$\mathbf{r}_{\mathbf{B}_{0j_y}}$	$\mathbf{r}_{\mathbf{B}_{0j_z}}$	$\mathbf{t}_{\mathbf{B}_{0j_x}}$	$\mathbf{t}_{\mathbf{B}_{0j_y}}$	$\mathbf{t}_{\mathbf{B}_{0j_z}}$
1	0	180	0	0	0	500
2	45	180	14	10	200	356
3	−43	−167	22	−318	−158	424
4	49	175	4	12	189	832
5	60	167	34	−321	376	294
6	−34	−163	−45	531	−269	756

Table 3
Intrinsic parameters of both rectified cameras, $\mathbf{K}$ and $\mathbf{K}'$, the baseline $\mathbf{t}$ and the parameters of Hand-Eye transformation, $\mathbf{X} = [\mathbf{R}_{\mathbf{X}}\,\mathbf{t}_{\mathbf{X}}]$, used for the generation of synthetic images (S.) and for the initial guess (I.) of the minimization process.

	α_u (–)	$\alpha_{u'}$ (–)	α_v (–)	u_0 (pix.)	u'_0 (pix.)	v_0 (pix.)	t (mm)
S.	195.84	191.61	201.58	137.39	26.95	58.58	−156.14
I.	200	200	200	82	82	64	−150
	$\mathbf{t}_{\mathbf{X}_x}$ (mm)	$\mathbf{t}_{\mathbf{X}_y}$ (mm)	$\mathbf{t}_{\mathbf{X}_z}$ (mm)	$\mathbf{r}_{\mathbf{X}_x}$ (°)	$\mathbf{r}_{\mathbf{X}_y}$ (°)	$\mathbf{r}_{\mathbf{X}_z}$ (°)	
S.	152	36	−19	6.53	4.89	−3.98	
I.	150	40	−20	0	0	0	

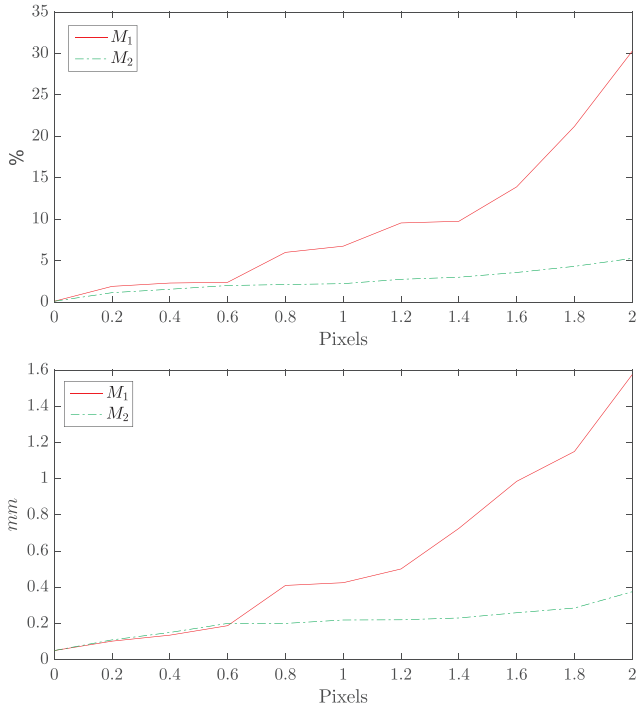


Fig. 5. Evolution of the mean value of $\bar{\epsilon}_{\mathbf{K}}^r$, the relative error of intrinsic parameters of camera $\mathbf{c}$, and ϵ_t^a, the relative error of translation parameter versus the standard deviation of noise for fits to Eqs. (32) and (40).

Fig. 5 gathers the mean value of the relative error of intrinsic parameters, $\bar{\epsilon}_{\mathbf{K}}^r$, and the absolute error of translation parameter ϵ_t^a.

With synthetic images with little noise ($\sigma_{\mathbf{q}} < 0.6$ pixels), the fit on intrinsic parameters with Method M_1 and Method M_2 are equivalents. In contrast, when the level of noise increases ($\sigma_{\mathbf{q}} > 0.6$ pixels), $\bar{\epsilon}_{\mathbf{K}}^r$ increases with method M_1 whereas it remains lower than 5% with Method M_2. The same behavior can be observed for the translation parameter t. The value of ϵ_t^a remains lower than 0.4 mm with Method M_1 and it rises above 1.5 mm with Method M_2 when $\sigma_{\mathbf{q}}$ increases from 0.6 to 2 pixels.

The mean value of the absolute error on translation and rotation hand-eye parameters, $\bar{\epsilon}_{\mathbf{t_X}}^a$ and $\bar{\epsilon}_{\mathbf{R_X}}^a$, shown in Fig. 6, confirms the previous trend. When the standard deviation of noise, $\sigma_{\mathbf{q}}$, is low, Method M_1 provides slightly lower values of $\bar{\epsilon}_{\mathbf{t_X}}^a$ and $\bar{\epsilon}_{\mathbf{R_X}}^a$. When the standard deviation of the noise, $\sigma_{\mathbf{q}}$, is higher than 0.6 pixels, Method M_2 performs significantly better than Method M_1 and provides a maximal value for $\bar{\epsilon}_{\mathbf{t_X}}^a$ (resp. $\bar{\epsilon}_{\mathbf{R_X}}^a$) of 4.5 mm (resp. of 0.2°).

Finally, it might seem that the two methods are equivalent in a context with a low standard deviation of noise ($\sigma_{\mathbf{q}}$) point extraction (less than 0.6 pixel). The Method M_2 is globally less sensitive to the noise level. In our application, the standard deviation of noise remains less than 0.6 pixel. The two methods can thus be used. However, the

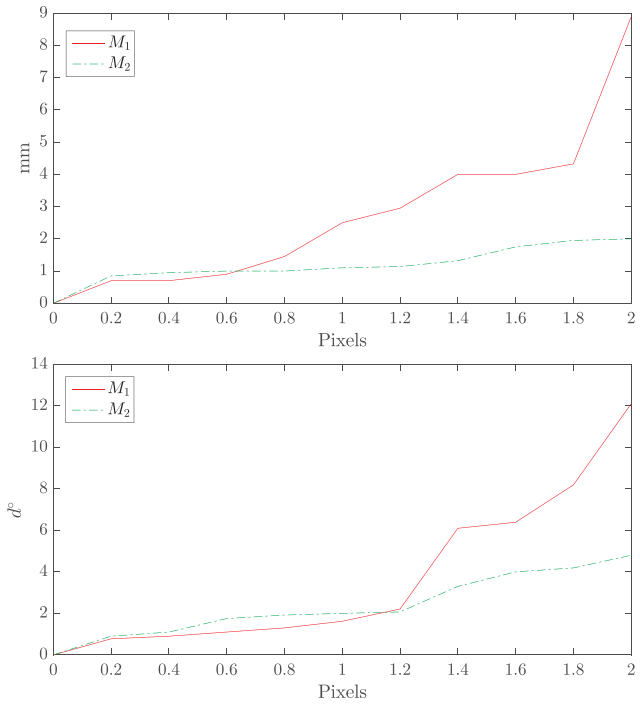


Fig. 6. Evolution of the mean value of the error of hand-eye translation parameters, $\bar{\epsilon}^a_{\mathbf{t_X}}$ (high figure) and hand-eye rotation parameters, $\bar{\epsilon}^a_{\mathbf{R_X}}$, (low figure) versus the noise parameter ν for fits to Eqs. (32) and (40).

advantage of Method M_1 is that it calculates the 3D points conjointly with the calibration parameters.

3.5. Conclusion

The first objective function (Method M_1) is very efficient with a low level of noise. When the noise level increases, it is necessary to introduce a second objective function in the projective space which takes advantage of the epipolar constraint between two images acquired by the same camera from two robot positions. However, Method M_1 is able to calculate the 3D points conjointly with the calibration parameters.

4. Automated thermal 3D reconstruction

The automated thermal 3D reconstruction will be implemented to monitor incremental forming processes [6]. Its principle is to locally and gradually deform a metal sheet using a hemispherical tip tool (small diameter compared to the dimensions of the sheet) until the desired shape. The trajectory of the tip tool is controlled by a numerical control machine. This process makes it possible to form sheets of geometries complex with an extremely simple and therefore inexpensive tooling. The automated multiview thermal 3D reconstruction method suggested aims at controlling both the local thermal gradient under the tip tool and guarantying the shape of the part during the forming. The section aims at testing and evaluating the shape measurement of a reference part, called pyramid, which is a test part of the incremental forming process.

The first section details the automatic NDT architecture composed by uncalibrated infrared cameras mounted on the Cartesian robot. The method, explained in the two previous sections, is summarized in the second subsection. The last subsection provides the results the 3D reconstruction of the reference part provided by the suggested thermal 3D reconstruction method. It also compares this 3D reconstruction with 3D reconstruction performed by a reference 3D Digitizer, Konica Minolta Range, with $\pm$ 50 μm accuracy.

4.1. Automatic NDT architecture for the automated thermal 3D reconstruction

The automatic NDT architecture is a Cartesian robot equipped only with stereovision bench composed of uncalibrated infrared cameras which is displayed in Fig. 7(a).

The Cartesian robot positions the end-effector (hand) at the desired position for the acquisitions in a working volume of 1000 × 1000 × 800 mm. The position of the end-effector is known and given by the calculation of euclidean transformations (translation and rotation, named $\mathbf{B}_{ij}$ in the previous section) between the world and the end-effector frames. A Cartesian robot offers a high rigidity (and so a good reproducibility of motion) and a good accuracy (thanks to the partial decoupling of the axes). The stereovision infrared cameras, installed at the end-effector, are compact and dedicated to embedded applications, such as drones. Their advantages are size and very low weight (around 400 g). These cameras are equipped with matrix sensors of uncooled microbolometers. They operate in 8–14 μm spectral band and work at 25 fps. Their resolution is 160 × 120 pixels with a pitch of 25 μm. The Noise Equivalent Temperature Difference (NETD) is approximately 100 mK. The focal length is 11 mm and acquisition distance is approximately 50 cm.

The automated multiview thermal 3D reconstruction is tested on the pyramid part (see Fig. 7(b)), manufactured by an incremental forming process. Its base is about thirty centimeters with sides around ten centimeters. Its height is twenty centimeters.

4.2. Overview of the process of automated thermal 3D reconstruction

As shown in Fig. 8, the first step of the automated thermal 3D reconstruction is the acquisition of multiviews from the robot displacement, so as to extract interest points on the part of the object to be reconstructed. Considering the displacements of the end effector of the robot expressed by the matrix $\mathbf{B}_j$ and $\mathbf{B}_{j+1}$, at the position j, the k

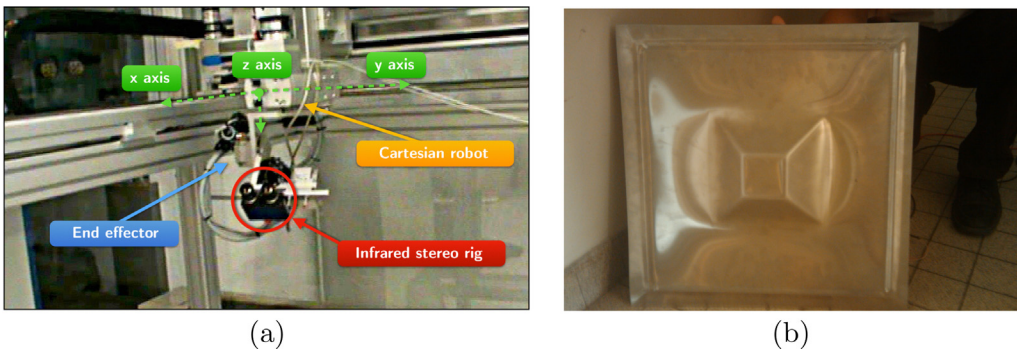


Fig. 7. (a) Thermal robotic set-up composed of a Cartesian robot and infrared cameras. (b) Reference part, called pyramid, used for testing the automated multiview thermal 3D reconstruction.

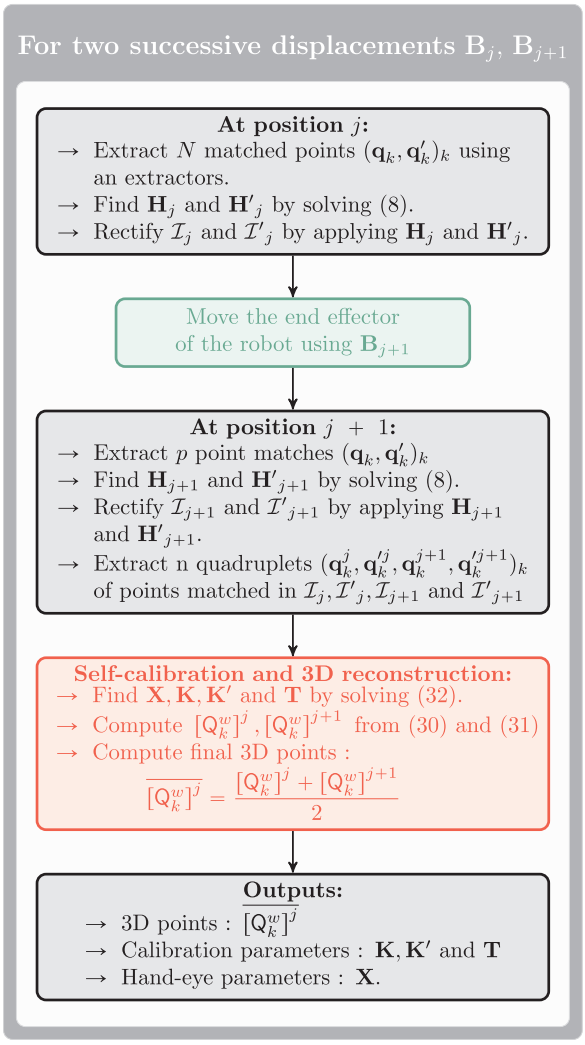


Fig. 8. Flowchart of the automated thermal 3*D* reconstruction.

matched points on each of the pairs of images and on each image between the displacement j and $j+1$, $(\mathbf{q}_k, \mathbf{q}'_k)_j$ are obtained with a combination of detectors.

Thus, the matched points on each of the pairs of images are the inputs to calculate the pair of rectification homographies of the stereovision bench which minimize the projective deformations according to the method described in Section 2. The two obtained homographies $\mathbf{H}^j$ and $\mathbf{H}^{j+1}$ are used to compute the two rectified images $\mathscr{I}^j$ and $\mathscr{I}'^j$. Next, the end effector is moved to the position $j+1$ using $\mathbf{B}^{j+1}$. At this new position, a new step of rectification is performed. Four rectified images $\mathscr{I}^j$, $\mathscr{I}'^j$, $\mathscr{I}^{j+1}$ and $\mathscr{I}'^{j+1}$ are then available. Edge detection is then performed on this set in order to obtain n quadruplets $(\mathbf{q}_k^j, \mathbf{q}_k'^j, \mathbf{q}_k^{j+1}, \mathbf{q}_k'^{j+1})_{k=1\ldots n}$ of matched points.

This set of quadruplets is used in the self-calibration and reconstruction step presented in Section 3 and to solve Eq. (32). Solving this problem provides the calibration parameters $\mathbf{K}$, $\mathbf{K}'$, $\mathbf{T}$, the hand-eye parameters $\mathbf{X}$ and two sets $([\mathrm{Q}_k^w]^j)_{k=1\ldots n}$, $([\mathrm{Q}_k^w]^{j+1})_{k=1\ldots n}$ of 3*D*-points. Note that the criterion of Eq. (32) is the sum of euclidean distance $\|[\mathrm{Q}_k^w]^j - [\mathrm{Q}_k^w]^{j+1}\|_2$. Consequently, the two sets of obtained 3*D*-points are very close at the end of the optimization process. Nevertheless, a choice must be made between the two sets. We define $(\overline{[\mathrm{Q}_k^w]^j})_{k=1\ldots n}$, the final set of 3*D* points, by the mean between $([\mathrm{Q}_k^w]^j)_{k=1\ldots n}$ and $([\mathrm{Q}_k^w]^{j+1})_{k=1\ldots n}$.

This process leads to the 3*D* partial reconstruction of the observed object summarized in Fig. 8.

4.3. Compared 3D reconstruction on the reference part

To assess the performances of the suggested approach, a 3*D* reconstruction comparison was carried out on the pyramid part presented in Fig. 7. The reference 3*D* reconstruction was performed using the 3*D* Digitizer, Konica Minolta Range. It measures the shape of the part by profilometry with ± 50 µm accuracy for a measurement distance of 60 cm.

The reference 3*D* reconstruction was compared to the 3*D* points cloud performed by the suggested automated thermal 3*D* reconstruction with $m = 15$ displacements of the robot. These fifteen displacements only provide a partial 3*D* reconstruction of the part.

4.3.1. Rectification of uncalibrated cameras

As presented in the flowchart of Fig. 8, the first step consists in matching N quadruplets $(\mathbf{q}_k^j, \mathbf{q}_k'^j, \mathbf{q}_k^{j+1}, \mathbf{q}_k'^{j+1})_{k=1\ldots N}$ of points on the stereo images acquired for the position j of the end-effector. The Harris's detector [58], coupled with ZNCC (Zero mean Normalized Cross-Correlation) method, provide the highest number of uniformly distributed matched points. Fig. 9a shows an example of the number of matched points ($N = 110$).

These matched points are the inputs of the rectification method suggested in Section 2. Fig. 9b displays the original and rectified images calculated with this set of matched points.

The criteria used to qualify the rectification image of Fig. 9 have already been presented in Section 2.3 and are tabulated in Table 4. The mean (respectively the standard deviation) on the rectification error E_r is very low: 0.09 pixels (respectively 0.03 pixels). The maximal orthogonality error $max(90° - E_o)$ (respectively the maximal aspect ratio error $max(1 - E_a)$) is 0.7° (respectively 0.02). These results, with an orthogonality error of less than one degree highlight a very small deformation of the image and will ensure the introduction of a small deformation in the next calibration step.

4.3.2. Extended self stereovision hand-eye calibration

As presented in the flowchart of Fig. 8, extended self stereovision hand-eye calibration is performed on $m = 15$ displacements of the robot. Between the position j and $j+1$ of the end effector, four rectified images $\mathscr{I}^j$, $\mathscr{I}'^j$, $\mathscr{I}^{j+1}$ and $\mathscr{I}'^{j+1}$ are computed in the step presented

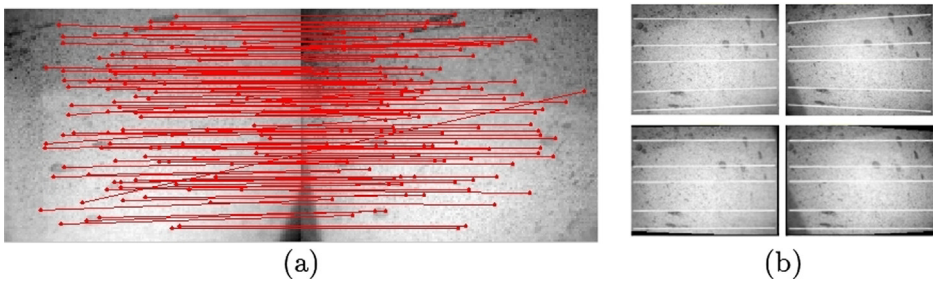


Fig. 9. Example of stereo infrared images of the pyramid part. (a) 110 Matched points between the stereo infrared images with Harris detector and ZNCC method. (b) Original (on the top) and rectified infrared images $\mathscr{I}^j$, $\mathscr{I}'^j$ (on the bottom).

Table 4
Rectification performances on the pyramid part.

Orthogonality E_o		Aspect Ratio E_a		Rectification error E_r	
$\mathbf{H}'$ (°)	$\mathbf{H}$ (°)	$\mathbf{H}'$ (no unit)	$\mathbf{H}$ (no unit)	*mean* (pix.)	σ (pix.)
89.29	90.07	0.9882	0.9991	0.09	0.06

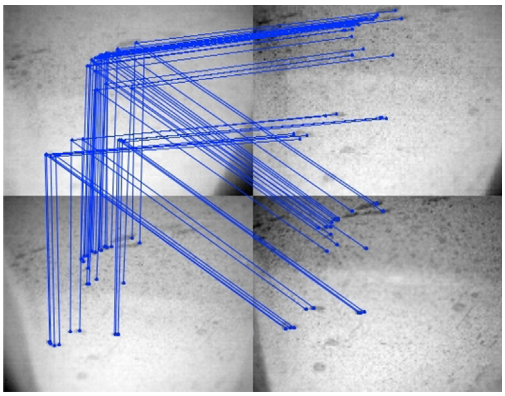


Fig. 10. Example of $n = 57$ matched points on four rectified images.

Table 5
Results on extended hand-eye self-calibration with the reference part.

Σ_d (mm)	$\overline{d}$ (mm)	σ_d (mm)	Σ_e (pix.)	$\overline{e}$ (pix.)	σ_e (pix.)
37.34	1.32	0.86	185.74	3.25	7.96

below. Next, fifty-seven quadruplets of matched points are extracted (i.e. $n = 57$) using phase congruency model [24,25] associated with ZNCC method (see Fig. 10).

The selected criteria to analyze the calibration-reconstruction quality are: $\overline{d}$, the mean value of reconstruction and $\overline{e}$, the mean value of the epipolar distance. They are tabulated in Table 5. The mean value of reconstruction error $\overline{d}$ remains lower than 1.32 mm. The mean value of the epipolar distances $\overline{e}$ is higher than one pixel at around 3.25 pixels. The extended hand-eye self-calibration is then accurate and consistent with the spatial resolution of the infrared cameras.

The mesh performed from the 3*D* point cloud calculated during this step is displayed in Fig. 11 where an image projection of the part is also plotted.

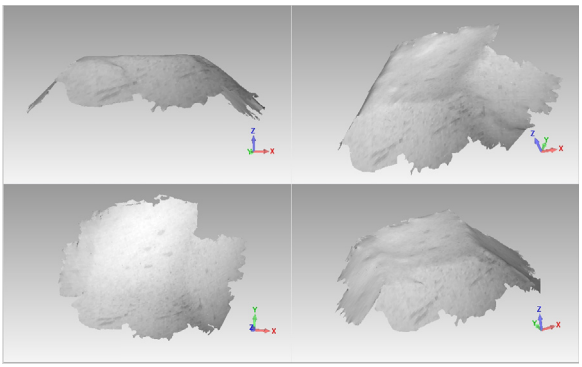


Fig. 11. Mesh from the 3*D* points cloud with a projection of the image part.

4.3.3. Comparison of 3D point cloud calculated by the suggested method and provided with the laser scanner (ground truth)

Fig. 12 provides the 3*D* deviations performed with Geomagic®software between the previous mesh and the reference model (ground truth) built with the laser scanner.

On this figure, the grey color represents the reference model. The over-printed color is the 3*D* deviation values between the 3*D* reconstruction performed by the method and the reference model. This colorbar provides the accuracy of the reconstruction for the different areas of the part. The left box indicates the colorbar associated to each 3*D* deviations and displays the histogram. The two images differs from the range of colorbar. For the first image, the range of colorbar is calculated using the maximal deviation. The image is called full-scale. The second image is displayed with a range containing the most of the deviation values. This image is named restricted-scale. These different scales make full use of the histogram of 3*D* deviations, in *mm*.

The first image with a full-scale highlights the maximum values of the reconstruction deviation. The maximal deviation value reaches 12.8 mm and is located on the left field of the part. Most of 3*D* deviations values are in the interval [−1.5 mm, 1.5 mm]. Thus, the part is mainly green. In the second image, the colorbar is then restricted to the interval [−1.5 mm, 1.5 mm] to analyze accurately deviation locations. The mean error on the top of the part is then equal to 1.7 mm with an standard deviation of 1.8 mm. The maximal deviation is recorded on the edges of the part where the area is not textured enough.

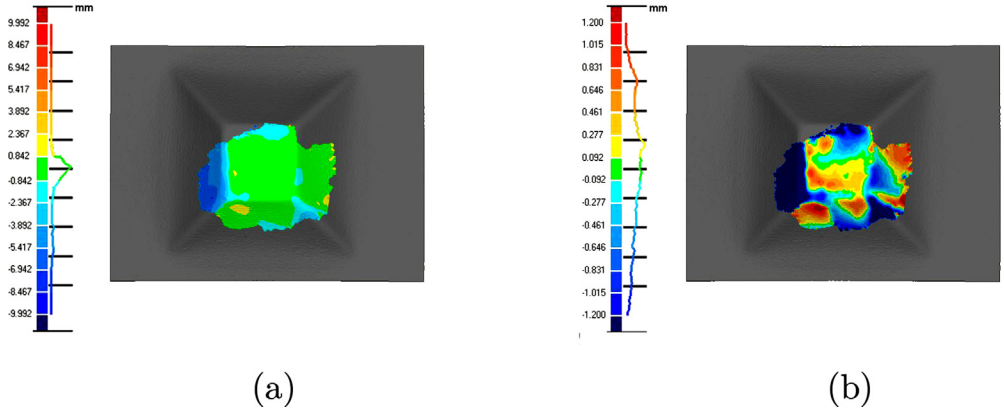
These deviations are comparable to the spatial resolution of the camera which is 1 mm (observation distance 50 mm, focal length 11 mm, pixel size 51 μm and a half correlation window of 4.5 pixel). This reconstruction performance on a plane with very low resolution cameras validates the 3*D* automated thermal reconstruction with a multi-view approach performed with the help of a robot equipped with two uncalibrated infrared cameras.

5. Conclusions and future works

This article presented algorithms required for performing an automated thermal 3*D* reconstruction of a part with a system composed of a stereovision rig of uncalibrated infrared cameras mounted on a six-axis Cartesian robot. This multi-view 3*D* thermal reconstruction relies on a geometric extended self-calibration for estimating both intrinsic parameters for each camera, relative position and orientation between the two cameras and the Euclidean transformation between the robot and the camera reference frames.

A new projective rectification method was first introduced on one hand for improving the stereo matching problem between pixels on the left and right images reducing the search space from two dimensions to a single one, and on the other hand for decreasing the number of parameters of the cost function of the calibration problem. Because the suggested method must cope with noisy and low-textured infrared images, the two homographies needed for the rectification were calculated in a single step, from only a few matched points and without previous information on the Fundamental Matrix. The suggested rectification keeps the aspect ratio and the orthogonality of images, and consequently, limits the creation or loss of pixels. The structure and the skewness of the images are preserved with an error of less than 0.8% on both criteria. Before demonstrating its effectiveness on infrared images in the last section, this method was benchmarked against recent rectification methods on images classically used in the vision community.

Next, the set of matched points from rectified images was used to perform simultaneously, also in a single step, the extended self-calibration and the multi-view infrared 3*D* reconstruction. The minimization problem of calibration was solved through two cost functions selected according to the noise level measured in infrared images. The first objective function aims at minimizing the reprojection errors (in pixels). When the noise level increases, a second objective function is expressed in the projective space, from the epipolar constraints between



(a) (b)

Fig. 12. 3*D* deviations at different scales (full and restricted scales) between the reference model and the mesh emerging from the 3*D* points cloud calculated by the suggested method. The box indicates the scale of each image in *mm* and the values of the displayed histogram.

two images acquired by the same camera from two robot positions. These two functions make explicit the intrinsic camera parameters and the essential matrix between the robot positions, which itself depends on the hand-eye parameters and the robot motion. The tests with simulated images, with noise on the extracted points, confirmed that the method based on the reprojection error is suitable, with a noise less than 1 pixel. Above this value, the second method is more suitable. The reconstruction based on a rectified stereovision rig gives low registration errors with an approach in a single step.

Finally, the automated thermal 3*D* reconstruction was validated by comparing its results with those obtained with a 3*D* scanner with an accuracy to 50 μm. An accurate 3*D* model of a hot object with an accuracy around $\pm$ 1 mm was achieved with several views acquired by uncalibrated infrared cameras.

References

[1] X. Maldague, Theory and Practice of Infrared Technology for Nondestructive Testing, John Wiley Interscience, 2001.

[2] A. Maynadier, M. Poncelet, K. Lavernhe-Taillard, S. Roux, One-shot measurement of thermal and kinematic fields: infrared image correlation (IRIC), Exp. Mech. 52 (2012) 241–255.

[3] S. Vidas, P. Moghadam, Heatwave: a handheld 3D thermography system for energy auditing, Energy Build. 66 (0) (2013) 445–460.

[4] R. Reichle, P. Andrew, G. Counsell, J.-M. Drevon, A. Encheva, G. Janeschitz, D. Johnson, Y. Kusama, B. Levesy, A. Martin, C.S. Pitcher, R. Pitts, D. Thomas, G. Vayakis, M. Walsh, Defining the infrared systems for ITER, Rev. Sci. Instrum. 81 (10) (2010).

[5] Y. Ham, M. Golparvar-Fard, Epar: energy performance augmented reality models for identification of building energy performance deviations between actual measurements and simulation results, Energy Build. 63 (0) (2013) 15–28.

[6] J. Jeswiet, F. Micari, G. Hirt, A. Bramley, J. Duflou, J. Allwood, Asymmetric single point incremental forming of sheet metal, CIRP Ann. Manuf. Technol. 54 (1) (2005) 88–114.

[7] D. Gonzalez-Aguilera, P. Rodriguez-Gonzalvez, J. Armesto, S. LagÃela, Novel approach to 3D thermography and energy efficiency evaluation, Energy Build. 54 (0) (2012) 436–443.

[8] D. Borrmann, A. Nuchter, M. ÃŘakulovic, I. Maurovic, Y. Petrovic, D. Osmankovic, J. Velagic, A mobile robot based system for fully automated thermal 3D mapping, Adv. Eng. Inform. (2014).

[9] K. Nagatani, K. Otake, K. Yoshida, Three-dimensional thermography mapping for mobile rescue robots, in: K. Yoshida, S. Tadokoro (Eds.), Field and Service Robotics, Springer Tracts in Advanced Robotics, vol. 92, Springer, Berlin Heidelberg, 2014, pp. 49–63.

[10] L. Zalud, P. Kocmanova, Fusion of thermal imaging and CCD camera-based data for stereovision visual telepresence, 2013 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), 2013, pp. 1–6.

[11] S. Vidas, P. Moghadam, M. Bosse, 3D thermal mapping of building interiors using an rgb-d and thermal camera, 2013 IEEE International Conference on Robotics and Automation (ICRA), 2013, pp. 2311–2318.

[12] P. Moghadam, S. Vidas, Heatwave: the next generation of thermography devices, Proc. SPIE 9105 (2014), http://dx.doi.org/10.1117/12.2053950 91050F–91050F-8.

[13] S. Prakash, P.Y. Lee, A. Robles-Kelly, Stereo techniques for 3D mapping of object surface temperatures, Quantit. InfraRed Thermogr. J. 4 (1) (2007) 63–84.

[14] T. Sentenac, R. Gilblas, D. Hernandez, Y.L. Maoult, Bi-color near infrared thermoreflectometry: a method for true temperature field measurement, Rev. Sci. Instrum. J. 83 (12) (2012) 124902.

[15] T. Sentenac, R. Gilblas, Noise effect on the interpolation equation for near infrared thermography, Metrologia 50 (3) (2013) 208.

[16] R. Gilblas, T. Sentenac, D. Hernandez, Y.L. Maoult, Quantitative temperature field measurements on a non-gray multi-materials scene by thermoreflectometry, Infrared Phys. Technol. 66 (0) (2014) 70–77.

[17] S. Vidas, R. Lakemond, S. Denman, C. Fookes, S. Sridharan, T. Wark, A mask-based approach for the geometric calibration of thermal-infrared cameras, IEEE Trans. Instrum. Meas. 61 (6) (2012) 1625–1635.

[18] Z. Yu, S. Lincheng, Z. Dianle, Z. Daibing, Y. Chengping, Camera calibration of thermal-infrared stereo vision system, 2013 Fourth International Conference on Intelligent Systems Design and Engineering Applications, 2013, pp. 197–201.

[19] H. Durrant-Whyte, T. Bailey, Simultaneous localization and mapping: part i, IEEE Robot. Autom. Mag. 13 (2) (2006) 99–110.

[20] R. Rusu, 3D robotic mapping: the simultaneous localization and mapping problem with six degrees of freedom, KI - KÃijnstliche Intell. 24 (3) (2010) 267.

[21] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2003.

[22] M.I.A. Lourakis, A.A. Argyros, Sba: a software package for generic sparse bundle adjustment, ACM Trans. Math. Softw. 36 (1) (2009) 2:1–2:30.

[23] Y. Furukawa, J. Ponce, Accurate, dense, and robust multiview stereopsis, IEEE Trans. Pattern Anal. Mach. Intell. 32 (8) (2010) 1362–1376.

[24] K. Hajebi, J. Zelek, Dense surface from infrared stereo, IEEE Workshop on Applications of Computer Vision, 2007. WACV '07, 2007, pp. 21–28.

[25] K. Hajebi, J.S. Zelek, Structure from infrared stereo images, Proceedings of the 2008 Canadian Conference on Computer and Robot Vision, CRV '08, IEEE Computer Society, Washington, DC, USA, 2008, pp. 105–112.

[26] F. Devernay, Vision stéréscopique et propriétés différentielles des surfaces (Ph.D. Thesis), Ecole polytechnique, 1997.

[27] C. Loop, Z. Zhang, Computing Rectifying Homographies for Stereo Vision, Tech. Rep. Microsoft Research, 1999.

[28] H.P. Trivedi, Estimation of stereo and motion parameters using a variational principle, Image Vision Comput. 5 (2) (1987) 181–183.

[29] R. Hartley, Theorie and practice of projective rectification, Int. J. Comput. Vision 35 (1999) 115–127.

[30] J. Gluckman, S.K. Nayar, Rectifying transformations that minimize resampling effects, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, 2001, p. 111.

[31] J. Mallon, P. Whelan, Projective rectification from the fondamental matrix, Image Vision Comput. 23 (2005) 643–650 W.P..

[32] R. Laganière, F. Kangni, Projective rectification of image triplets, Signal, Image Video Process. (2009) 389–397.

[33] P. Monasse, J.-M. Morel, Z. Tang, Three-step image rectification, Proceedings of the British Machine Vision Conference, BMVA Press, 2010, pp. 89.1–89.10.

[34] F. Zilly, M. MÃijller, P. Kauff, P. Eisert, Three-step image rectification, Proceedings of the 3D Data Processing, Visualization and Transmission Conference, 2010.

[35] C. Lawrence, A. Tits, A computationally efficient feasible sequential quadratic programming algorithm, SIAM J. Optim. (2001) 1092–1118.

[36] J.Z.C.T. Lawrence, A. Tits, User's Guide for cfsqp version 2.5, Tech. Rep. Electrical Engineering Department and Institute for Systems Research, University of Maryland, 1997.

[37] E.R. Panier, A.L. Tits, On combining feasibility, descent and superlinear convergence in inequality constrained optimization, Math. Program. (1993) 261–276.

[38] M. Shah, R.D. Eastman, T. Hong, An overview of robot-sensor calibration methods for evaluation of perception systems, Proceedings of the Workshop on Performance Metrics for Intelligent Systems, 2012, pp. 15–20.

[39] R. Tsai, R. Lenz, Real time versatile robotics hand-eye calibration using 3D machine vision, IEEE International Conference on Robotics and Automation, 1988. Proceedings, vol. 1, 1988, pp. 554–561.

[40] Y.C. Shiu, S. Ahmad, Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX = XB, IEEE Trans. Robot. Autom.

5 (1) (1989) 16–29.
[41] J.C.K. Chou, M. Kamel, Finding the position and orientation of a sensor on a robot manipulator using quaternions, Int. J. Rob. Res. 10 (1991) 240–254.
[42] R.P. Horaud, F. Dornaika, Hand-eye calibration, Int. J. Robot. Res. 14 (3) (1995) 195–210.
[43] H. Chen, A screw motion approach to uniqueness analysis of head-eye geometry, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91, 1991, pp. 145–151.
[44] Z. Zhao, Y. Weng, A flexible method combining camera calibration and hand-eye calibration, Robotica 31 (2013) 747–756.
[45] A. Malti, J.P. Barreto, Hand-eye and radial distortion calibration for rigid endoscopes, Int. J. Med. Robot. Comput. Assisted Surg. 9 (4) (2013) 441–454.
[46] A. Malti, Hand-eye calibration with epipolar constraints: application to endoscopy, Robot. Auton. Syst. 61 (2) (2013) 161–169.
[47] J. Heller, M. Havlena, A. Sugimoto, T. Pajdla, Structure-from-motion based hand-eye calibration using l_∞ minimization, 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.
[48] S. Seo, Y.-J. Choi, S.W. Lee, A branch-and-bound algorithm for globally optimal calibration of a camera-and-rotation-sensor system, 2009 IEEE 12th International Conference on Computer Vision, 2009.
[49] J. Heller, M. Havlena, T. Pajdla, Globally optimal hand-eye calibration using branch-and-bound, IEEE Trans. Pattern Anal. Mach. Intell. (99) (2015) 1.
[50] Z. Zhao, Hand-eye calibration using convex optimization, 2011 IEEE International Conference on Robotics and Automation (ICRA), 2011, pp. 2947–2952.
[51] J. Heller, D. Henrion, T. Pajdla, Hand-eye and robot-world calibration by global polynomial optimization, 2014 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2014, pp. 3157–3164.
[52] N. Andreff, R. Horaud, B. Espiau, On-line hand-eye calibration, International Conference on 3D Digital Imaging and Modeling, 1999, p. 0430.
[53] K. Daniilidis, E. Bayro-Corrochano, The dual quaternion approach to hand-eye calibration, Proceedings of the 13th International Conference on Pattern Recognition, vol. 1, 1996.
[54] J. Schmidt, F. Vogt, H. Niemann, Calibration free hand-eye calibration: a structure from motion approach, in: W.G. Kropatsch, R. Sablatnig, A. Hanbury (Eds.), Pattern Recognition, Lecture Notes in Computer Science, vol. 3663, Springer, Berlin/Heidelberg, 2005, pp. 67–74.
[55] G. Qing Wei, K. Arbter, G. Hirzinger, Active self-calibration of robotic eyes and hand-eye relationships with model identification, IEEE Trans. Robot. Autom (1998).
[56] A. Jordt, N. Siebel, G. Sommer, Automatic high-precision self-calibration of camera-robot systems, IEEE International Conference on Robotics and Automation, 2009. ICRA '09, 2009, pp. 1244–1249.
[57] A.S. Lewis, M.L. Overton, Non-smooth optimization via quasi-newton methods, Math. Program. 141 (1) (2013) 135–163.
[58] C. Harris, M. Stephens, A combined corner and edge detector, Proc. of Fourth Alvey Vision Conference, 1988, pp. 147–151.