Multi-camera parameter tracking

F.Pedersini, A.Sarti and S.Tubaro

Abstract: The quality of 3-D reconstructions with multi-camera acquisition systems is strongly influenced by the accuracy of the camera calibration procedure. In fact, when acquiring a long sequence of views, mechanical shocks, vibrations and thermal gradients could cause a significant drift of the camera parameters. The authors propose a method for tracking the camera parameters and, whenever possible, correcting them accordingly. This technique does not need any *a priori* knowledge or test objects to be positioned in the scene, as it exploits natural scene features. The approach is based on accurate detection, matching and back-projection of luminance corners and spots in the scene space. Such features are then tracked over time to detect unexpected parameter changes or drifts, and to apply corrections to them. Experimental results on real sequences are reported in order to prove the effectiveness of the proposed technique. It is shown that changes in the calibration parameter are correctly detected and, when this happens, the camera system can be re-calibrated with an accuracy that increases with the number of tracked feature points.

1 Introduction

The problem of the 3-D reconstruction of objects from the analysis of a number of digital images has been long studied and the available solutions can greatly differ from each other depending on the application for which they are designed. Three main types of camera systems are usually considered: a single camera that freely moves around the scene acquiring images from different viewpoints, a set of two or more cameras mounted on a rigid frame that can move around the scene, or a fixed set of many cameras mounted on a rigid dome-like frame that surrounds the scene to be reconstructed. In the first case, the camera parameters (camera position and orientation, focal length, pixel size and lens distortion coefficients) must be estimated together with the 3-D scene description using the available images [1–3]. Conversely, in the last case the position and the orientation of the cameras are fixed (the camera frame is not only rigid but is a fixed structure) and so are their intrinsic parameters [4]. This enables a separation between the estimation of the camera parameters and the estimation of the 3-D scene shape, making it possible to calibrate only when the acquisition structure is set up. Indeed, while the main advantage of this approach is to allow the complete reconstruction of a dynamic scene, the invasivity of the acquisition structure represents its main drawback.

In between such extreme cases is the portable multi-camera rig, whose camera parameters are also expected to be fixed [5, 6]. Indeed, in this situation we can always decide to adopt an uncalibrated approach and jointly estimate camera parameters and 3-D scene shape. However, the rigidity constraint between the positional parameters of the cameras suggests to us that it would be best to estimate them separately through calibration, together with the other intrinsic parameters. This operation is done by acquiring one or more multi-views of a calibration target-set, with targets of known shape and 3-D location [7, 8]. This way of decoupling camera calibration and 3-D scene reconstruction usually leads to a better estimation of the 3-D information compared to the single camera case, as the calibration is carried out on targets that are optimally scattered in the scene and whose shape is known beforehand, which makes their image detection and localisation far more accurate. All this, of course, is done at the price of a heavier acquisition system. One problem to be aware of, however, is that the stability of the initially estimated camera parameters can become critical when acquiring a long sequence of images. The camera calibration is, in fact, rather sensitive to mechanical shocks, vibrations and even thermal changes of both cameras and frame. This drift of calibration parameters can easily cause a significant 3-D reconstruction error, as the 3-D back-projection process is an ill-conditioned operation with respect to the camera parameters. To minimise the parameter drift, we could adopt a heavy and rigid camera frame, with a significant increment of cost. This, however, would make the acquisition system even more cumbersome to handle. A more reasonable solution is thus to try to detect and track any changes in the acquisition system and, if possible, to correct the camera parameters 'on the fly'. This way the calibration will hold accurate throughout the whole acquisition session.

In this article we will focus on calibrated multi-camera acquisition systems (particularly trinocular systems) and we will show how to detect camera parameter changes through the analysis of scene features. To achieve this goal, we will describe our approach based on detection, matching, back-projection (onto the object space) and tracking of natural point-like features. The method does not need any special test objects to be placed in the scene or any *a priori* knowledge about it, but exploits luminance features that are already present in the scene (e.g. corners and spots),

IEE Proceedings online no. 20010140

DOI: 10.1049/ip-vis: 20010140

Paper received 3rd December 1999

The authors are with the Image and Sound Processing Group (ISPG), D.E.I. - Politecnico di Milano, Piazza L. da Vinci 32, I-20133, Milano, Italy

which are accurately localised on the image plane. The super-resolution localisation process is followed by a matching procedure that returns n-tuples (n being the number of cameras) of homologous feature points. The matched n-tuples are then back-projected onto the 3-D scene space. As any camera parameter change causes an unexpected back-projection error, meaning larger than the predicted error after pre-calibration, we can reveal and characterize incidental changes of the camera parameters through a proper analysis of the magnitude and the temporal behavior of the back-projection errors. We adopted and extended this idea in a rather straightforward fashion to correct the calibration parameters 'on the fly', whenever possible. Currently, our technique is able to correct the parameters in two possible ways, depending on the situation. If the camera system remains still during the acquisition session, then parameter re-calibration is done on the scene's 'fixed points', otherwise the scene's stable points are tracked and used to perform self-calibration.

The proposed technique has been extensively tested on real sequences acquired with a trinocular camera system, with both simulated and real changes of the camera parameters, providing very encouraging results. In all the experiments we conducted the algorithm was able to correctly detect the camera parameters changes. Moreover, with all the typical parameter changes associated with accidental shocks, changes of focal length, etc., the algorithm was able to correctly quantify the parameter drift and re-calibrate the system.

2 Camera models and camera calibration

A camera model is defined as the set of mathematical relationships that link the 3-D co-ordinates of a point in the scene space to the 2-D co-ordinates of its projection on the acquired image. Such relationships can be defined in a number of different ways, as the literature shows. Among them, a distinction could be made between those that define an operator (e.g. a projection matrix) that links the co-ordinates of a 3-D point to the co-ordinates of its image projection using homogeneous co-ordinates [5, 9, 10], and those that define a model by directly using all the optical and geometric parameters of the camera [11]. The camera model that we adopted belongs to this latter group, which is represented in Fig. 1. This choice was motivated by our interest in assigning a precise physical meaning to each parameter that belongs to the camera model. This can be particularly useful when we need to straightforwardly use all the *a priori* information on the camera parameters. For example, if the adopted acquisition system uses a lens with nominal focal length of 16 mm, this information can be effectively used to improve the reliability and the accuracy of the calibration procedure. Moreover, the adoption of a 'physical' model allows us to directly and immediately judge the calibration results through a comparison between estimated parameters and our rough knowledge of the physical camera characteristics, such as position, orientation, focal length, etc.

A generic camera model is specified by a vector of parameters [7]:

$$\boldsymbol{CP} = [\varphi, \theta, \psi, t_X, t_Y, t_Z, f, \boldsymbol{k}, c_X, c_Y] \quad (1)$$

whose first six elements are the extrinsic camera parameters, i.e. the Euler angles that specify the camera orientations and the three world co-ordinates of the camera's optical centre. The other (intrinsic) parameters are: the focal length, the lens distortion coefficients and the image location of the principal point, which is the intersection between the optical axis and the image plane. Often only radial lens distortion is considered, in which case $\boldsymbol{k}$ (see eqn. 1) has only one or two coefficients [7]. It is important to notice that the pixel size is assumed as known [11]. This is a reasonable assumption for most commercial digital cameras. When using an analogue CCD camera with known pixel size, however, we also need to know the ratio between pixel frequencies of the CCD sensor and the frame grabber. The optimal solution is to lock the frame grabber's pixel clock to the internal pixel clock of the camera sensor [12].

The estimation of the camera parameters is carried out through the analysis of views of a test object (calibration target-set). The target set usually consists of a set of fiducial marks, also called targets, positioned within the 3-D volume that is being imaged by the camera system (see Fig. 2).

A simple calibration approach can be used and trusted when the knowledge on the 3-D co-ordinates of the targets is complete and accurate. When, on the contrary, the information on the target positions is very little or absent, some self-calibration approach is used (see [7]).

It is important to emphasise that the estimated parameters of the acquisition system are expected to hold accurate only for measurements within the 3-D volume 'spanned' by the specific calibration target-set [13]. In fact, roughly speaking, the target-set plays the role of a training set for the simple calibration procedure, and therefore it should be chosen in such a way to be 'statistically representative' of the scene to be reconstructed. As a consequence, to achieve high accuracy in the calibration and in the 3-D reconstruction, it is important for the targets to properly 'fill up' the entire volume that will be later occupied by the object to be measured.

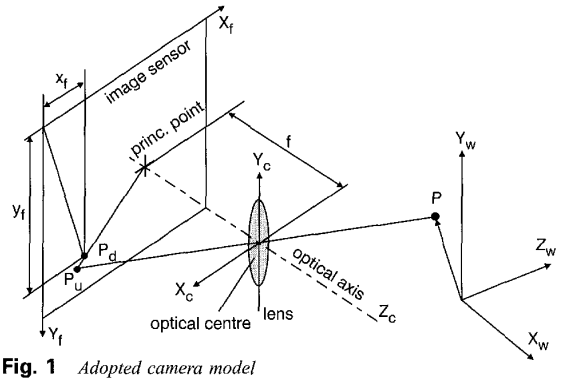

Fig. 1 *Adopted camera model*

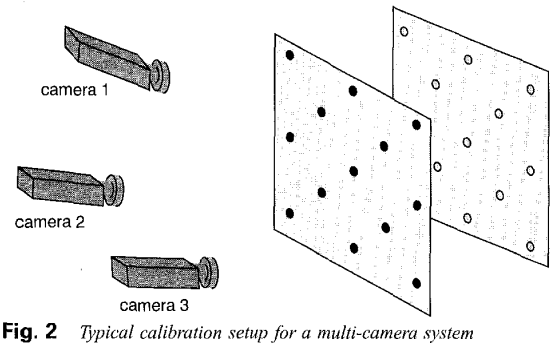

Fig. 2 *Typical calibration setup for a multi-camera system*

A set of fiducial marks are placed in front of the acquisition system, to fill the 3-D scene space

From a practical standpoint, both simple calibration and self-calibration can be seen as a way of exploiting a large number of constraints that cumulate in a space made of a large number of co-ordinates. The constraint equations are those that force the projection of a target onto an image plane, computed through a camera model equation, to correspond to its actual image co-ordinates. In fact, the projection of a 3-D point onto an image plane gives rise to a pair of equations (one per image co-ordinate). It is customary (and advisable) to use a redundant number of fiducial points with respect to the number of unknowns, so that the model space will result in being overconstrained [11].

In between the two extremes of traditional calibration and self-calibration there are solutions that work like self-calibration in less of a strict sense. For example, a rigid planar surface with circular marks is used as a target-set in [7], where the relative locations of the targets are only partially known. In that case, several views of the target-set are taken from unknown positions, so that the union of all targets fills up all the whole calibration volume. This idea can be pushed to the extreme of a simple bar with only two targets [14], which can be used for calibrating the acquisition system through the analysis of a sequence of multi-views, acquired while the bar was moving throughout the calibration volume.

3 Calibrated multi-view reconstruction

A calibrated multi-camera system can be used for determining the 3-D co-ordinates of scene features through a back-projection of their image location [5]; see Fig. 3. Because of the unavoidable localisation errors and the limited calibration accuracy, however, the optical rays associated with corresponding image features fail to meet exactly in one point (see Fig. 3), therefore the scene feature has to be localised as the point of minimum distance from three such rays [15]. Indeed, the back-projection's accuracy is strongly and directly influenced by the accuracy of the camera calibration. The influence of camera calibration on the reconstruction's quality, however, can also be indirect as, due to the modelled nonlinear lens distortion, what ideally should be an epipolar line is warped into a curve of an amount that depends on the distortion parameters. We should not forget, in fact, that the distance between a point and the epipolar line associated with its homologous is an index of accuracy of the calibration parameters, and the temporal tracking of the calibration's accuracy is, in turn, of crucial importance if we need to guarantee a good 3-D reconstruction quality throughout a long acquisition session.

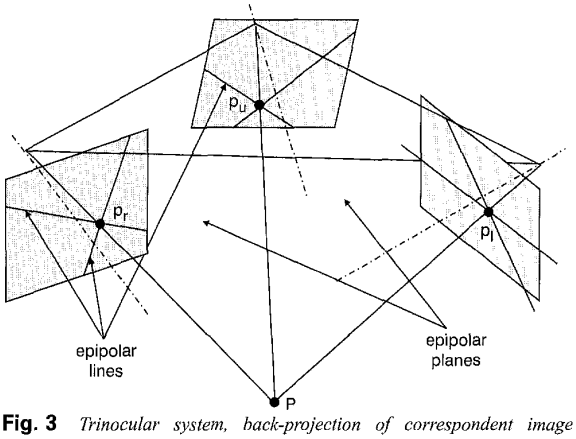

Fig. 3 *Trinocular system, back-projection of correspondent image points (p_l, p_u, p_r) to identify the correspondent 3-D point (P)*

We assume that the image co-ordinates have been corrected from lens distortion effects. For simplicity the image plane has been designed in front of the correspondent optical centre rather than behind these points. For each image point the corresponding epipolar line is drawn on the other images

4 Camera parameter tracking

As already mentioned above, when using a calibrated multi-camera system, the correctness of the model parameters can be evaluated by looking at the accuracy of the back-projection of stereo-corresponding image features. When a parameter drift (due, for example, to significant thermal changes or mechanical vibrations) or a sudden but modest parameter change (due, for example, to a mild mechanical shock) is detected, recalibration can be performed to correct our knowledge of the epipolar geometry. To do so, feature localisation and tracking need to be performed throughout the acquisition session. Notice that, although well-localisable features can always be artificially added to the scene, it is usually preferable to use natural image features to minimise the invasivity of the acquisition. Indeed, this is only possible if an accurate (sub-pixel) feature localisation technique is available.

The strategy that we present in this Section is not strictly based on the assumption that there is temporal continuity in the images. As a matter of fact, we only assume that our multi-camera system acquires a number of still images from different (but not too different) viewpoints. In fact, the only continuity assumption that we rely on is in the calibration parameters, which may undergo a significant drift or be subjected to an abrupt change of modest entity. However, if the acquisition consisted of a video sequence, then the temporal continuity could be quite easily exploited in the determination of the feature correspondences, by adopting some feature tracking approach.

The proposed parameter tracking technique consists of the following steps:

(i) *Pre-calibration* — Aimed at determining the initial calibration parameters, the calibration approach that we adopt is based on a planar target-set [7, 16], although other types of calibration procedures can be employed instead. This choice requires the acquisition of a set of views of the target-set before the actual acquisition session begins. This preliminary acquisition could be avoided by running self-calibration on natural image features, provided that a sufficient number of them be available for this purpose.

(ii) *Feature detection and localisation* — The natural image features that we consider are corners, as they can be well identified and localised and they are very likely to be viewer-independent. This allows us to reliably perform the initial feature matching [17] and track them throughout the acquisition session.

(iii) *Matching and backprojection* — If we assume that the camera calibration is still fairly correct, we can help a correlation-based matching approach with epipolar constraints to determine the correspondences between image features. Matched points are then back-projected in the space to estimate their 3-D positions.

(iv) *Accuracy evaluation* — The validity of the camera parameters is checked through the analysis of the back-projection accuracy, which describes how well corresponding optical rays intersect in 3-D space, as seen in Fig. 3.

(v) *Accuracy analysis* — The temporal evolution of the back-projection accuracy is analysed in order to reveal an increment of the back-projection error that could likely

denote a change in the parameters of the acquisition system.

(vi) *Camera parameters update* — If parameter correction is needed and an adequate number of accurately detected and matched image features is available, then camera re-calibration is performed. To do so, when the 3-D location of the image features is known (i.e. obtained when the system was still calibrated) and referenced to the current camera position by using the time feature tracking capabilities of the system, we use simple calibration. It is also possible to refine (or estimate) this 3-D information by using a self-calibration approach.

4.1 Feature extraction and matching

The accurate detection of image features (to be used as control points) is often required in applications of 3-D reconstruction [17]. Spot detection is encountered when dealing with features that have been artificially added to the scene, and can be performed through template matching [12]. The method that we developed for detecting features that are naturally present in the scene searches for vertices (crossings between edges that can be extracted from the luminance image profiles). Several approaches have been proposed in the literature for the extraction of this type of image feature, which can be broadly divided into two groups. Algorithms belonging to the first group are based on the extraction of luminance edges and on the detection of points of maximum edge curvature. The second group of algorithms works directly with a grey-level image. In general these techniques works on the analysis of the gradient and/or curvature of the surface that represents the shape of the luminance profiles (see [18] for a review).

In our work we have developed a corner detector that is an extension of the one proposed in [18] that belongs to the second class of algorithms. This choice is due to the fact that with this class of algorithm it is possible to estimate, with sub-pixel accuracy, the position of the feature points even if they are very localised. In fact, using the intersection of straight line segments to locate vertex positions it is necessary that this segment will be very long to ensure high quality estimations. Normally, when we refer to a luminance corner or vertex, we consider intersection of edges with V, Y and T configurations as shown in Fig. 4.

If we model the luminance transitions as smoothed (by means of a null phase filter) step edges, the vertex point is characterised by the fact that the Laplacian of the luminance function is always zero independently of the specific configuration of the edges that meet in the vertex [19]. Furthermore, the Baudet operator

$$DET = \det\begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix} = I_{xx}I_{yy} - I_{xy}^2 \qquad (2)$$

has a relative maximum (in all directions) in the proximity of vertices and, when applied to a set of progressively more filtered versions of the image, the maxima can be shown to lie on a line that intersects the vertex point. Two such constraints can be used jointly for determining a vertex with super-resolution accuracy. To do so, it is possible to look for the zero-crossing of the Laplacian along the line of the maxima of the DET. In [18] the image is, at first, filtered with a low-pass two-dimensional Gaussian filter (f_1), which is optimal for an accurate edge detection in the presence of noise [20]. This filter can be seen as the separable product of two 1-D filters (in horizontal and vertical directions) which are characterised by an impulse response with standard deviation (scale factor) σ_1. After

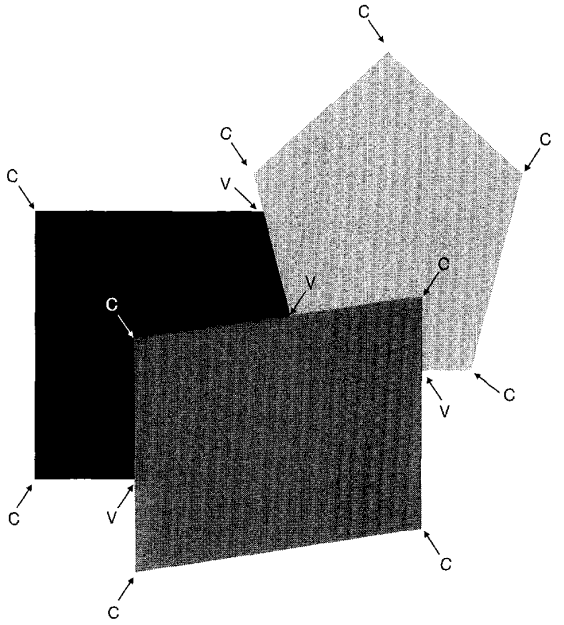

Fig. 4 *Definition of corners and vertex points*
C, corners; V, vertex points

filtering the image, the elliptical maximum (above a certain threshold) of the DET operator is determined. Then a heavier filtering (f_2, with $\sigma_2 < \sigma_1$) is applied to the original images in the proximity of the previously detected DET's maximum, to determine its new location. At this point we can search for the Laplacian's zero-crossing along the line that connects such two maxima, which corresponds to the image location of the corner/vertex. This process can be implemented in such a way as to achieve sub-pixel accuracy. Some experiments have been carried out in the literature to determine the filters to be used for best performance. What was found was that, to extract a significant number of feature points, the magnitudes of σ_1 and σ_2 must be kept modest. This, however, turns out to limit the achievable sub-pixel localisation accuracy of the algorithm.

To overcome this difficulty, we consider four different DET's maxima for each corner, which correspond to four differently filtered versions of the image. The search for the Laplacian's zero crossing can thus be limited to the line of collinearity of such maxima, determined through linear regression. The sub-pixel co-ordinates of each maximum are determined using quadric interpolation about the pixel position of the maximum. Interpolation is performed over a grid of 3×3 samples as larger regions would not improve the result due to the very complex profile of the DET in the proximity of a vertex. To determine the exact location of the Laplacian's zero-crossing along the line of maxima, we adopt a polynomial approximation of a Laplacian's 1-D profile. Furthermore, to reduce the impact of the noise, instead of computing the Laplacian of the original luminance profile, we use the most mildly Gaussian-filtered version of this profile [18].

One important aspect to consider when implementing a high-accuracy corner detector is the isotropy of the spectral response of the acquisition system. As a matter of fact, it often happens that the cameras exhibit different spectral characteristics in the horizontal and the vertical directions. The acquisition system used for this work, for example, was based on a set of three black-and-white CCD cameras

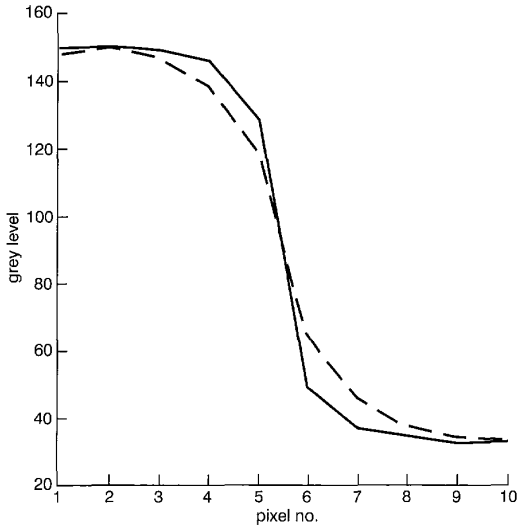

Fig. 5 *Typical luminance profiles associated with horizontal and vertical edges*

—— horizontal edges
– – – vertical edges

(Sony XC77CE) of standard TV resolution, connected to a frame grabber. Fig. 5 shows the luminance profiles, as generated by our camera system, when imaging an abrupt luminance transition in the vertical and the horizontal directions. As we can see, the horizontal profile (vertical edge) turns out to be smoother compared with the vertical one (horizontal edge). This difference in bandwidth, which is to be attributed to the fact that scan lines are subject to analogue filtering in the camera circuitry [21], can cause a significant reduction in the localisation accuracy of the corner/vertex. One simple way to avoid this problem is to use different scale factors in the horizontal and vertical Gaussian filters.

To quantify the accuracy of our corner extractor, we conducted a number of experiments on images acquired by a well calibrated trinocular system based on TV-resolution cameras. After feature detection, we selected a number of stereo-corresponding corners/vertices on the three images. From the detected image co-ordinates of two corresponding points, we predicted the location of the homologous point on the third image using calibration information. The difference between detected and predicted image locations of the homologous feature point on the third image was taken as a quality descriptor of our vertex/corner extractor. We obtained the best results using the following parameters: $\sigma_{h1} = 1.8$ pixels, $\sigma_{h2} = 1.5$ pixels, $\sigma_{h3} = 1.2$ pixels, $\sigma_{h4} = 0.9$ pixels, $\sigma_{v1} = 2.1$ pixels, $\sigma_{v2} = 1.8$ pixels, $\sigma_{v3} = 1.5$ pixels and $\sigma_{v4} = 1.2$ pixels, where the subscripts h and v denote horizontal and vertical filtering, respectively. As we can see, the vertical spread factors σ_v are 0.3 pixel wider than the corresponding horizontal ones. The corner deviations that we obtained with the above parameters are 0.15–0.2 pixels.

Fig. 6 shows the various steps of our corner localisation algorithm, while Figs. 7 and 8 show the results of our feature extractor on two typical images acquired with our trinocular system.

Once the features are correctly extracted, we can apply an n-partite matching algorithm to automatically determine the stereo-corresponding n-tuples [22]. The matching criterion is based on the similarity of the local luminance profiles (correlation-based matching) [17]. The epipolar

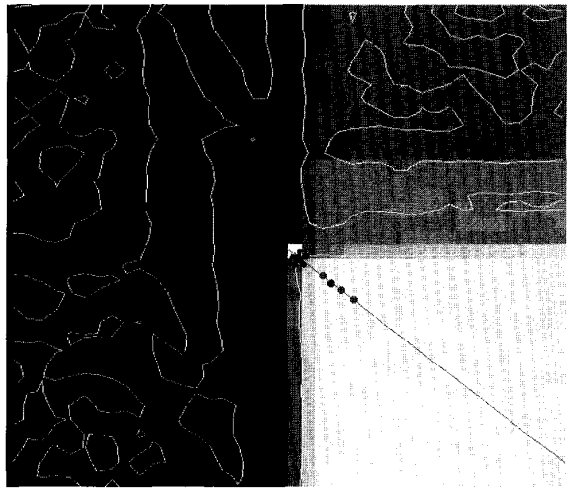

Fig. 6 *Zoomed-in details in neighbourhood of a luminance corner*

Circles denote detected sub-pixel locations of DET's maxima; closed curves are Laplacian's zero-crossings; white square denotes pixel location of corner; star denotes its sub-pixel location as detected by the algorithm

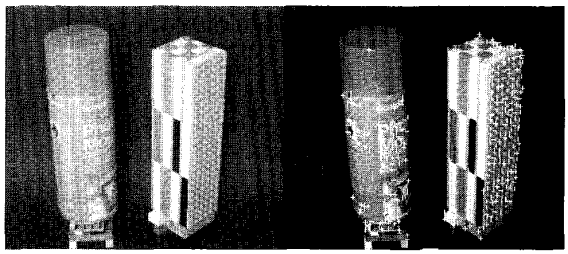

Fig. 7 *Example of application of our corner detector*

Original view (left) and detected feature points (right)

Fig. 8 *Example of application of our corner detector*

Original view (left) and detected feature points (right)

geometry defined by the current calibration is used only to reduce the search space for corresponding features, as is cannot considered as reliable. In fact, if we consider a feature point on one image, the corresponding features on the other views are searched for in wide regions surrounding the epipolar lines.

4.2 Analysis of back-projection's quality

Ideally, stereo-corresponding visual rays should meet exactly in one point in 3-D space (see Fig. 3). However, problems of inaccuracy in the camera calibration parameters and in the feature localisation, together with problems of noise in the available images, prevents this from happening. If we keep track of how 'close' the visual rays are to intersecting each other, we can detect unexpected changes in the position/orientation of the cameras.

To do so, an 'accuracy index' is computed from the back-projection error (average minimum distance between visual rays) associated with each matched n-tuple of points (in our experiments, triplets of points). For each acquisition time, the statistical distribution of the back-projection error over the matched n-tuples and the temporal evolution of the global 'accuracy index' are analysed to reveal any anomalous increment of these indexes that could very likely denote a change in the system parameters.

4.3 System re-calibration

When a parameter change is detected, re-calibration is triggered. To do so, some sort of temporal tracking of the image features can be of help. What we do is to first perform corner extraction on all available still pictures. Correspondences between features are then determined between consecutive stills of the same camera, using a correlation-based approach. We consider the luminance profile within a square region (typically 8×8 pixel) centred on a corner of the previous image, and look for the homologous region in a given search area of the current image, still centred on the same feature. As we can see, the temporal correspondences are determined using an approach that is quite similar to that used for the determination of correspondences between features of simultaneous views. Some congruence checks, however, are in order:

(*a*) *Consistency*: The displacements of all corners between consecutive views of the same camera must be consistent. This means that the motion vectors must agree with a single model. Our choice of motion model, which is adequate for describing pan and zoom displacement [23], is as follows:

$$x_n = x_{n-1} + \alpha + \beta x_{n-1}$$
$$y_n = y_{n-1} + \alpha + \gamma y_{n-1}$$

where (x_{n-1}, y_{n-1}) and (x_n, y_n) are the co-ordinates of a feature on the previous and current images, respectively, while α, β, γ are the model parameters. The motion parameters of this model can be estimated through linear regression [24] on the feature displacements. This estimation process is made particularly robust by a phase of elimination of the 'outliers' from the list of the tracked features, which are those displacement vectors that disagree with the estimated model. Should the number of 'inliers' become insufficient, the tracking would be declared as unreliable.

(*b*) *Smoothness*: If a n-tuple of features is declared as stereo-corresponding at a certain time instance, then the tracked features at the next time instance is taken as stereo-corresponding.

(*c*) *Rigidity*: The distance between tracked and back-projected features must be preserved, otherwise the tracking is restarted.

As far as the re-calibration process is concerned, two different situations are here considered. The first and simplest scenario is based on the assumption that the camera system is not subjected to a significant rigid motion with respect to the scene throughout the acquisition session. Therefore, in the first part of the acquisition session, some features whose 3-D back-projections appear as stable in time are automatically selected and used as control points. These are the candidate points to be used as primary 3-D targets for parameter correction (re-calibration). When a parameter change is detected, the entire current set of matched n-tuples of image features is used for recovering the new camera parameters. Depending on the available knowledge of the 3-D position of the matched points, the algorithm adopts either a calibration or a self-calibration approach. More precisely, if the time tracking of the image features worked correctly (as it normally occurs), then the 3-D positions of the primary targets and of the other features present on the current images, and whose positions have been well estimated at the previous time instances (when the system was calibrated), are used as inputs for a calibration procedure like the one used before the beginning of the image acquisition [7]. Obviously the primary targets have more weight in the calibration process with respect the other considered features. Moreover, the calibration process is speed-up by the fact that the new camera parameters are not searched from scratch, but from the values that they had before the detected change. If, on the contrary, due to problems in the feature temporal tracking (for example caused by significant changes in the scene illumination), no reliable information is available about the real 3-D positions of the features matched among the images available at the current time instance, then a self-calibration approach [3, 7] must be used to update the camera parameters. Self-calibration allows us to simultaneously determine the camera parameters and the 3-D positions of the fiducial points, but it is not a completely 'blind' procedure. In fact, also in this case, the new camera parameters can be searched for starting from the previously available values. This corresponds to assuming that it is important to be able to recalibrate the system without interrupting the acquisition session, provided that changes in camera position, orientation or focal length are of modest magnitude. This guarantees that self-calibration will not turn out to be ill-conditioned, although the system performance will be a bit worse than with simple calibration.

The second scenario is the one in which the camera undergoes a significant rigid motion. A typical example is the one in which the entire camera system is mounted on a dolly and moved around the scene. The situation, in this case, is slightly more complex, as each image acquisition corresponds to a different set of camera parameters. However, since the motion estimation of the multi-camera system between consecutive acquisitions [25] is separate from the calibration, the changes of positional parameters that we need to analyse are the relative displacements between cameras. Once again, the procedure for the detection of the parameter changes and for the system's recalibration is very similar to the one proposed in the previous sub-sections. As seen with the first scenario, time-tracking of the feature points can be of help in ensuring better performance. Obviously, when the camera system moves, some features that are visible in the current n-tuple of views could disappear in the next one, and new features could appear. In this case, the primary targets used with a fixed camera system become useless, and therefore the algorithm must be able to deal with these situations.

5 Experimental results

To validate the proposed technique, we tested it with two different trinocular systems. The first test was conducted with an acquisition system based on three Sony DCX950 TV-resolution colour cameras, each with a $1/2''$ 3-CD sensor. The cameras were mounted on a rigid frame at the vertices of a triangle with a baseline of 800 mm and the other two sides of ~ 500 mm. The volume to be calibrated was about ~ 2 m wide, 1.5 m deep and 1.5 m tall, placed at

an average distance of 2 m from the camera set. To precalibrate this camera setup, we used a high-quality target-set, made of a grid of 29 × 20 fiducial points (centres of black circular stickers with a radius of 12.5 mm), placed on the surface of an aluminium 'wafer' with honeycomb structure (for improved rigidity and light weight). The exact world co-ordinates of the targets had been measured through classical photogrammetric methods (whose accuracy was better than 0.1 mm).

The second test was performed with a low-cost system based on three B/W Sony XC77CE cameras, with 2/3 CCD sensors, and a nominal focal length of 16 mm, placed on a rigid frame at the vertices of a triangle that was approximately 40 cm tall and had a baseline of ~60 cm. We considered a scene volume of ~60 × 60 × 60 cm, placed at an average distance of ~80 cm from the camera set. In order to pre-calibrate this camera system, we used an inexpensive target-set made of a printed paper sheet glued on a flat 60 × 60 cm surface. The targets were nominally positioned at the cross-points of a square grid with a step size of 40 mm and the extracted image co-ordinates were those corresponding to the centres of the circular dots.

As the images were collected by a PC through a frame grabber, this second system had an acquisition frame-rate of about on image triplet every 15 s. Conversely, the first camera system was connected to digital video recorders, and therefore the frame rate was up to 25 frames per second. In Figs. 7 and 8, two examples of scenes considered for the experiments are shown. The first scene was acquired with the low-cost B/W trinocular system, while the second scene was acquired with the colour camera system. Fig. 9 shows the features with which a reliable stereo-correspondence is detected at a certain time instance.

Two types of tests have been performed on the acquired trinocular sequences: in the former case some calibration parameters are artificially modified at a certain time instance to simulate the effects due to a change of the acquisition set-up characteristics (geometrical and/or optical). In the latter case, during the acquisition of the sequence, the camera set-up is physically modified, by changing the relative pose of the cameras on their rigid frame and by slightly changing their focal length. In both cases the parameters of only one camera had been modified; more specifically, the orientation angles had been changed by ~2–3°, while a change of ~2–3% had been introduced on the focal length.

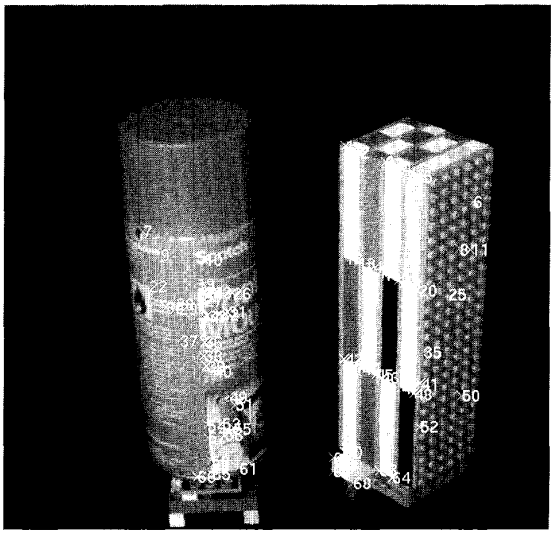

Fig. 9 *Image features for which a stereo-correspondence match has been identified at a certain time instance, denoted by crosses*

Table 1: Results of the analysis of the 3-D back-projection errors relative to the trinocular system using TV-resolution B/W cameras

Calibration method	Mean BP error, mm	Std. dev. — BP error, mm
Standard calibration	0.0520	0.0240
Recalibration	0.1020	0.0401
Self-calibration	0.1016	0.0400

The results achieved show that, in both cases, the system was able to immediately detect the changes in the camera parameters. This was revealed by a significant increase in the average accuracy index. It is worth noticing that the number of matched points was not significantly reduced by the parameter change, and therefore we could use most of the matched points as fiducial points and correct the calibration parameters using such points. To test the accuracy of the corrected calibration parameters, a new feature matching was performed and the accuracy index was again evaluated and compared with the initial one. The 3-D back-projection errors relative to the trinocular system that uses B/W cameras are collected in Table 1. The back-projection errors are here written in the form of average errors (on the considered image triplet) and of standard deviations. The first row of the table shows the results that are obtained with a standard calibration, which is performed using an ad hoc target structure (see [7]); the second row shows what happens after system recalibration, which is possible when 3-D information on the 'natural target points' is available; finally, the third row corresponds to self-calibration, which is applied when no information on the features is available. These results confirm that the accuracy of the corrected parameters is comparable with that of the original calibration in both cases of re-calibration and self-calibration. Moreover, self-calibration is able to achieve slightly better results than re-calibration, because the greater degree of freedom of the self-calibration is exploited in the final part of the search of the best-fitting solution, which often leads to a better minimum.

6 Conclusions

In this paper we have proposed a technique for tracking the camera parameters through the analysis of luminance features that are naturally present in the scene. The method is based on sub-pixel feature localisation, followed by feature matching. The accuracy of the back-projection of homologous features onto the 3-D space is used as an index of quality for deciding whether or not to proceed with the correction of the parameters of the acquisition system.

The proposed technique was proven effective through tests performed on real sequences acquire with trinocular camera systems. Successful experiments were conducted with both simulated and real camera parameter drift, proving the method suitable for adaptive calibration.

Further research is being conducted to improve the performance of the accurate feature detection strategy and add features of different nature. We are also focusing on self-calibration without any calibration target-set.

7 References

1 VAN GOOL, L., and ZISSERMAN, A.: 'Automatic 3D model building from video sequences', *Eur. Trans. Telecommun.*, 1997, **8**, (4), pp. 369–378

2 ZHANG, Z., and FAUGERAS, O.: '3D dynamic scene analysis' (Springer-Verlag, 1992)

3 GRUEN, A., and BEYER, H.: 'System calibration through self-calibration'. Invited paper, Workshop on *Camera Calibration and Orientation in Computer Vision* XVII ISPRS Congress, August 1992, Washington DC

4 SING BING KANG, , WEBB, J.A., ZITNICK, C.L., and KANADE, T.: 'A multibaseline stereo system with active illumination and real-time image acquisition'. Proceedings of IEEE Int. Conf. on *Computer Vision*, June 1995, Cambridge, MA, USA, pp. 88–93

5 AYACHE, N.: 'Artificial vision for mobile robots' (MIT Press, 1991)

6 PIGAZZINI, P., PEDERSINI, F., SARTI, A., and TUBARO, S.: '3D area matching with arbitrary multiview geometry', *Signal Process. Image Commun.*, 1998, **14**, (1, 2), pp. 71–94

7 PEDERSINI, F., SARTI, A., and TUBARO, S.: 'Accurate and low-cost calibration and self-calibration ofmulti-camera acquisition systems', *Signal Process.*, 1999, **77**, (3), pp. 309–334

8 WENG, J., COHEN, P., and HERNIOU, M.: 'Camera calibration with distortion model and accuracy evaluation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1992, **14**, (10), pp. 965–980

9 YAKIMOWSKY, Y., and CUNNINGHAM, R.: 'A system for extracting three-dimensional measurements from a stereo pair of TV cameras', *Comput. Graph. Image Process.*, 1978, **7**, pp. 195–210

10 WEI, G.Q., and DE MA, S.: 'Implicit and explicit camera calibration: Theory and experiments', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1994, **16**, (5), pp. 469–480

11 TSAI, R.Y.: 'A versatile camera calibration technique forhigh-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses', *IEEE J. Robot. Autom.*, 1987, **RA-3**, (4), pp. 323–344

12 BEYER, H.A.: 'Geometric and ratiometric analysis of a CCD-camera based photogrammetric close-range system'. PhD thesis no. 51, Institut für Geodäsie und Photogrammetrie, ETH, Zürich, May 1992

13 FERRIGNO, , BORGHESE, N.A., and PEDOTTI, A.: 'Patterns recognition in 3D automatic human motion analysis', *ISPRS J. Photogram. Remote Sens.*, 1990, **45**, pp. 227–246

14 BORGHESE, N.A., and CERVERI, P.: 'Calibrating a video camera pair with a rigid bar', *Pattern Recognit.*, 2000, **33**, (1), pp. 81–95

15 PEDERSINI, F., PIGAZZINI, P., SARTI, A., and TUBARO, S.: 'Multi-camera motion estimation for high-accuracy 3D reconstruction', *Signal Process.*, 2000, **80**, (1), pp. 1–21

16 PEDERSINI, F., SARTI, A., and TUBARO, S.: 'Multi-camera systems: calibration and applications', *IEEE Signal Process. Mag.*, 1999, **16**, (3), pp. 55–65

17 ZHANG, Z., DERICHE, R., FAUGERAS, O., and LUONG, Q.T.: 'A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry'. RR-2273, INRIA Research Report, May 1994

18 DERICHE, R., and GIRAUDON, G.: 'A computational approach for corner and vertex detection', *Int J. Comput. Vis.*, 1993, **10**, (2), pp. 101–124

19 DE MICHELI, E., CAPRILE, B., OTTONELLO, P., and TORRE, V.: 'Localization and noise in edge detection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1989, **11**, pp. 1106–1117

20 CANNY, J.: 'A computational approach to edge detection', *IEEE Tran. Pattern Anal. Mach. Intell.*, 1996, **8**, (6), pp. 679–698

21 BARBE, D.: 'Imaging devices using the charge-coupled concept', *Proc. IEEE*, 1975, **63**, (1), pp. 38–67

22 MALUCELLI, F., and PRETOLANI, D.: 'Efficient labelling algorithms for the Maximum Non Crossing Matching Problem'. Internal Report, Department of Informatics, University of Pisa, Italy, 1990

23 MIGLIORATI, P., and TUBARO, S.: 'Multistage motion estimation for image interpolation', *Signal Process. Image Commun.*, 1995, **7**, pp. 187–199

24 STAUDTE, R.G., and SHEATHER, S.J.: 'Robust estimation and testing' (Wiley, 1990)

25 PEDERSINI, F., SARTI, A., and TUBARO, S.: 'Egomotion estimation of a multicamera system through line correspondence'. Proceedings of IEEE International Conference on *Image Processing* 1997 (ICIP-97), October 1997, Santa Barbara, CA, USA, pp. 26–29