Wand-based Multiple Camera Studio Calibration

Joel Mitchelson
Charnwood Dynamics Ltd.
Rothley, Leicestershire, UK
joel@charndyn.com

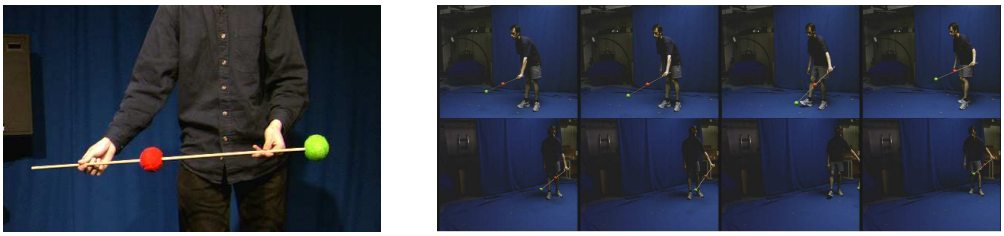
Adrian Hilton
Centre for Vision, Speech, and Signal Processing
University of Surrey, Guildford, UK
a.hilton@surrey.ac.uk

Abstract

To meet the demands of the many emerging multiple camera studio systems in entertainment content production, a novel wand-based system is presented for calibration of both intrinsic (focal length,lens distortion) and extrinsic (position, orientation) parameters of multiple cameras. Full metric calibration is obtained solely from observations of a wand comprising two visible markers at a known fixed distance. It is not necessary for all cameras to see the wand simultaneously, cameras may face each other, and have non-overlapping fields of view. High accuracy is achieved by using iterative bundle adjustment of tracked feature points across multiple views to refine calibration parameters until re-projection errors are minimised over the required measurement volume. The approach involves a new automatic initialisation procedure and novel application of bundle adjustment to refine calibration estimates. Evaluation of wand-calibration is performed using an eight-camera system. Results demonstrate a reprojection error of approximately 0.5 pixels rms and 3D reconstruction error of less than 2mm rms for a capture volume of 2x3x2m. Advantages of wand-based calibration over conventional chart-based calibration include time-efficient calibration of multiple camera systems and calibration of camera configurations without all cameras having to view the same objects or having overlapping fields of view.

I. Introduction

Many emerging applications require calibrated of multiple synchronised video cameras. Multiple camera systems are increasingly used for broadcast and film production in both studios and outdoors for shooting on location and sports events such as football. Simultaneous capture of multiple view video of events allows viewpoint control in post-production for special effects such as slow motion camera moves and 3D content production. Typical applications involve



(a) Close-up of calibration wand (b) Calibration images of wand

Fig. 1. Example images from wand input for two cameras

reconstructing or tracking complex structures, which suffer from problems of self-occlusion if only seen from one point of view, such as a moving person. The accuracy of reconstruction is heavily dependent on the accuracy to which internal camera parameters and relative camera pose are known. For studio production applications cameras may be re-configured many times in one day with changes in location and zoom, so ease of use and speed of computation are of key practical importance.

A typical paradigm for visual calibration is extraction of corresponding point features from images, and use of these to compute the camera parameters. For a limited number of camera views (often a stereo rig) it has become common to use a calibration chart of known structure placed in the field of view to establish point correspondences automatically [1], [2], [3]. Unfortunately the use of a chart is impractical for many multi-view scenarios. This is because the purpose of multiple views is to image the scene from many different angles - some camera views may not overlap and cameras may face each other. So a chart would have to be multi-faceted to be imaged by several cameras simultaneously. Even a planar chart is an unwieldy object to move around a studio.

In this work, we replace the chart with a **wand** - a short rod with two coloured markers on it, shown in figure 1(a). The wand based system introduced in this paper allows rapid, flexible and accurate calibration of multiple camera systems and does not require the cameras to all face in the same direction or have overlapping fields of view. Typical calibration times for a studio system are 5 minutes compared to several hours for chart based calibration. By acquiring video *sequences* of the moving wand, it is possible to build up large sets of point correspondences between views in a short time. These replace the role of multiple feature points on a chart.

In this paper we introduce a wand-based calibration of both intrinsic (focal-length, lens distortion) and extrinsic (position, orientation) parameters of multiple cameras. The calibration algorithm uses iterative bundle adjustment of tracked feature points across multiple views to refine calibration parameters until re-projection errors are minimised over the required measurement volume. Our method is constructed so that it is not necessary for all cameras to see both wand markers in every video image, and cameras may face each other and have non-overlapping fields of view. An evaluation of the calibration accuracy and comparison with conventional chart-based calibration is presented for two and eight camera studio systems.

Contributions of this work include a novel application of bundle adjustment techniques [4], and a new but straightforward way to initialise the system based on self-calibration methods [5], [6], [7]. The wand-based calibration presented allows simultaneous estimation of extrinsic parameters (location, orientation) together with focal length, centre-of-projection and lens distortion for camera configurations with non-overlapping fields of view. Wand-based calibration with observations of two markers at a known distance apart allows metric calibration of multiple cameras without prior estimation of intrinsic parameters. This approach allows rapid and flexible calibration of multiple camera systems with no requirement for the field of view of all cameras to overlap. An experimental evaluation of wand and chart-based calibration is presented for 2-camera and 8-camera systems. Wand-based calibration achieves a reconstruction accuracy comparable to chart calibration and requires an order of magnitude less-time to perform. A public implementation of the method can be downloaded from *http://www.ee.surrey.ac.uk/CVSSP/VMRG/WandCalibration*.

The wand calibration method comprises the following stages:

Multiple View Video Capture: Video of two markers separated by a known fixed distance is acquired as they move through the capture volume. Capture is performed simultaneously with synchronised cameras. Note that it is *not* necessary for all cameras to see markers simultaneously in all frames.

Feature Extraction: Automatically estimate the location of the centre of each visible marker in each image. Accurate estimation of marker location is required for accurate calibration.

Initial Estimation of Camera Configuration: Two-view geometry of marker image locations and knowledge of the distance between the markers is used to estimate an initial

approximation of camera parameters.

Refinement of Camera Calibration: Bundle adjustment is used to refine the initial estimate of camera calibration.

This paper is organised as follows: Section II reviews previous research on multiple camera calibration. Background theory and nomenclature are presented in section III. Robust feature point extraction for a wand with two colour markers is presented in section IV. In section V we develop an initialisation procedure using pair-wise correspondence. The refinement algorithm using a global bundle-adjustment is presented in section VI. Finally a comparative performance evaluation is presented in section VII.

II. Previous Work

In 1980, Tsai developed a simple automated visual method for calibration, using images of a planar chart [1]. Subsequent work has developed this concept to allow more accurate initial estimates to be made [2], [3]. This has coincided with a good understanding of how descent based optimisation methods may be used to refine a calibration [4]. Accurate chart-based systems are now freely available on the internet.[1]

Also of note is the development of 'self-calibration' methods. These were introduced by Hartley [5], [6] and Faugeras [7]. They allow camera parameters to be determined from point correspondences alone, and can be used to solve the 'structure-from-motion' problem, whereby the 3D geometry of a natural scene is recovered using a 2D image sequence from a single moving camera [8], [9], [10]. Factorisation methods [11], [12], [13] use a different mathematical formulation to provide iterative solutions to the same problem. We will see that self-calibration algorithms can be adapted to give an initial solution to the problem of calibration using point correspondences from a moving wand.

The idea of wand calibration has been around for some years, as seen in work by Maas [14], but it was not fully developed or evaluated. Similar techniques also appear to have been used by some commercial motion capture companies. Work developed independently by Baker and Aloimonos

[1] `http://www.vision.caltech.edu/bouguetj/calib_doc`
`http://sourceforge.net/projects/opencvlibrary/`

[15] presents a related technique using an active marker. They use an iterative factorisation method which effectively minimises an algebraic error over multiple views. This approach has recently been extended by Svoboda et al. [16] to Euclidean reconstruction using a single point marker. Their current implementation uses a visible-light LED active marker and so the method is best performed in the dark. Since a single point is used, an additional application-specific distance measurement step is required before full metric calibration can be achieved. A different application of the wand-based approach was recently demonstrated by Zhang [17]. A wand with 3 points, fixed at one end, was used to solve the problem of internal parameter calibration when only one camera is present.

III. Background Theory and Notation

A. Camera Model

We use the standard **pinhole model** for cameras. Points in the world are specified by 3D co-ordinates $\begin{pmatrix} x & y & z \end{pmatrix} \in \mathbb{R}^3$. The corresponding point in a camera image is described by 2D co-ordinates $\begin{pmatrix} u & v \end{pmatrix}$. It is convenient to specify such points in homogeneous co-ordinates, $\mathbf{x} \in \mathbb{R}^4$ for world points and $\mathbf{u} \in \mathbb{R}^3$ for image points, where $\mathbf{x} = \begin{bmatrix} x & y & z & 1 \end{bmatrix}^\mathsf{T}$ and $\mathbf{u} = \begin{bmatrix} u & v & 1 \end{bmatrix}^\mathsf{T}$. The pinhole model states that, in the absence of noise, a camera maps world points to image points as follows:

$$\lambda \mathbf{u} = \mathbf{P}\mathbf{x} \tag{1}$$

$\mathbf{P}$ is a 3x4 matrix of rank 3, known as the camera **projection matrix**. Saying that a camera is **calibrated** is equivalent to saying that we have an estimate of $\mathbf{P}$. Given a world point and a camera projection matrix, this relation allows us to calculate the corresponding image point in homogeneous co-ordinates up to a scalar factor, λ. Since $\lambda \mathbf{u} = \lambda \begin{bmatrix} u & v & 1 \end{bmatrix}^\mathsf{T}$, it follows that we can extract the non-homogeneous co-ordinates as follows:

$$u = \frac{\mathbf{p}_1 \mathbf{x}}{\mathbf{p}_3 \mathbf{x}} \quad v = \frac{\mathbf{p}_2 \mathbf{x}}{\mathbf{p}_3 \mathbf{x}} \tag{2}$$

Where $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$ are the rows of $\mathbf{P}$. The need for this division shows that the pinhole model is non-linear, despite the apparent linearity of equation 1.

A 3x4 matrix $\mathbf{P}$ of rank 3 may be decomposed into the form $\mathbf{P} = \mathbf{K}\left[\begin{array}{c|c} \mathbf{R} & \mathbf{t} \end{array}\right]$, where $\mathbf{t} \in \mathbb{R}^3$, $\mathbf{K}^{3\times3}$ is upper triangular, and $\mathbf{R} \in SO(3)$ is orthonormal. These components have physical interpretation. Rotation matrix $\mathbf{R}$, and translation vector $\mathbf{t}$ represent a 3D rigid transform from world co-ordinates to a co-ordinate frame centred on the camera. These are the **external** camera parameters. $\mathbf{K}$ represents the camera's **internal** parameters. We are concerned with CCD cameras which have rectilinear arrays of pixels. Such internal parameters have the form:

$$\mathbf{K} = \begin{bmatrix} f & 0 & u_0 \\ 0 & af & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

f is the effective focal length of the camera (in units of pixels), a is the aspect ratio, and $\begin{pmatrix} u_0 & v_0 \end{pmatrix}$ is the centre of projection in image co-ordinates.

In this work the centre of projection is not considered as a variable, and fixed default values are used throughout. Justification for this simplification is given in section VII-G.

The pinhole model (equation 1) is an elegant formulation which is used throughout this paper. It does not, however, model any distortion artefacts due to camera lenses. This is considered as a special case in the evaluation (section VII-H).

B. Calibration from Point Correspondences

Suppose we have N cameras, so there are N projection matrices $\mathbf{P}_1 \ldots \mathbf{P}_N$ to be found. Now suppose that by some means we have identified a set of points whose observed positions in each camera view are known. Let the i^{th} observation of point j be $\mathbf{u}_{ij}$. At this stage we know neither $\mathbf{P}_i$ nor $\mathbf{x}_j$. Since some noise will always be present, we formulate calibration as a minimisation problem. We wish to find the $\mathbf{P}_i$ and $\mathbf{x}_j$ which minimise the geometric error function:

$$E_{geom} = \sum_{i=1}^{N}\sum_{j=1}^{M} \left| u_{ij} - \frac{\mathbf{p_{i}}_1\mathbf{x}_j}{\mathbf{p_{i}}_3\mathbf{x}_j} \right|^2 + \left| v_{ij} - \frac{\mathbf{p_{i}}_1\mathbf{x}_j}{\mathbf{p_{i}}_3\mathbf{x}_j} \right|^2 \tag{4}$$

This expresses the sum of squared errors between reconstructed points and actual points, measured as Euclidean distance in 2D image space. It can be minimised using descent methods

(bundle adjustment), but this requires a good initial estimate for the process to converge. To show that such an initial estimate exists, consider the following approximation to equation 4:

$$E_{alg} = \sum_{ij} |\lambda \mathbf{u}_{ij} - \mathbf{P}_i \mathbf{x}_j|^2 \tag{5}$$

which measures the error as a distance in homogeneous image co-ordinates. Hartley [5] showed that minimising such equations for two or more cameras *is sufficient* to determine projection matrices up to a non-singular transform of homogeneous world co-ordinates, $\mathbf{H} \in \mathrm{GL}(4)$. This is known as a **projective reconstruction** of the world points. Hartley [6] also shows that if the internal parameters are known to have the form of equation 3 with known centre of projection and aspect ratio, then for two cameras the elements of $\mathbf{H}$ will have at most two degrees of freedom. These correspond to scene scale and a one-parameter family of choices for the two focal lengths[2]. Now suppose that there is a known Euclidean distance d between a pair of points $\mathbf{x}$ and $\mathbf{y}$ in the scene. This places the following additional constraint on the elements of $\mathbf{H}$:

$$|(\mathbf{h}_4\mathbf{y})\mathbf{H}\mathbf{x} - (\mathbf{h}_4\mathbf{x})\mathbf{H}\mathbf{y}|^2 = (\mathbf{h}_4\mathbf{x})^2(\mathbf{h}_4\mathbf{y})^2 d^2 \tag{6}$$

where $\mathbf{h}_4$ is the 4^{th} row of $\mathbf{H}$. It can be shown that this is quadratic in the cross-terms of elements of $\mathbf{H}$. Using a wand with two points separated by known distance, each image will in general contribute an independent constraint on the elements of $\mathbf{H}$, which will constrain the two unknown degrees of freedom. This suggests that calibration is possible using data from a wand of known length.

C. Two-View Geometry

Here we summarise without proof some important results in two-view geometry which are useful for finding approximate estimates of camera parameters, [18], [6]. In the case of two cameras a pairwise estimate of calibration can be found from a **fundamental matrix** computed using the **8-point algorithm**. This concept was defined by Longuet-Higgins [19] and extended to the uncalibrated case by Faugeras *et al.* [7] and Hartley [5].

[2]Constraints on the degrees of freedom for $\mathbf{H}$ in a given projective reconstruction may be established using the principal of the Image of the Absolute Conic (IAC). See [18] for details.

Theorem 3.1 (Fundamental Matrix): Suppose we have two cameras and have identified corresponding sets of image points seen from each camera $\{\mathbf{u}_{1j}\} \leftrightarrow \{\mathbf{u}_{2j}\}$. Then in the absence of noise, there exists a matrix $\mathbf{F} \in \mathbb{R}^{3\times3}$ with rank$(\mathbb{F}) \leq 2$ such that: $\mathbf{u}_{1j}^{\mathsf{T}}\mathbf{F}\mathbf{u}_{2j} = 0 \quad \forall\, j$ and this matrix is determined completely by 8 or more correspondence points (except for a small subspace of degenerate cases).

Theorem 3.2 (Decomposition of Fundamental Matrix): A fundamental matrix $\mathbf{F}$ is composed of the following matrices:

$$\mathbf{F} = \mathbf{K}_1^{\mathsf{T}}\mathbf{R}[\mu\mathbf{t}]_\times\mathbf{K}_2 \tag{7}$$

Where $\mathbf{K}_1$, $\mathbf{K}_2$ are the internal parameters of each camera, $(\mathbf{R}\ \mathbf{t})$ is relative rigid transform from the co-ordinate system of one camera to the other, μ is a non-zero scalar and $[\mathbf{t}]_\times$ denotes the skew-symmetric matrix:

$$[\mathbf{t}]_\times = \begin{bmatrix} 0 & -t_3 & t_2 \\ -t_1 & 0 & t_3 \\ -t_2 & t_1 & 0 \end{bmatrix} \tag{8}$$

It follows that $\mathbf{R}$ and $\mu\mathbf{t}$ may be extracted from $\mathbf{F}$, given $\mathbf{K}_1$ and $\mathbf{K}_2$ [18].

Theorem 3.3 (Self-Calibration): The focal lengths of two cameras may be estimated from their fundamental matrix if aspect ratios and centre of projections are known, except if the principal rays of the cameras meet. In this case a mean focal length may be found.

Theorem 3.4 (Reconstruction): Given 2D locations of a point in two or more images from cameras whose principal rays are not co-linear and whose projection matrices are known, it is possible to estimate the 3D location of the point.

D. Bundle Adjustment

Let us assume for now we have a way to get an approximate estimate of camera parameters. We require a method to update those parameters to find the minimum of the function in equation 4. This may be expressed as a non-linear optimisation problem, with several constraints. It may be solved iteratively using **descent methods** of optimisation. When a descent method is tuned to exploit the specific structure of a calibration problem, it is known as a **bundle adjustment** approach [4].

In general, a descent method assumes there is a known function $\mathbf{f} : \mathbb{R}^A \mapsto \mathbb{R}^B$, and a known value $\mathbf{Z} \in \mathbb{R}^B$. The problem is to find $\mathbf{X} \in \mathbb{R}^A$ such that $|\mathbf{Z} - \mathbf{f}(\mathbf{X})|$ is as small as possible. Descent methods iteratively refine the estimate of parameters X by evaluating an increment $\Delta\mathbf{X}$ such that $|\mathbf{f}(\mathbf{X}+\Delta\mathbf{X}) - \mathbf{Z}| < |\mathbf{f}(\mathbf{X}) - \mathbf{Z}|$. In this work we use a weighted Levenberg-Marquardt method [20] to evaluate the increment $\Delta\mathbf{X}$ as the solution to the linear equations:

$$(\mathbf{J}^\mathsf{T} W \mathbf{J} + \lambda \mathbf{I})\Delta\mathbf{X} = \mathbf{J}^\mathsf{T} W(\mathbf{f}(\mathbf{X}) - \mathbf{Z}) \tag{9}$$

where $\mathbf{J}$ is the Jacobian matrix of partial derivatives:

$$\mathbf{J} = \left[\frac{\partial \mathbf{f}}{\partial X_1} \; \frac{\partial \mathbf{f}}{\partial X_2} \; \dots \right] \tag{10}$$

The parameter λ is set as part of the algorithm [20] - if small it results in large Gauss-Newton (quadratic) updates and if large it results in smaller steepest-descent (linear) updates. In this work λ is initialised to 0.001, increased by a factor 10 if $\Delta\mathbf{X}$ reduces the cost and decreased by a factor 10 if $\Delta\mathbf{X}$ increases the cost. The weight matrix W determines the influence of each observation on the solution and is used to turn off the effect of any point which is not visible in a given view.

The parameters may be required to satisfy some constraints, $\mathbf{c}(\mathbf{X}) = \mathbf{0}$. At each iteration step, we consider only a linearised model of the Constraints $\mathbf{c}(\mathbf{X} + \Delta\mathbf{X}) \simeq \mathbf{c}(\mathbf{X}) + C\Delta\mathbf{X}$, where C is the Jacobian matrix of $\mathbf{c}$ at $\mathbf{X}$. The linearised constraints are introduced into the optimisation using the method of **Lagrange Multipliers** [21]. The Levenberg-Marquardt update is now found by solving the equations:

$$\begin{bmatrix} \mathbf{J}^\mathsf{T} W \mathbf{J} + \lambda \mathbf{I} & C^\mathsf{T} \\ C & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{X} \\ \alpha \end{bmatrix} = \begin{bmatrix} \mathbf{J}^\mathsf{T} W(\mathbf{f}(\mathbf{X}) - \mathbf{Z}) \\ \mathbf{c}(\mathbf{X}) \end{bmatrix} \tag{11}$$

We will return to this formula in section VI to design a specific bundle adjustment method.

IV. Wand Feature Extraction

In this work a wand with two coloured spherical markers at a known distance is used to allow accurate extraction of corresponding feature points across multiple views. The use of spherical

markers allows simultaneous visibility and accurate localisation from cameras with opposing viewing directions. Observation of markers from widely spaced views is important to ensure accurate estimation of their relative extrinsic parameters. Colour facilitates the segmentation of markers in the captured video. Alternatives such as colour LED's provide small distinct markers but do not allow simultaneous visibility from all directions. It should be noted that the wand-calibration algorithm does not require markers to be simultaneously visible in all views, allowing calibration of non-overlapping camera configurations and marker occlusion.

Accurate localisation of the spherical marker centres is required for accurate calibration. An outline of the approach developed for accurate marker localisation is presented in this section. Full details of the implementation of marker extraction and localisation are presented in [22]. Marker localisation is performed in the following steps:

1) **Colour Segmentation:** Pixels corresponding to each marker colour (red and green) are extracted using an adaptive thresholding technique [23].
2) **Centroid Estimation:** The centroid of pixels corresponding to each marker is calculated.
3) **Refinement:** Robust circle fitting is used to refine the centroid estimates in the following steps:
    a) **Marker Edge Detection:** Colour edge detection is performed in a 64×64 region about the initial centroid estimate using Sobel edge detection on the red and blue chrominance.
    b) **Circle Fitting:** RANSAC [24] is used to robustly estimate the marker centroid from the edge images. Triples of edge points are used to estimate the marker centre and radius. The edge strength around the estimated circle is then summed to evaluate the observation support. K edge pixel triples are sampled and the circle with the highest support is taken as the initial estimate.
    c) **Circle Optimisation** The circle parameters are refined to optimise the sum of edge strength around the circumference. To avoid local minima of the derivative in the objective function importance sampling in a 1.5 pixel neighbourhood is used to search for the best solution. In this work 400 samples for both triplet selection and sampling have been found to give reliable results.
4) **Marker Verification:** To eliminate unreliable marker locations tests are performed to iden-

(a) (b) (c) (d) (e) (f)

Fig. 2. Wand marker colour segmentation (green marker) and location refinement using edge information (a) Input image (b) Colour correspondence(c) Marker pixels (d) Weighted edge image (e) Extracted edges (f) Estimated marker location.

tify markers going out of view, partial occlusion and colour distortion due to shadowing. Markers within 64 pixels of the image boundary are eliminated to avoid partial visibility within the field of view. The number of pixels identified for a marker, edge strength and colour are used to eliminate marker that are partially occluded or in shadow.

Wand marker feature extraction provides estimates of the centroid for all markers visible at each video frame from each camera view. The set of estimated marker centroids provide the input for camera calibration. Throughout this work a wand with two spherical markers of approximately 7cm diameter with a spacing of 40cm has been used. The wand used for our experiments is shown in figure 1. The marker localisation method was found to perform with sub-pixel resolution of approximately 0.2 pixel rms noise at a distance of $3-5$m. Marker colour segmentation and location refinement for a typical image frame are shown in figure 2. Further details of implementation and evaluation can be found in [22].

V. Calibration Initialisation

Having automatically identified corresponding marker points over multiple frames from all cameras, an initial estimate of camera calibration can be made. Some methods exist for simultaneous solution of equation 5 over an arbitrary number of cameras. These are known as **factorisation methods** [11], [12], [13] and require iterative evaluation of eigenvector problems together with simultaneous feature visibility in all frames. For efficiency and ease of implementation, we use closed-form *pairwise* estimates of geometry as described in section III-C. These are combined to form a globally consistent calibration initialisation. A pairwise approach was also used by Sturm and Triggs [12].

We begin by finding suitable parameters for successive pairs of cameras, (1 and 2, 2 and 3 ... $(N-1)$ and N). Theorem 3.1 shows that we can extract the fundamental matrix $\mathbf{F}$ for a pair of cameras from point correspondences. The standard algorithm for doing this is given in [5], [18]. In this work we use a straightforward application of the RANSAC algorithm [24] to obtain a robust initial estimate of the fundamental matrix, where the error metric is the reprojection error for reconstructed points. Zhang [25] reviews robust methods for fundamental matrix estimation.

The next step is to establish initial estimates for focal length. In practice approximate focal lengths are often known in which case no computation is necessary. If focal lengths are unknown, they can be computed efficiently using theorem 3.3, which is effective provided camera principal rays do not come close to intersection. For this degenerate case an alternative is to make use of equation 6 to perform an exhaustive search to minimise variation in measured wand-length from the two cameras. In experiments a coarse search using increments of 50 pixels was found to be sufficient to ensure successful bundle adjustment (see section VII-F). This search takes approximately two seconds per camera-pair on a 1GHz machine.

Given initial estimates of fundamental matrix and focal length the next step is to extract estimates for relative rotation $\mathbf{R}$ and translation direction $\mu\mathbf{t}$. Once focal lengths are known, the non-singular matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ are determined, so theorem 3.2 implies that $\mathbf{R}$ and $\mu\mathbf{t}$ can be calculated. At this point the parameter μ is unknown, since we have not specified the scale of the scene.

Remark 5.1: If $\mathbf{x}_\mu = \begin{bmatrix} x_\mu & y_\mu & z_\mu & 1 \end{bmatrix}$ is the 3D reconstruction of image point $\mathbf{u} = \begin{bmatrix} u & v & 1 \end{bmatrix}$, according to camera model $P_\mu = K[\mathbf{R} \mid \mu\mathbf{t}]$, then $\mathbf{x} = \begin{bmatrix} \frac{x_\mu}{\mu} & \frac{y_\mu}{\mu} & \frac{z_\mu}{\mu} & 1 \end{bmatrix}$ is a reconstruction of the same point according to camera model $P = K[\mathbf{R} \mid \mathbf{t}]$.

Proof: $\mathbf{x}_\mu$ satisfies $\lambda\mathbf{u} = P_\mu\mathbf{x}_\mu$ for some $\lambda \in \mathbb{R}$.

So $\lambda\mathbf{u} = K[\mathbf{R} \mid \mu\mathbf{t}]\,\mathbf{x}_\mu = K \begin{bmatrix} \mathbf{r}_1 & \mu t_1 \\ \mathbf{r}_2 & \mu t_2 \\ \mathbf{r}_3 & \mu t_3 \end{bmatrix} \begin{bmatrix} x_\mu \\ y_\mu \\ z_\mu \\ 1 \end{bmatrix} = K \begin{bmatrix} \mu\mathbf{r}_1 & \mu t_1 \\ \mu\mathbf{r}_2 & \mu t_2 \\ \mu\mathbf{r}_3 & \mu t_3 \end{bmatrix} \begin{bmatrix} \frac{x_\mu}{\mu} \\ \frac{y_\mu}{\mu} \\ \frac{z_\mu}{\mu} \\ 1 \end{bmatrix}$.

Letting $\lambda' = \frac{\lambda}{\mu}$ gives $\lambda'\mathbf{u} = K[\mathbf{R} \mid \mathbf{t}]\,\mathbf{x}$ as required $\square$

Lemma 5.2 (Scene Scale): We can set the scale of the scene using measurements of two points

separated by known distance.

Proof: Let us call the two points A and B, and distance d. Let $\mathbf{u}^{\mathrm{A}}_{1k}$, $\mathbf{u}^{\mathrm{B}}_{1k}$ be the location of points A and B in the k^{th} image in the video sequence from camera 1. Let $\mathbf{u}^{\mathrm{A}}_{2k}$, $\mathbf{u}^{\mathrm{B}}_{2k}$ be the corresponding point locations from camera 2. Let $\mu\mathbf{t}$ be a known initial estimate of translation with arbitary μ, and $\mathbf{t}$ be the unknown appropriately scaled value. We can define the following projection matrices:

$$P_{\text{pair1}} = \mathbf{K}_1 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{12}$$

$$P_{\text{pair2}} = \mathbf{K}_2 \left[\mathbf{R} \mid \mu\mathbf{t}\right] \tag{13}$$

According to theorem 3.4 P_{pair1}, P_{pair2}, $\mathbf{u}^{\mathrm{A}}_{1k}$, $\mathbf{u}^{\mathrm{A}}_{2k}$, are sufficient to determine the 3D location $\mathbf{x}^{\mathrm{A}}_k$ of marker A (for computations see appendix A). $\mathbf{x}^{\mathrm{B}}_k$ is computed in the same way. It follows from remark 5.1 that $|\mathbf{x}^{\mathrm{A}}_k - \mathbf{x}^{\mathrm{B}}_k| = \mu d \quad \square$

To reduce the effects of noise, we use all such pairs of points in the video sequence and take the mean. Hence:

$$\mathbf{t} = \left(\frac{1}{Nd} \sum_k \left| \mathbf{x}^{\mathrm{A}}_k - \mathbf{x}^{\mathrm{B}}_k \right| \right) (\mu\mathbf{t}) \tag{14}$$

Now suppose we have N cameras, and have computed relative transform $(\mathbf{R}_{1,2}\ \mathbf{t}_{1,2})$ for the first pair, $(\mathbf{R}_{2,3}\ \mathbf{t}_{2,3})$ for the next pair, and so on. We may define a global co-ordinate system, which has the reference frame of the first camera, as follows:

$$\begin{array}{rclrcl} \mathbf{R}_1 & = & I & \mathbf{t}_1 & = & \mathbf{0} \\ \mathbf{R}_{n+1} & = & \mathbf{R}_n\mathbf{R}_{n,n+1} & \mathbf{t}_{n+1} & = & \mathbf{R}_n\mathbf{t}_{n,n+1} + \mathbf{t}_n \end{array} \tag{15}$$

Initialisation by pair-wise camera calibration has been found to provide initial estimates of sufficient accuracy for global refinement using bundle adjustment with all camera configurations tested. If errors accumulate in the initial pair-wise estimation then refinement could be applied

to subsets of cameras prior to integration in a global system. This has not been found to be necessary in this work.

VI. Application of Bundle Adjustment

We present a novel application of bundle adjustment which optimises over focal lengths whilst using wand length as a constraint. We have made a careful choice of parameterisation to suit our problem.

A. Parameterisation

Here we formulate the minimisation of 4 as an optimisation problem of the form used in section III-D. The unknowns in equation 4 are the rotation, translation and focal length of each camera, and the real-world location of each observed point. All of these quantities must be parameters in our optimisation. A rotation matrix $\mathbf{R} \in \mathrm{SO}(3)$ comprises 9 numbers, and must satisfy the polynomial constraints $\mathbf{R}^{\mathsf{T}}\mathbf{R} = \mathbf{I}$, $\det(\mathbf{R}) = 1$. We can choose parameterisation to eliminate these constraints. A popular choice is the **quarternion**, which has 4 parameters and 1 constraint[4]. We use the **exponential map**, which has 3 parameters and no constraints [26]. It is defined:

$$\exp : \mathbb{R}^3 \mapsto \mathrm{SO}(3) \qquad \exp(\omega) = \cos(|\omega|)\mathbf{I} + \frac{\sin(\omega)}{|\omega|}\left[\omega\right]_\times + \frac{1-\cos(|\omega|)}{|\omega|^2}\omega\omega^{\mathsf{T}} \tag{16}$$

Some authors say that 3-parameter representations of rotation impose singularities which cause bundle adjustment to fail [4]. Although this is true of the **Euler angles**, it is not true of the exponential map. We represent camera translation simply using the 3D vector $\mathbf{t} \in \mathbb{R}^3$. Focal length is scaled by constant f_0 so as to have an order of magnitude of 1. The projection matrix $\mathbf{P}_i$ for the i$^{\text{th}}$ camera is thus represented using 7 parameters $q_{i1} \ldots q_{i7}$, as follows:

$$\mathbf{P}\begin{pmatrix} q_{i1} \\ q_{i2} \\ \vdots \\ q_{i7} \end{pmatrix} = \begin{bmatrix} q_{i7}f_0 & 0 & u_{i0} \\ 0 & q_{i7}\alpha_i f_0 & v_{i1} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \exp\begin{pmatrix} q_{i1} \\ q_{i2} \\ q_{i3} \end{pmatrix} & \begin{matrix} q_{i4} \\ q_{i5} \\ q_{i6} \end{matrix} \end{bmatrix} \tag{17}$$

Point locations are also 3D vectors $\mathbf{x}_j \in \mathbb{R}^3$. It has been suggested that points should be represented in homogeneous co-ordinates using 4D vectors, since this allows stable estimation for points a long distance from the camera (said to be at infinity). Since all of our points lie within

a fixed capture volume, we do not require such a representation. To initialise the estimates of 3D point location, we can use our initial estimate of camera projection matrices with the method given in Appendix A.

We can now re-write equation 4 to give an optimisation problem of the form shown in section III-D. Let $\mathbf{p}_{i1}$, $\mathbf{p}_{i2}$, $\mathbf{p}_{i3}$ be the rows of the projection matrix of the i$^{\text{th}}$ camera. Define a function $\mathbf{f}_{ij} : \mathbb{R}^{10} \mapsto \mathbb{R}^2$ as:

$$\mathbf{f}_{ij}\begin{pmatrix} q_{i1} \\ \vdots \\ q_{i7} \\ x_j \\ y_j \\ z_j \end{pmatrix} = \begin{bmatrix} \dfrac{\mathbf{p}_{i1}\begin{bmatrix} x_j & y_j & z_j & 1 \end{bmatrix}^{\mathsf{T}}}{\mathbf{p}_{i3}\begin{bmatrix} x_j & y_j & z_j & 1 \end{bmatrix}^{\mathsf{T}}} \\ \\ \dfrac{\mathbf{p}_{i2}\begin{bmatrix} x_j & y_j & z_j & 1 \end{bmatrix}^{\mathsf{T}}}{\mathbf{p}_{i3}\begin{bmatrix} x_j & y_j & z_j & 1 \end{bmatrix}^{\mathsf{T}}} \end{bmatrix} \tag{18}$$

This is the projection of the current estimate of point location j into image co-ordinates according to the current estimate of camera parameters i. We can define a function $\mathbf{f} : \mathbb{R}^{7N+3M} \mapsto \mathbb{R}^{2MN}$ by writing:

$$\mathbf{f}\begin{pmatrix} \mathbf{q}_1 \\ \vdots \\ \mathbf{q}_N \\ \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_M \end{pmatrix} = \begin{bmatrix} \mathbf{f}_{11}\begin{pmatrix} \mathbf{q}_1 \\ \mathbf{x}_1 \end{pmatrix}^{\mathsf{T}} & \dots & \mathbf{f}_{N1}\begin{pmatrix} \mathbf{q}_N \\ \mathbf{x}_1 \end{pmatrix}^{\mathsf{T}} & \mathbf{f}_{12}\begin{pmatrix} \mathbf{q}_1 \\ \mathbf{x}_2 \end{pmatrix}^{\mathsf{T}} & \dots\dots & \mathbf{f}_{NM}\begin{pmatrix} \mathbf{q}_N \\ \mathbf{x}_M \end{pmatrix}^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}} \tag{19}$$

Now let $\mathbf{Z} = [u_{11}\, v_{11}\, \dots\, u_{N1}\, v_{N1}\, v_{12}\, \dots\, u_{N2}\, v_{N2}\, \dots\dots\, u_{NM}\, v_{NM}]^{\mathsf{T}}$ and $\mathbf{X} = [\mathbf{q}_1\, \dots\, \mathbf{q}_7\, \mathbf{x}_1\, \dots\, \mathbf{x}_N]$. This enables us to express equation 4 using the required form:

$$E_{\text{geom}} = |\mathbf{f}(\mathbf{X}) - \mathbf{Z}|^2 \tag{20}$$

Introducing a suitable weight matrix W gives us the modified error function:

$$E_{\text{weighted}} \;=\; (\mathbf{f}(\mathbf{X}) - \mathbf{Z})^{\mathsf{T}} W (\mathbf{f}(\mathbf{X}) - \mathbf{Z}) \tag{21}$$

We can also impose the constraints that wand markers are separated by unit length. That means that for each consecutive pair of points, $\mathbf{x}_{2k-1}$, $\mathbf{x}_{2k}$ we have a constraint function:

$$c_k \left(\begin{array}{c} \mathbf{x}_{2k-1} \\ \mathbf{x}_{2k} \end{array} \right) = \left(\mathbf{x}_{2k} - \mathbf{x}_{2k-1}\right)^{\mathsf{T}} \left(\mathbf{x}_{2k} - \mathbf{x}_{2k-1}\right) - 1 \tag{22}$$

The linearised constraint is:

$$c_k \left(\begin{array}{c} \mathbf{x}_{2k-1} + \Delta\mathbf{x}_{2k-1} \\ \mathbf{x}_{2k} + \Delta\mathbf{x}_{2k} \end{array} \right) \simeq c_{k0} + C_k \left[\begin{array}{c} \Delta\mathbf{x}_{2k-1} \\ \Delta\mathbf{x}_{2k} \end{array} \right] \tag{23}$$

where $c_{k0} = \left(\mathbf{x}_{2k} - \mathbf{x}_{2k-1}\right)^{\mathsf{T}} \left(\mathbf{x}_{2k} - \mathbf{x}_{2k-1}\right) - 1$ and $C_k = 2\left[\left(\mathbf{x}_{2k-1} - \mathbf{x}_{2k}\right)^{\mathsf{T}} \ \left(\mathbf{x}_{2k} - \mathbf{x}_{2k-1}\right)^{\mathsf{T}}\right]$

B. Implementation

The formula for the model function $\mathbf{f}$ from equation 19 can be differentiated (by hand) to get a formula for associated Jacobian matrix $\mathbf{J}$, in terms of the camera parameters and estimated point locations. To perform a Levenberg-Marquardt update, the calculated values of $\mathbf{f}$ and $\mathbf{J}$ are substituted into equation 18, together with the linearised constraint values C_k calculated using equation 23. The matrix equation must then be solved to obtain the update, $\Delta\mathbf{X}$.

There are well-known standard methods for solving a general matrix equation [27], so the obvious solution would be to use one of these. Typical dimensions of the matrix involved may be $\sim 1000 \times 1000$. There are two problems with this approach: there may be numerical round-off errors; and the computation time may be several minutes per iteration on current computing hardware. A better implementation respects the sparsity of the matrices which naturally arises in this problem.

Entry J_{mn} in the Jacobian matrix $\mathbf{J}$ expresses the dependency of observation m on parameter n. If parameter n is unrelated to observation m, then $J_{mn} = 0$. For our problem, we can see from equation 19 that observations of a point j are dependent on the parameters of the cameras $\mathbf{q}_1 \ldots \mathbf{q}_N$, and the parameters of the corresponding 3D point location $\mathbf{x}_j$, but are *independent*

of all other parameters. So for our problem, $\mathbf{J}$ has the following structure:

$$\mathbf{J} = \left[\begin{array}{c|ccccc} \mathbf{J}_{\text{cam } 1} & \mathbf{J}_{\text{pt } 1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{J}_{\text{cam } 2} & \mathbf{0} & \mathbf{J}_{\text{pt } 2} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{J}_{\text{cam } 3} & \mathbf{0} & \mathbf{0} & \mathbf{J}_{\text{pt } 3} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{J}_{\text{cam } M} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{J}_{\text{pt } M} \end{array}\right] \tag{24}$$

For all image observations of point j, the matrices $\mathbf{J}_{\text{cam } j}$ with dimensions $2N \times 7N$ encode derivatives with respect to all camera parameters $\mathbf{q}_1 \dots \mathbf{q}_N$, and $\mathbf{J}_{\text{pt } j}$ with dimensions $2N \times 3$ encode derivatives with respect to corresponding 3D point parameters $\mathbf{x}_j$. $\mathbf{J}$ is said to be **sparse**, because a large number of elements are always zero [3]. It follows that $\mathbf{J}^\mathsf{T} W \mathbf{J} + \lambda \mathbf{I}$ of equation 9 is also sparse if the weighting matrix W is block diagonal. The lower right hand corner of this matrix is composed of block diagonal elements $\mathbf{J}_{\text{pt i}}^\mathsf{T} W_i \mathbf{J}_{\text{pt i}} + \lambda \mathbf{I}$ with a 3×3 block size. An efficient method for solving the linear system is presented in Appendix B.

Constraints can be included whilst maintaining the block diagonal structure by reordering the rows of equation 11. We ensure that each constraint $\mathbf{c}_k$ is put in a row adjacent to the points to which it refers. This arrangement retains a block structure, although the block size is now increased to 7×7 instead of 3×3.

VII. Evaluation

Experiments were designed to test the convergence, stability, model-fitting accuracy, and positional accuracy of calibrations produced using the wand. Where possible, accuracy was compared with a standard chart-based calibration method. An additional experiment was constructed to illustrate the use of a lens distortion model.

The cameras used were Sony DXC-9100P 3-CCD PAL cameras, running in 25Hz progressive scan mode with image size 720x576 pixels and shutter speed 1/125s. The multiple cameras were electronically synchronised.

[3] In fact $\mathbf{J}_{\text{cam } j}$ are themselves sparse, but we do not use this fact since they are relatively small in dimension.

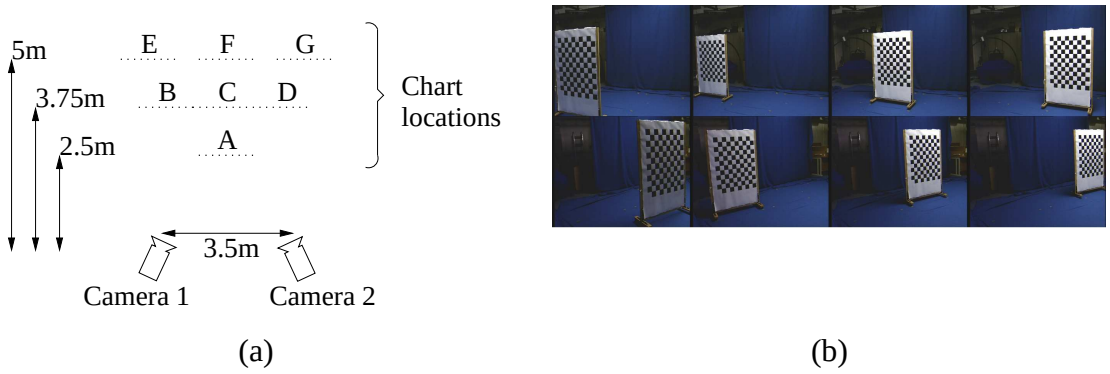


Fig. 3. Plan view of 2-Camera system for wand calibration evaluation (a), and example test images (b)

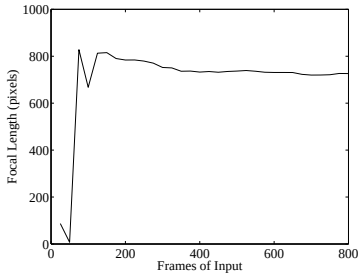


Fig. 4. Focal length estimate as a function of number of input frames

A. Stability with respect to number of frames

Initial tests were needed to find a suitable number of frames of wand data to use for calibration when using real data. This was done using a two-camera system. The cameras were placed approximately 3.5m apart, with viewing directions parallel to the floor, angled in towards the centre of the viewing volume by $\sim$25 degrees, as shown in figure 3(a). The usable measurement volume extended up to $5m$ away from the cameras, and was approximately $3.5m$ wide at this distance.

The wand was waved at a ‘fast walking pace’ through the capture volume for a total of 800 frames (32 seconds of capture). Example input data is shown in figure 1.

Figure 4 shows a typical plot of estimated focal length of one of the cameras versus number of frames used as input. The large variations at low frame counts are caused by initialisation errors, which occur due to instabilities in the eight-point algorithm. Calibration appears stable

after as few as 200 frames.

Refinement of the camera parameters using bundle adjustment usually converges within 2-5 iterations, although up to 15 iterations were needed for some cases where less than 400 points were used and initialisation was poor. To allow a safe margin, 600 frames (24 seconds) was chosen for further experiments.

Noise on a static marker location was also measured, and was found to be 0.18 pixels RMS at a distance of 3m from a camera, under normal studio illumination. The RMS residual error of wand calibration for two cameras was found to be up to 0.5 pixels, which is larger than one would expect based on static marker noise alone.

B. Chart Calibration

The chart calibration method used for comparison was the ‘Camera Calibration Toolbox for MATLAB’ of J. Bouget [4]. Calibration of a multi-camera system requires several images of a known calibration chart for each camera as shown in Figure 3(b) of squares with side length 80mm.

The corners of squares on the chart are extracted using a method based around the Harris corner detector [28] and some manual intervention to identify chart orientation. An initialisation stage based on the work of Zhang [2] gives estimates for internal camera parameters, which are refined using Levenberg-Marquardt bundle adjustment (see section III-D). For a consistent global calibration, one chart location must be common to all cameras. This is used to define a global co-ordinate system and camera extrinsics. It should be noted that estimation of extrinsics for chart calibration could be improved using multiple charts images visible from subsets of cameras. For the purposes of this evaluation publicly available chart calibration using a single chart to estimate extrinsics has been used for comparison.

We use 10 different locations of the chart for each camera, plus an 11th location common to all cameras. An example input for one of the cameras is shown in figure 3. In all cases the method converged to give a residual pixel reprojection error less than 0.3 pixels. Bouget’s toolbox allows

[4] `http://www.vision.caltech.edu/bouguetj/calib_doc`

for models of lens distortion and principal point location estimation, these were disabled for purposes of comparison with the wand calibration. Chart calibration is expected to give the most accurate calibration of intrinsic parameters such as lens distortion. If chart calibration is used to estimate the intrinsic camera parameters estimates of lens distortion and centre of projection could be included in the wand-calibration for estimation of extrinsic parameters. Chart calibration provides an estimate of the projection matrix P_i for sets of cameras which can observe the same chart. This estimate is used to compare the accuracy of chart and wand calibration, noting the limitations of the chart calibration used.

C. Accuracy Tests

Two types of test for positional accuracy are used, one using the chart and the other using the wand. To give a fair test independent data sets are used for calibration and accuracy evaluation.

1) Chart Reconstruction Accuracy: The 2D rms reprojection error of the reconstructed points for a new set of chart observations provides a measure of the reconstruction accuracy:

$$E_{\text{rms}} = \left(\frac{1}{MN} \sum_{i=1}^{N} \sum_{j=1}^{M} \left(u_{ij} - \frac{\mathbf{p}_{i1}\mathbf{x}_j}{\mathbf{p}_{i3}\mathbf{x}_j} \right)^2 + \left(v_{ij} - \frac{\mathbf{p}_{i2}\mathbf{x}_j}{\mathbf{p}_{i3}\mathbf{x}_j} \right)^2 \right)^{\frac{1}{2}} \tag{25}$$

The true 3D point locations $\mathbf{y}_j$ are known with respect to the co-ordinate system of the calibration object. These two point sets can be compared to give an estimate of positional accuracy, but their co-ordinate systems must first be aligned. This is done using the least squares method of Arun *et al.* [29]. The residual error E_{chart} of this fitting is taken as a suitable indication of accuracy of calibration. It is defined:

$$E_{\text{chart}} = \left(\frac{1}{M} \sum_{j=1}^{M} \left| \mathbf{x}_j - \mathbf{y}'_j \right|^2 \right)^{\frac{1}{2}} \tag{26}$$

where $\mathbf{y}'_j$ is 3D point location after alignment. The locations chosen to perform this test were changed depending on camera configuration (see sections VII-D and VII-E).

2) Wand Length Measurement Accuracy: The wand is waved through the volume to be tested, and novel 2D marker points extracted using the same method as used for calibration, described in section IV. The 3D location of each wand marker is calculated, again using appendix A followed by gradient-descent refinement. The distance d_i between the estimated 3D locations is

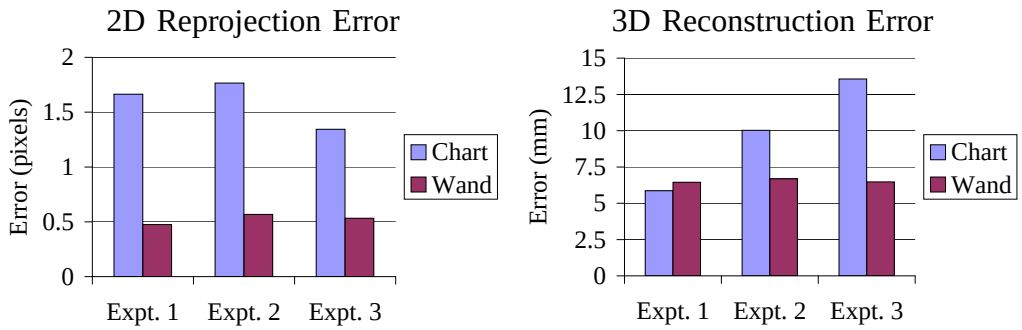


Fig. 5. Chart reconstruction error for 2-camera chart and wand calibration methods

calculated. The indication of error E_{wand} is taken to be the root-mean-squared difference between estimated marker distance and the true marker distance d_0 which was measured using a ruler to $\pm$0.5mm. It is defined:

$$E_{\text{wand}} = \left(\frac{1}{N} \sum_{i=1}^{N} (d_i - d_0)^2 \right)^{\frac{1}{2}} \tag{27}$$

D. Accuracy Tests using Two Cameras

Since chart calibration is really designed for small numbers of cameras, usually facing the same direction, our first accuracy test uses the two-camera system described in section VII-A. The system was calibrated using the wand and using the chart. The chart was then placed upright in the 7 different locations shown in figure 3 in order to evaluate rms reprojection error E_{rms} and chart reconstruction accuracy E_{chart}. To check consistency in calibration and evaluation, the whole experiment was repeated 3 times. Overall 2D reprojection errors and 3D reconstruction errors for each experiment are shown in fig 5. This result indicates that in the 2-camera case the wand calibration achieves a reconstruction error comparable to a chart calibration which estimates extrinsics using a single chart location visible from all cameras. Wand calibration for the two camera case gives an rms reprojection error of 0.5 pixels and 3D reconstruction error of 6mm at $3 - 5m$.

Wand calibration took around 5 minutes to complete (including data capture and storage), whereas a typical time for chart calibration is around 25 mins due to the need to physically move the chart between data captures, and manual extraction of chart points.

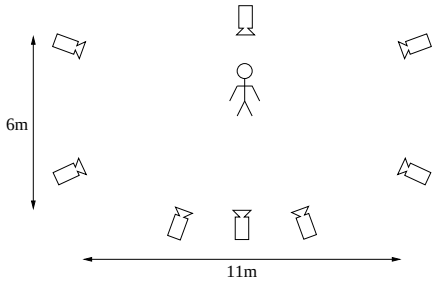
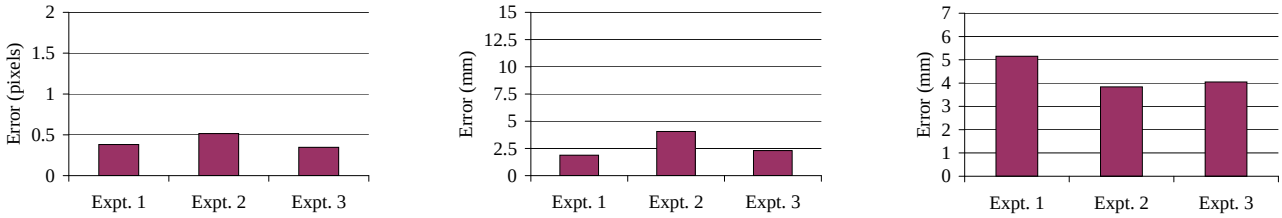


Fig. 6. Plan view of 8-Camera system



(a) Chart 2D reprojection error (b) Chart 3D reconstruction error (c) Wand 3D wand length error

Fig. 7. Evaluation of reconstruction error for 8-camera wand calibration using the chart (a),(b) and wand length (c).

E. Accuracy Tests using Eight Cameras

A system of 8 cameras was arranged around the capture volume, according to the plan view in fig 6. All cameras were placed at a height of $\sim$2m from the floor, angled downwards and towards the centre of the volume. The capture volume occupied a region of around 3m $\times$ 2m in the centre of the cameras.

The system was calibrated using the wand which took around 7 minutes due to the extra computational overhead created by the additional cameras. The residual error was between 1.0 and 1.5 pixels RMS.

With this camera setup, it was not feasible to place the chart upright in view of all cameras. So only 3-camera subsets could be used to evaluate errors E_{rms} and E_{chart}. A more informative error evaluation is the measurement of wand length error E_{wand} which was calculated using a new set of 600 frames as test data as the wand moved throughout the whole volume. As before, the experiment was repeated three times, and the results shown in figures 7.

F. Convergence with respect to inaccurate initial estimates

Simulated data was used to examine the performance of bundle adjustment when incorrect initial estimates are used. This may happen when applying theorem 3.3 in the case where principal rays come close to intersecting, or if manual initial estimates are used. Cameras were simulated with a separation of 3.5m, the second one having an orientation of $\exp(0.1, 0.3, 0.2)$ with respect to the first, where exp is the function given in equation 16. Simulated data for a wand of length 0.5m was created, by positioning the centre of the wand at psuedo-random uniformly distributed points in a cuboid, and uniformly distributed angular orientation. Cameras were given simulated focal lengths of 600 and 900 pixels respectively.

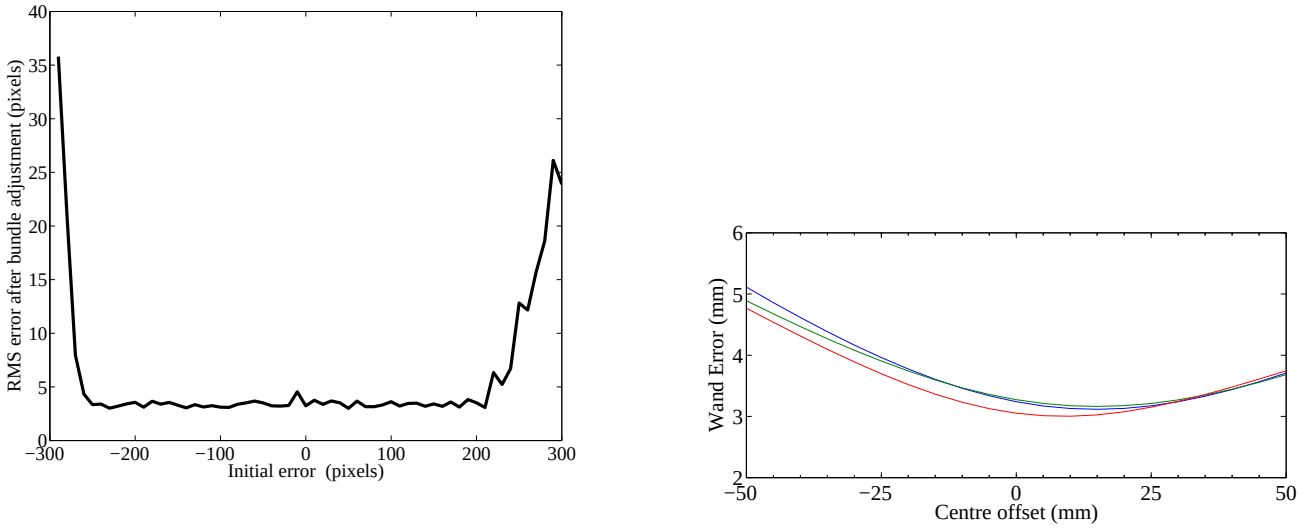
Experiments were conducted whereby bundle adjustment was applied to a simulated set of 400 wand measurements, using an incorrect initial estimate of focal length. Such an experiment was repeated using 100 different sets of pseudo-random data for each simulated initial error. Simulated initial errors of -300 to +300 pixels were used. Figure 8(a) shows the results of root-mean-squared deviation in focal length from the true value, after bundle adjustment. It can be seen that in this case the bundle adjustment converges to the true value even when the initial errors are as large as 280 pixels.

In practice the bundle adjustment does appear stable. In all practical experiments using 600 frames of input data, it was found that the focal length estimation from theorem 3.3 could be bypassed and all initial focal lengths set to a nominal value of 600 pixels. Bundle adjustment still converged within ten iterations to the correct values which lay over the range 400 to 800 pixels.

G. Dependence on Centre of Projection

So far it has been assumed that the centre of projection of all cameras is known or that a suitable approximation can be made, such as setting the centre of projection to be the image centre. An experiment was conducted to evaluate the sensitivity of the method to this choice of parameter.

A set of 600 frames each were acquired using a fixed eight-camera setup similar to the one shown in figure 6, and used as input to the calibration algorithm. A further 1200 frames of



(a) Stability to errors in initial focal length (b) Stability to errors in centre-of-projection

Fig. 8. Evaluation of stability of calibration to (a) errors in initial focal length and (b) errors centre of projection for three real data sets

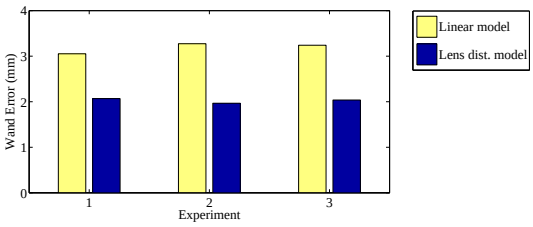


Fig. 9. Comparison of linear model with first-order lens distortion model

data were then acquired to evaluate the quality of the calibration, according to the wand length accuracy measure from equation 27.

This experiment was repeated three times, and results in figure 8(b) show the wand length accuracy as a function of centre of projection offset applied to all cameras. In these experiments the centre of projection is offset along the diagonal line $u = v$ in each camera, but similar results are obtained for offsets in any given direction. Figure 8(b) shows that the calibration accuracy is insensitive to the centre of projection over a wide range of values. Leading to the conclusion that accurate calibration can be achieved with a fixed estimate of the centre of projection.

H. Lens Distortion Model

In all lens-based optical systems some lens distortion is present, and can be incorporated into the camera model as required. To demonstrate this, some results are shown using a radial lens distortion function of image co-ordinates:

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + L(r) \begin{bmatrix} u - u_0 \\ v - v_0 \end{bmatrix} \tag{28}$$

Here (u, v) is the pixel location resulting from the pinhole model (equation 1), and $(\tilde{u}, \tilde{v})$ are the final pixel co-ordinates after lens distortion. $L : \mathbb{R} \mapsto \mathbb{R}$ may be any analytical function $L(r) = 1 + \kappa_1 r + \kappa_2 r^2 + \ldots$, where the values of κ_i are model parameters which must be estimated during calibration. In this work a first-order model, $L(r) = 1 + \kappa_1 r$ is used. This was incorporated in the bundle-adjustment by extending each function $\mathbf{f}_{ij}$ (equation 18) to include the lens function from equation 28, so that $\mathbf{f}_{ij}$ has the additional lens distortion paramter κ_{i1} corresponding to camera i. Note that the corresponding changes must be made to the formulae for the Jacobian $\mathbf{J}$.

Again an eight-camera system was used to acquire data sets of 600 points. These were given as input to both the linear projection model and the model with lens distortion. The residual RMS reprojection errors were reduced by around 30% when the lens distortion model was used. Estimated values of κ_1 were of order $-0.0001\text{pixels}^{-1}$. Further sets of 1200 points were used as input to evaluate E_{wand} using the linear and lens-distortion models. This experiment was repeated three times, and results are shown in figure 9. The results show metric errors are reduced by around 30% when lens distortion is modelled to first-order. This demonstrates that the wand calibration can be extended to estimate lens distortion parameters in the bundle-adjustment refinement leading to improved calibration accuracy.

I. Discussion

Evaluation of the two camera calibration (figure 5) indicates that reconstruction errors are greater when using a chart calibration than using a wand calibration. This does not imply that the chart is a fundamentally less accurate calibration tool - it is thought to be a reflection of the fact that only a single pair of images is used to define the external parameters of both cameras. Use of

multiple images for external parameter estimation may provide superior results from the chart calibration approach. Nevertheless, wand calibration compares well with chart calibration using a single chart for extrinsic parameter estimation used in this experiment.

It is significant that for neither chart nor wand calibration did the residual error approach the estimated error of feature extraction. This is a reflection of imperfections in the calibration model, which may be due to lens distortion, inaccuracies in the choice of centre of projection, or errors in feature extraction due to artefacts such as motion blur. It was found that reprojection errors and accuracy estimates were improved by incorporation of a first-order lens distortion model, and higher order models may improve this further.

There was considerable variability in results of chart reconstruction using the 8-camera system, as suggested by figure 7(a) and (b). Manual inspection revealed noisy feature extraction caused by the oblique angle presented by the chart to some cameras. Measurement of wand length shown in figure 7(c) appears to be a more consistent and meaningful indication of accuracy. Again this is the result of using a single chart for estimation of extrinsics and does not provide a general indication of the performance of chart calibration.

For wand-calibration the rms error for wand-length is approximately 5mm. This value represents the combined positional error from both markers, and so the accuracy in estimation of a single point is approximately $\frac{1}{\sqrt{2}}E_{\text{wand}}$ indicating an rms error in 3D of approximately 3mm. Our figures include measurements near the edge of the capture volume, where lens distortion effects are likely to be most significant, and measurement resolution may be lowest.

Calibration of an eight camera system using the wand was relatively fast (7 minutes). An equivalent chart-based calibration is likely to take 30-60 mins, since up to 81 different chart locations may be required. However, use of chart-calibration for the eight camera configuration was not possible due to the requirement for the chart to be simultaneously visible in all views or overlapping sets of views.

The results in section VII-F show the method performing well even with poor initialisation. This is attributed to the choice of parameterisation and the use of the wand-length constraint. The computation of initial estimates using theorem 3.3 or exhaustive search may therefore be seen as an ‘optional extra’, necessary only when no reasonable initial estimates are known.

Figure 8(b) shows the influence of the choice of principal point on the accuracy of the final result. We note that an offset of 30 pixels from the nominal value increases the wand length deviations by 1.0mm, corresponding to an increase of 0.7mm on resolution of position measurement. Clearly it is desirable to obtain correct centre of projections where possible, but the contribution due to a 30 pixel error is 0.023% of the total field of view, which for many applications would not be significant. Figure 9 shows that introduction a lens distortion model in the bundle-adjustment refinement increases calibration accuracy.

VIII. Conclusions

We have introduced a new wand-based method for calibration of a multi-camera studio. Accuracy for estimation of extrinsic parameters and focal length gives a sub-pixel reprojection error of approximately 0.5 pixels rms with a marker localisation error of 0.2 pixels rms. The overall positional accuracy was $\pm$3mm rms over 3m$\times$2m$\times$2m using an 8-camera system, using measurements taken throughout the capture volume. This level of accuracy is achieved using the pinhole camera model, calibrated for focal length and camera extrinsics using the wand. Further results show that performance is insensitive to the choice of camera centres of projection, and that calibration remains stable in the presence of significant errors in the initial parameter estimates. Camera lens distortion may also be estimated as part of the method, and results show this gives further improvements in accuracy. A major advantage of the wand calibration method is speed of use, taking 7 minutes for an 8-camera system. Unlike chart techniques, the method allows cameras which face each other or have non-overlapping fields of view. Wand-based calibration is therefore a versatile approach for rapid calibration of multiple camera systems, with an accuracy comparable to widely-used chart-based techniques.

Appendix

A. Appendix: Simple reconstruction from calibrated cameras

Suppose we have multiple cameras whose projection matrices $\mathbf{P}_1 \dots \mathbf{P}_N$ are known, and let $\mathbf{u}_1 \dots \mathbf{u}_N$ be the locations of a 3D point in the 2D images from all camera. It is possible to

estimate the 3D location $\mathbf{x}$ of the point by minimising the algebraic error:

$$E_{\text{alg}} = \sum_{i=1}^{N} |\lambda_i \mathbf{u}_i - \mathbf{P}_i \mathbf{x}|^2 \tag{29}$$

Let $\mathbf{p}_{i1}$, $\mathbf{p}_{i2}$, $\mathbf{p}_{i3}$ be the rows of projection matrix $\mathbf{P}_i$. Now define a matrix $A \in \mathbb{R}^{2N \times 4}$ with rows:

$$\mathbf{a}_{2i-1} = \mathbf{p}_{i1} - u_i \mathbf{p}_{i3} \tag{30}$$

$$\mathbf{a}_{2i} = \mathbf{p}_{i2} - v_i \mathbf{p}_{i3} \tag{31}$$

Then the value of $\mathbf{x}$ which minimises equation 29 is the eigenvector of matrix $A^{\mathsf{T}}A$ which has the smallest eigenvalue.

B. Appendix: Solution of Sparse Matrix System

This is a fast method to solve the system of equations:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{32}$$

where D is a block diagonal matrix of the form:

$$D = \begin{bmatrix} d_1 & 0 & \dots \\ 0 & d_2 & \dots \\ 0 & 0 & \ddots \end{bmatrix} \tag{33}$$

First, compute D^{-1} by inverting each block d_i:

$$D^{-1} = \begin{bmatrix} d_1^{-1} & 0 & \dots \\ 0 & d_2^{-1} & \dots \\ 0 & 0 & \ddots \end{bmatrix} \tag{34}$$

This is fast since each block has low dimension. Now compute matrix $\overline{A} = A - BD^{-1}C$ and vector $\overline{b} = b_1 - BD^{-1}b_2$ and solve the system

$$\overline{A}x_1 = \overline{b} \tag{35}$$

to get x_1. Using this value of x_1, solve the system:

$$Bx_2 = b_1 - Ax_1 \tag{36}$$

to get x_2.

It is stated in the literature that a more numerically stable method is to perform a complete numerical factorisation rather than block matrix inversion [4], but the method described here was found adequate to allow Levenberg-Marquardt bundle adjustment to converge for any valid set of input data.

REFERENCES

[1] R. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.

[2] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[3] J. Heikkil and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Proc. Computer Vision and Pattern Recognition*, 1997, pp. 1106–1112.

[4] B. Triggs, P.McLauchlan, R. Hartley, and A. Fitzgibbon, *Bundle Adjustment – A Modern Synthesis*, ser. LNCS. Springer Verlag, 2000.

[5] R. Hartley, "Estimation of relative camera positions for uncalibrated cameras," in *Proc. European Conference on Computer Vision*, 1992, pp. 579–587.

[6] R.Hartley, *Extraction of Focal Lengths from the Fundamental Matrix*. GE CRD, Schenectady, NY, 1993.

[7] O. Faugeras, Q.-T. Luong, and S. Maybank, "Camera self-calibration: theory and experiments," in *Proc. European Conference on Computer Vision*, 1992, pp. 321–334.

[8] J. Weng, T. S. Huang, and N. Ahuja, *Motion and Structure from Image Sequences*. Springer-Verlag, 1993.

[9] Q.-T. Luong and O. Faugeras, "Camera calibration, scene motion and structure recovery from point correspondences and fundamental matrices," INRIA, 2004 route des Lucioles, B.P. 93, 06902 Sophia-Antipolis, France, Tech. Rep., 1994.

[10] M. Pollefeys, R. Koch, and L. V. Gool, "Self calibration and metric reconstruction in spite of varying and unknown internal camera parameters," *International Journal of Computer Vision*, vol. 32, no. 1, pp. 7–25, 1999.

[11] S. Mahamud, M. Herbert, Y. Omori, and J. Ponce, "Provably-convergent iterative methods for projective structure from motion," in *Proc. Computer Vision and Pattern Recognition*, 2001, pp. 1018–1025.

[12] P. Sturm and B. Triggs, "A factorization-based algorithm for multi-image projective structure and motion," in *Proc. European Conference on Computer Vision*, 1996, pp. 709–720.

[13] T. Kanade and D. Morris, "Factorisation methods for structure from motion," *Phil. Trans. Royal Society of London*, vol. A, 1997.

[14] H.-G. Maas, "Dynamic photogrammetric calibration of industrial robots," in *Proc. SPIE Videometrics V*, vol. 3174, 1997, pp. 106–112.

[15] P. Baker and Y. Aloimonos, "Complete calibration of a multi-camera network," in *Proc. IEEE Workshop on Omnidirectional Vision (OMNIVIS00)*, 2000, pp. 11–20.

[16] T. Svoboda, "Quick guide to multi-camera self-calibration," ETH, Swiss Federal Institute of Technology, Zurich, Tech. Rep. BiWi-TR-263, http://www.vision.ee.ethz.ch/ svoboda/SelfCal, 2003.

[17] Z. Zhang, “Camera calibration with one-dimensional objects,” in *Proc. European Conference on Computer Vision*, 2002, pp. 161–174.

[18] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2000.

[19] H. Longuet-Higgins, “A computer algorithm for reconstructing a scene from two projections,” *Nature*, vol. 293, 1981.

[20] J.J.More, “Levenberg-Marquardt Algorithm: Implementation and Theory,” *Springer Lec.Notes in Mathematics*, vol. 630, 1977.

[21] R. Buck, *Advanced Calculus*. McGraw, 1965.

[22] J. Mitchelson, “Multiple Camera Studio Methods for Automated Measurement of Human Motion,” PhD Thesis, University of Surrey, UK, Tech. Rep. http://www.ee.surrey.ac.uk/CVSSP/VMRG/Publications/mitchelon03phd.pdf, 2003.

[23] D. Forsyth and J. Ponce, *Computer Vision - A Modern Approach*. Prentice Hall, 2003.

[24] M. Fischler and R. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Communication ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[25] Z. Zhang, “Determining the epipolar geometry and its uncertainty: A review,” *International Journal of Computer Vision*, vol. 27, no. 2, pp. 161–195, 1998.

[26] R. Murray, Z. Li, and S. Sastry, *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994, ch. Appendix A: Lie Groups and Robot Kinematics.

[27] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C : The Art of Scientific Computing*, 2nd ed. Cambridge University Press, 1992.

[28] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Proc. Alvey Vision Conference, Manchester*, 1988, pp. 147–151.

[29] K. Arun, T. Huang, and S. Blostein, “Least squares fitting of two 3-D point sets,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 5, pp. 698–700, 1987.