Journal of Cultural Heritage 1 (2000) 301–313

1296-2074(00)01082-7/FLA

Automatic monitoring and 3D reconstruction applied to cultural heritage

Federico Pedersini*, Augusto Sarti, Stefano Tubaro

Dipartimento di Elettronica e Informazione (DEI), Politecnico di Milano, Piazza L. Da Vinci 32, 20133 Milan, Italy

Received 12 April 2000; accepted 20 August 2000

Abstract – In this article we present our global approach to the problem of accurate 3D measurement and reconstruction of 3D works of art using a calibrated multi-camera system. In particular, we illustrate a simple and effective adaptive technique for the self-calibration of CCD-based multi-camera acquisition systems with minimum *a-priori* information. We also propose a general and robust approach to the problem of close-range partial 3D reconstruction of objects from stereo-correspondences. Finally, we introduce a method for performing an accurate patchworking of the partial reconstructions, based on 3D curve matching. © 2000 Éditions scientifiques et médicales Elsevier SAS

Keywords: 3D reconstruction / CDD-based / multi-camera acquisition / works of art

1. Introduction

Three-dimensional works of art are characterized by such a diversity of shapes, materials and physical structures that, as of now, their accurate 3D relief can only be performed through expensive and time-consuming ad-hoc methods. This fact, together with the need of employing specialized personnel, reflects negatively on the whole costs of the 3D modeling process, which ends up being applied to the most urgent cases only.

Due to the large number of historical objects and monuments that need to be inspected, monitored, modeled, restored, and protected, the need of developing low-cost, automatic, and accurate methods for 3D modeling is becoming urgent. As a consequence, one need that is becoming more urgent is that of developing and implementing non-invasive systems that allow non-expert users to construct a 3D model of a monument, a statue, a portion of a building, or a medium/large-sized work of art, in an automated fashion and with photogrammetric accuracy, through the analysis of images of the object itself, acquired on-site.

* Correspondence and reprints:
E-mail address: pedersin@elet.polimi.it (F. Pedersini).

The rush to develop methods for the 3D reconstruction of objects from the analysis of camera images has been particularly intense in the past two decades. Particularly in the past few years, a large number of these applications have been developed with the goal to approach the problem of content creation for the multi-media market. There is a considerable number of applications, however, in which the accuracy of the 3D reconstruction plays a crucial role, especially those applications of close-range digital photogrammetry aimed at the preservation and restoration of 3D works of art. Such methods, require effective techniques for accurate, quantitative, reproducible, and repeatable 3D reconstruction. In fact, suitable 3D modeling methods should be sufficiently accurate as to match the performance of the methods that are commonly adopted for the 3D relief of works of art; and to guarantee that such measurements will be reproducible and can be repeated a long time for monitoring purposes.

One aspect that heavily influences the types of methods that can be used for 3D reconstruction from images is the size of the object under examination. Usually, objects that are rather small in size (fractions of a meter) exhibit peculiarities and com-

plex characteristics of complexity that dramatically differ from case to case. In order to model such objects, a variety of systems that are not image-based have been developed and can be used successfully in controled environments. Large-sized objects (several meters or more), on the other hand, are more difficult to model as they need to be imaged from viewpoints that are sometimes far apart from each other. On the other hand, such objects are usually rather well-structured (e.g. architectural or historical buildings), therefore they are best dealt with by exploiting some *a-priori* knowledge on their structure. In between these two extremes is a wide variety of mid-sized objects (between some fractions of a meter and several meters). Examples of 3D objects that are usually within this range, and normally they very modestly structured, are historical monuments and 3D works of art, for which we cannot exploit *a-priori* information.

The most popular non-invasive approaches to 3D reconstruction of mid-sized objects are based on stereo-correspondences. Such methods consist of the detection of special features (e.g. points, edges) on the available images of the object. When the camera parameters (position, orientation, and other intrinsic physical parameters) are known (*calibrated* case), the process of determining the correspondences is helped by some rigidity constraints such as the coplanarity of corresponding visual rays (epipolar constraint), and the 3D coordinates of the features can then be determined through geometric triangulation [1, 2]. When, on the contrary, the camera parameters are not known (*uncalibrated* analysis), the determination of the feature correspondences becomes more difficult. In this case, in fact, the determination of the correspondences becomes a statistical matching process based on heuristic rules, while the epipolar constraints is now used for the joint estimation of the camera pose and the 3D coordinates of the features.

Another correspondence-based approach is stereopsys, i.e. on the matching of luminance profiles that pertain to small image areas in the available views [3]. Once again, the 3D coordinates of the surface patch that originated the corresponding luminance profiles are determined through geometric triangulation, while the matching process is performed by maximizing a similarity function between the luminance profiles. These methods usually suffer from problems of matching ambiguity when only two views are employed, while they become more accurate with more cameras, provided that geometric and radiometric distortions are taken into account [4, 5].

The problem of matching ambiguity is present in all correspondence-based methods (feature, and area matching) that use an uncalibrated approach or a calibrated pair of cameras. In order to overcome this difficulty, a calibrated set of three or more cameras mounted on a rigid frame can be used. In fact, three is the minimum number of views with which it is possible to exploit the multi-ocular invariance [1], i.e. the constraint according to which each one of three corresponding points in three views is bound to lie on the intersection of the epipolar lines (i.e. the views of the optical rays) associated to the other points. This is, in fact, quite a reasonably strong constraint and can be used for all calibrated feature matching methods. Other forms of multi-ocular invariance can be found, for example, for line matching, and used for removing ambiguities in the matching process [1].

In general, the 3D reconstruction methods based on feature matching can be classified into two categories:

- *monocular approach*: a series of uncalibrated views are taken in as a sequence or randomly, and then processed all together (*global* approach) or in subgroups (*local* approach) in order to jointly estimate camera motion and object's structure. In the global approach, one or more cameras are employed for acquiring a number of images of the object from a variety of viewpoints [6]. The pose of the cameras and the 3D coordinates of the features are found through a joint analysis of feature correspondences between all available views. In the local approach a video sequence of the object is acquired in such a way to 'cover' all portions of the object. Then the views are partitioned into groups to be processed independently in the same way, using uncalibrated methods and concepts of invariance theory.
- *multi-ocular approach*: a set of cameras are mounted on a rigid support and calibrated, so that all camera parameters are known beforehand. A number of medium-high resolution multi-views of the object are acquired from a variety of views. Each multi-view is analyzsed individually and generates a 'local' surface. All local reconstructions are then fused together into a single surface, by using some global constraints [6–8]

In general, the global monocular approach is the one that estimates the 3D coordinates of some object's features with the best accuracy, as it is based on a joint analysis of all available views. Due to its global treatment of the data, however, this method

produces a sparse set of 3D features that cannot easily be interpolated into a global surface unless some *a-priori* information on the object is available. Partitioning the views into 'good' subsets for a more 'local' approach would result in a heavy reduction of the accuracy and would be difficult to perform on an automatic basis. This partitioning, however, must be applied to the methods based on the analysis of monocular sequences (local monocular approach). Consequently, the simplicity of the acquisition process of such methods is paid for in terms of a reduced optimal positioning of the views that constitute the subgroup of images for the local reconstruction. In fact, consecutive views of video sequences are likely to be 'aligned' with each other and, therefore, not optimally positioned for feature matching purposes.

The local multi-camera approach, exhibits some characteristics that make it quite interesting:

- the multi-camera acquisition system induces a 'natural' partition of the views; if the cameras are well-positioned on the rigid frame (e.g. three cameras at the vertices of a regular triangle), such partitioning will be optimal;
- the acquisition system can be quite easily calibrated, and the estimated parameters can be used for to safely determining feature correspondences between the views through the epipolar constraint; furthermore the calibration can be made adaptive in order to compensate for the drift of the parameters throughout the acquisition process;
- although the highest accuracy can only be reached through a joint simultaneous analysis of all available views, the local 3D reconstruction accuracy resulting from the analysis of a well-calibrated triplet of views is very close to it (photogrammetric accuracy);
- each calibrated triplet generates a partial 3D surface model (*patch*) that corresponds to the imaged portion of the whole surface. This 'local' 3D modeling approach is suitable for a high level of automation as the 3D patches are topologically easier to deal with than the whole surface. In fact, once all 3D patches are available, the global surface can be obtained through 'patchworking'.

In this article we present a summary of the results of our research activity on problems of 3D reconstruction from multiple camera views, conducted within projects related to the areas of cultural heritage and multi-media applications. In particular, the most relevant results have been achieved within the European ACTS-PANORAMA Project (Package for New OpeRational Autostereoscopic Multiview systems and Applications), the strategic project '*Conoscenza per immagini: un'applicazione ai beni culturali*' (knowledge through images: an application to cultural heritage) of the Italian National Research Council, and the project '*Elaborazione e codifica di segnali per sistemi multidimensionali di telecomunicazione'* (signal processing and coding for multidimensional telecommunication systems) of the MURST (Italian Ministry for the University and Scientific and Technological Research).

All calibrated 3D reconstruction methods are critically dependent on the accuracy with which the camera parameters, i.e. the geometrical, optical and electric characteristics of the camera system (camera position and orientation, focal length, pixel size, location of the optical center, nonlinear distortion coefficients, etc.) are known.

In the past few years several approaches to the calibration problem have been proposed. Such methods apply to electronic cameras the same techniques that were traditionally used for the calibration of photogrammetric cameras [9–11]. The camera characteristics are, in fact, computed through a proper processing of the image of a test object (calibration target-frame) placed in the scene. The accuracy of the camera model can be arbitrarily improved by employing an adequate number of parameters, therefore, when the goal is that of improving the calibration accuracy as much as possible, the pattern's accuracy becomes the major bottleneck. For this reason, we developed an advanced photogrammetric method that jointly estimate the camera parameters and the geometry of the calibration target-set in a more accurate fashion. This method is based on a *multi-camera*, *multi-view* calibration approach, and performs an accurate estimation of the parameters of the multi-camera system from the analysis of several views of a simpler calibration target-frame, such as a marked planar surface (a printed sheet of paper glued on a glass surface) or some other even simpler structure. In fact, not only is this technique able to estimate the camera parameters, but it can also determine the 3D position of the targets on the calibration frame, which can be just roughly known or, in some situations, not known at all. Since the methods performs a refinement of the 3D coordinates of the targets, we will refer to it as a *self-calibration* method. Finally, we developed a method for making the calibration robust against the inevitable parameter drift that takes place during the acquisition process. Such method detects and tracks some 'safe' features that are naturally present in the scene, and use their image coordinates for making the calibration process adaptive.

Among the numerous approaches to close-range photogrammetry available today, those that are based on stereo matching seem to be particularly promising. Such methods, however, can usually provide a reconstruction of just a portion of the scene surfaces, while it would be desirable to reconstruct the surfaces of the whole scene. As a matter of fact, automatic 3D reconstruction systems based on stereo-matching can only reconstruct the *visible* portion of the surface. Such systems, in fact, typically provide a description of just the *front side* of the imaged scene or, when the surface is too large to fit simultaneously in all views, of just a limited portion of it. In conclusion, in order to obtain a complete scene reconstruction through stereometry, it is necessary to observe the scene from several significant viewpoints and put together the final reconstruction like a *patchwork* of partial reconstructions.

In order to be able to merge 3D data coming from different reconstructions, we need to accurately estimate the rigid motion that the acquisition system undergoes between two partial reconstructions. In order to do so, one could employ high-precision mechanical devices for positioning the camera system (or the object) before acquiring a multi-view. This *a-priori* solution of the ego-motion problem, however, is usually quite expensive and not very flexible. In alternative, one can perform detection and tracking of some image features throughout the acquisition process, and use the location of such features for estimating the camera motion. This last approach becomes particularly interesting when the features to be extracted are part of the scene to be reconstructed rather than being artificially added to it. Adding special *markers* to the imaged scene is, in fact, common practice in photogrammetry but, besides making the egomotion retrieval more invasive, it requires a certain expertise and slows down the acquisition process [8]. Scene features that can be quite safely detected and are commonly present in natural scenes are luminance edges [6]. These features are more likely to be naturally present in the scene and rather easy to detect, which makes them good candidate features for egomotion estimation.

In order to safely perform patchworking, we developed a method for estimating the egomotion of a multi-camera system from the analysis of 3D contours in the imaged scene. Being as the method is based on a calibrated multi-ocular camera system [9, 11], the estimation is performed entirely in 3D space. In fact, all edges of each one of the multi-views are previously localized, matched and back-projected onto the object space [12]. Roughly speaking, the method searches for the rigid motion that best merges the sets of 3D edges that are extracted from each one of the multiple views

2. Materials and methods

2.1. Calibration

Camera calibration is usually carried out through the analysis of the views of a test object (calibration *target-set*), which usually consists of a set of *fiducial marks* (*targets*), positioned within the 3D volume that is being imaged by the camera system. If the geometrical characteristics of this target-set are only partially known or not known at all, then the calibration process must include the refinement or the blind estimation of the 3D coordinates of the targets.

We developed an advanced calibration method that, besides estimating the parameters of the multi-camera acquisition system, is able to refine a rough *a-priori* estimate of the geometry of the calibration target-set to produce accurate 3D coordinates of the targets (*self-calibration*). The method is based on the analysis of several views of a simple (planar or even linear) calibration target-frame, which is moved throughout the scene in order to emulate a 3D target-frame, and is robust against the inevitable parameter drift that takes place during the acquisition process. In fact, detects and tracks some 'safe' features that are naturally present in the scene, and use their image coordinates for making the calibration process adaptive.

2.2. Calibration strategy

The calibration target-set that we use for calibration is planar as the pixel size is assumed to be known [9]. A planar target-set is much simpler to build with respect to a 3D target-frame as it can be easily constructed, for example, by gluing a laser-printed sheet of paper on a rigid planar surface. This procedure also gives us some *a-priori* information on the coordinates of the targets (and their uncertainty), relative to a frame attached to the surface. A 3D calibration target-frame, on the other hand, would require an accurate 3D measurement of the coordinates of the targets (generally through some photogrammetric technique [11]).

The main drawback of 2D target-sets is the fact that they can only occupy a rather limited volume of the 3D scene. It is well known, in fact, that a reliable camera calibration can only be performed if the targets are not only numerous enough, but also

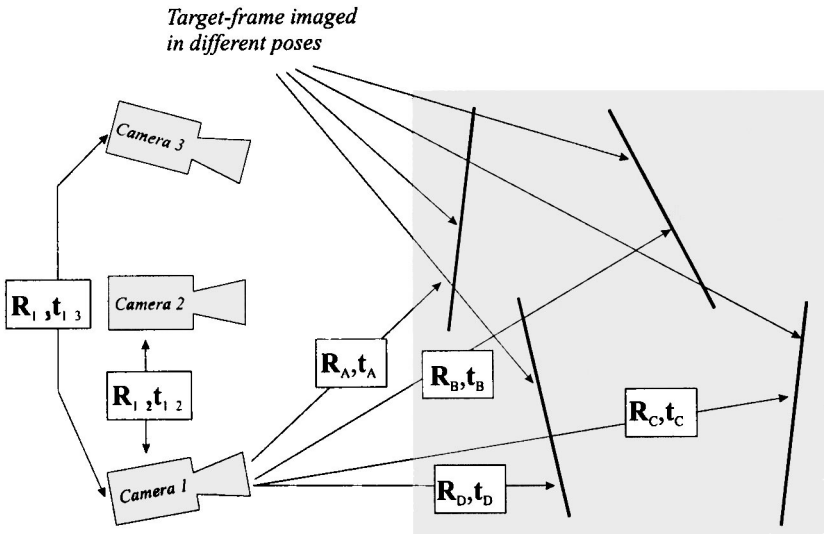

Figure 1. General scheme for the multi-view multi-camera approach to self-calibration.

well distributed in the 3D space that will later be occupied by the object to be measured. In order to overcome this limitation, we proceeded by virtually *enlarging* the planar target-set through the acquisition of several of its views (see *figure 1*). The poses of the target-frame are chosen in such a way that the union of all targets will fill-up the volume of interest in a rather uniform fashion. This strategy, quite clearly, modifies the calibration problem as the relative motion that the target undergoes between acquisitions is not known and needs to be *a-posteriori* determined. In order to do so, the position and the orientation of the target-set (relative to the world reference frame) will be added to the model parameters that need to be estimated for each pose of the target-frame. An example of the application of our parameter estimation approach is reported in *figure 2*, where a laser-printed sheet of paper, glued to a flat surface, is imaged by our multi-camera system. The targets (printed circular dots) are only known in their nominal 2D coordinates, which are then corrected *a-posteriori* by the self-calibration method. The orientation of the (magnified) correction vectors denotes the deformation of the sheet of paper due to the action of the dragging mechanism of the laser printer.

2.3. Adaptive calibration

In order to extract 3D information from the scene views, the camera parameters must be known with good accuracy throughout the whole acquisition campaign procedure. As *camera calibration* is performed before the beginning of an acquisition session, problems of parameter drift may occur. In fact, when long video sequences are acquired, the stability of the camera parameters measured at the beginning becomes a crucial problem as mechanical shocks, vibrations or thermal effects on cameras and

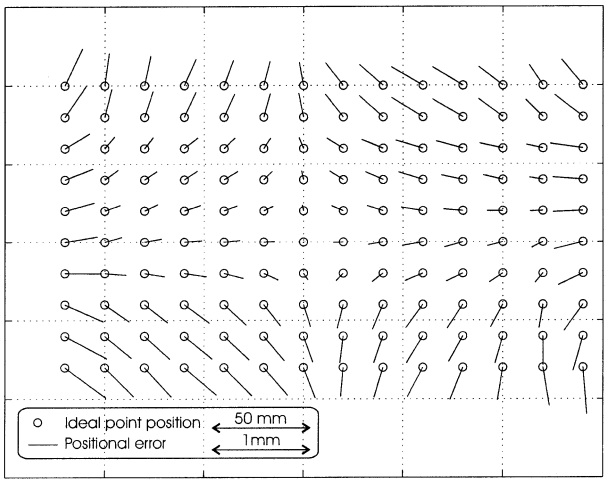

Figure 2. *A-priori* coordinates of the fiducial points of the target-set (laser-printed circles on a sheet of A4 paper, glued to a flat surface) and corresponding *a-posteriori* corrections estimated through self-calibration. The orientation of the (magnified) correction vectors denotes the deformation of the sheet of paper due to the action of the dragging mechanism of the laser printer.

supports, can cause small variations of the initial camera set-up. This drift of the camera parameters leads to significant 3D reconstruction errors, as the 3D back-projection is rather ill-conditioned with respect to the camera parameters. In order to overcome this problem, we detect and track any changes in the acquisition system and, whenever possible, we apply an on-the-fly correction of the camera parameters. By doing so, the calibration holds remains accurate throughout the acquisition campaign.

Our approach does not need any targets to be placed in the scene or any *a-prioria* knowledge, but exploits luminance features that are already present in the scene (e.g. corners and spots) which can be located placed in the image with high precision. After the localization process, which is performed with sub-pixel accuracy, a matching operation is performed among the n sets (n being the number of cameras) of feature points, which returns a set of n-tuples of homologous points. The matched n-tuples will then be back-projected into the 3D scene space. If the camera parameters change, then the back-projection will be affected by larger errors, with respect to the predicted pre-calibration accuracy. A proper analysis of the magnitude and the temporal changes of the back-projection error allows us to reveal and characterize any incidental modifications of the camera parameters. Furthermore, if the set of matched n-tuples is informative enough, the proposed technique allows to accurately measure the occurred modification and, therefore, to re-calibrate the system.

Our approach can be seen as composed of two main steps:

- check on the validity of the current camera parameters through the estimation of the back-projection's accuracy
- analysis of the temporal changes of the back-projection's accuracy, in order to reveal increments in the reconstruction error that could likely denote a change in the system parameter

The first step of the algorithm consists in the detection of the significant image features that will be used as control points. Our method is based on the techniques presented in [13–15]. In order to obtain super-resolution in the image localization accuracy, an algorithm for the local modeling of the image Laplacian function has been developed and employed in the localization procedure. The obtained results show that the introduced improvements has allowed to reach a localization accuracy better greater than 0.2 pixel [16].

Over the obtained sets of image points, a n-partite matching algorithm is applied, in order to find the stereo-corresponding n-tuples. The matching criterion is based not only on the epipolar geometry defined by the current calibration, as the calibration is not considered as reliable in this application, but also on the similarity of the local luminance profiles. All the matched n-tuples are then back-projected in the 3D scene space, and an 'accuracy index' is computed for each match, based on the back-projection error. The statistical distribution of this index over the matched points and its temporal behavior is then analyzed, in order to reveal any increment of the accuracy index that could very likely denote a change in the system parameters. Moreover, at the beginning of the sequence, the back-projected points that are most accurate and are fixed in the scene are selected as *control points*. These are the points that could then be used as 3D fiducial points for the re-calibration of the system. In fact, if the number of matched points is sufficient, it is possible to perform a reliable re-calibration of the system. When a change in the camera system has been detected, the current set of matched n-tuples of image features is exploited, in order to recover the new camera parameters.

The proposed technique was tested on real sequences acquired with different trinocular camera systems, with both simulated and real variations of the camera parameters. In all experimental situations, the algorithm was able to detect the modification of the camera parameters. Moreover, after artificial modifications of the camera system, of the same characteristics and entity of accidental ones (artificial shocks, change of focal length, etc.), the algorithm was able to measure the drift of the parameters, thus allowing the re-calibration of the system. The results showed that the accuracy of the re-calibration, in all cases, reached the same accuracy as the original calibration.

3. Results

3.1. Local reconstruction

The availability of accurate camera parameters, guaranteed by the above adaptive calibration process, allows us to perform a series of 'local' multi-camera acquisitions of the object to be reconstructed. Such local multiple views are taken from different viewing angles, so that all portions of the object surfaces will be visible on at least one of them. Each multiple view will contribute with a

'local' patch of surface, and all local reconstructions will then be glued together in a sort of a global *patchwork*. In this Section we will illustrate our approach to local reconstruction which, in order to produce accurate results, is based on feature correspondences.

Image features that are most often used for 3D reconstruction are points, luminance edges, and luminance patches. These two types of features tend to provide information of a different nature. The edge matching/backprojection process is generally very precise and reliable, but it usually results in a sparse set of 3D points. Conversely, the matching/back-projection of the luminance profile of small image patches tends to provide much denser sets of 3D points but it is rather sensitive to the unavoidable viewer-dependent perspective/radiometric distortions, therefore this approach tends to be less stable and reliable. For this reason we developed a general and robust solution to the problem of 3D reconstruction from stereo correspondence of luminance patches. The method is largely independent on the camera geometry, and employs a calibrated set of three or more standard TV-resolution CCD cameras, which provides enough redundancy for removing possible matching ambiguities. The robustness of the approach can also be attributed to the *physicality* of the matching process, which is actually performed in the 3D space rather than on the image plane. In order to do so, besides the 3D location of the surface patches, it estimates their local orientation in 3D space as well, so that the geometric distortion of the luminance patch can be included in the model. Finally, the method takes into account the viewer-dependent radiometric distortion.

3.2. Edge-based approach

As a preliminary step we perform partial reconstruction from the edge matching, in order to obtain reliable and accurate 3D data to begin with. Furthermore we can use the same type of features for egomotion estimation as well. In fact, partial reconstruction is based on 2D edge matching (stereo correspondence on the image planes) [1, 2, 4, 5], while motion estimation is based on 3D contour matching (edge correspondence in object space) [6, 8]. It is important to emphasize the fact that, in order to be able to use edges for accurate egomotion estimation, we need them to be detected with great accuracy. We do this by first using a traditional edge detector, we then retrieve the subpixel location of the edge points through an interpolation process which takes the luminance gradient into account. Finally, a rule-based contour tracking method is employed for determining the correct connection between edge points.

The search for homologous edges on different views is performed along *epipolar lines*. Notice that using more than two cameras allows us to avoid problems of matching ambiguity. For example, with three cameras, not only can we always select the best pair of views for a specific stereo-correspondence (sharp intersection between edge and epipolar lines), but we can validate the matching through a check on the third view. In fact, the edge point must lie on the intersection of the two epipolar lines associated to the homologous edge points on the other views. Once the stereo correspondences are found, each set of corresponding contours is back-projected onto the 3D scene space by looking for the point at minimum distance from the three homologous visual rays.

3.3. Area-based approach

The luminance patches used by most area-matching techniques are normally assumed to have the same shape in all views. It is quite clear, however, that this hypothesis is acceptable only when the angles between the viewing directions of the three cameras are not too wide, which is not our case. As a consequence, we need to take into account the perspective distortion of the shape of the patch, when back-projected onto the object surface and then re-projected onto the other image plane. In order to do so, we assume the 3D surface to be locally flat, which means that it can be approximated by a plane within the back-projected surface patch.

In the other view we search, along the distorted (due to radial distortion) epipolar line, for the patch that best matches the first one. The projective distortion of the patch is accounted for by estimating, together with the position of the patch, the normal to the object surface, according to which the shape and texture of the patch are most likely to be warped. In practice, the minimum of a *similarity function* between a patch of the actual image and a re-projected patch after perspective warping is searched for as a function of position and local orientation of the tangent plane of the object surface. As far as the radiometric distortion is concerned, an additional pair of variables (luminance *offset* and *gain*) is included in the similarity function.

If a reference patch produces reliable 3D information, then it can be used for 3D surface reconstruction. Once all reference regions have been considered, surface interpolation is carried out and the area matching process can start over with a smaller patch size. In this case the previously estimated surface can be used for initializing the search in the next step and speeding up the process.

As a general rule, we need to make sure that the maximum size of the patch is small enough to guarantee a limited matching error. On the other hand, we know that the area matching process is based on the minimization of a highly nonlinear similarity function, therefore we can expect the process to be quite sensitive to local minima. In order to avoid such a problem, we can use an initial *guess* of the surface shape, which helps the minimization process converge to a global minimum and dramatically speeds up the matching process by reducing the size of the search space.

In principle, any method can be used for obtaining a first guess of the surface shape. In our implementation we opted for an edge-based approach [1, 2], whose reliability is guaranteed by the accuracy of the camera model and the calibration procedure. We adopted a calibrated trinocular camera system [2, 9, 11], which allowed us to select the best pair of views for a specific edge correspondence and validate it through a check on the third view. As the result of the edge-matching approach is usually a sparse, though accurate, set of 3D points. Such data must be interpolated in order to obtain a first guess of the surface to be reconstructed. We interpolated the 3D data by means of a modified and optimized version of the edge-preserving discrete smooth interpolator (DSI) [17]. This interpolation process is used at each step of the surface refinement as well.

Some experiments of 3D scene reconstruction have been carried out on a number of test scenes. The first test presented in this paper concerns the measurement of the accuracy of the area matching using a flat object placed at a distance of about 1.2 m from the camera system. The surface reconstruction resulted to be flat with 0.1 mm of standard deviation (see *figure 3*) [4].

Another reconstruction experiment concerned a large stone (about 1×1 m) of the Roman Amphitheater of Aosta, Italy (see *figure 4*). Due to the size of the object, a patchworking of several partial reconstructions was performed. We also performed a comparative evaluation of the quality of the results and found our local reconstruction results to agree with the measurements taken with classical photogrammetric methods. Notice, however, that the nature of our reconstruction is quite different from that obtained with a photogrammetric relief. In fact, the photogrammetric data were much sparser than ours (our final mesh consisted of approximately 400 000 triangles), therefore a quantitative comparison between the two reconstructions could be misleading. In order to give an idea of the accuracy of the reconstruction, a close inspection of the resulting surface mesh shows that even the borders of the round paper stickers of *figure 4a*, *b*, and *c*, applied to the surface of the stone for the photogrammetric survey, are accurately reconstructed. It is important to notice that the cameras employed for this 3D reconstruction are standard TV-resolution CCD cameras. Using higher resolution digital photo-camera the accuracy of the results would greatly improve.

4. Global reconstruction

The fusion of partial reconstructions into a global 3D model can be performed by estimating the rigid motion of the camera system between acquisitions, and by referring all 3D data to a common global frame. We perform this operation by looking for the rigid camera motion that best merges the 3D data that are in common between views.

The egomotion estimation method that we developed and implemented for accurate patchworking purposes is organized in two mains steps. After having partitioned the available 3D contours in lines and curves, we proceed as follows:

1. rough egomotion estimation from straight contours:
   - matching of straight contours
   - motion estimation through minimization of the distance between homologous contours
2. egomotion refinement using curved contours:
   - matching of curved contours
   - motion estimation through a minimization of the distance between homologous curved contours.

Notice that, as a first approximation of the egomotion is already available, the matching of curved contours is a rather simple operation compared with the matching of straight lines.

4.1. Egomotion from straight lines

Line matching in a 3D space is performed through a hypothesize-and-test type of procedure [18]. The first step of this method procedure consists of formulating hypotheses on the possible couplings by

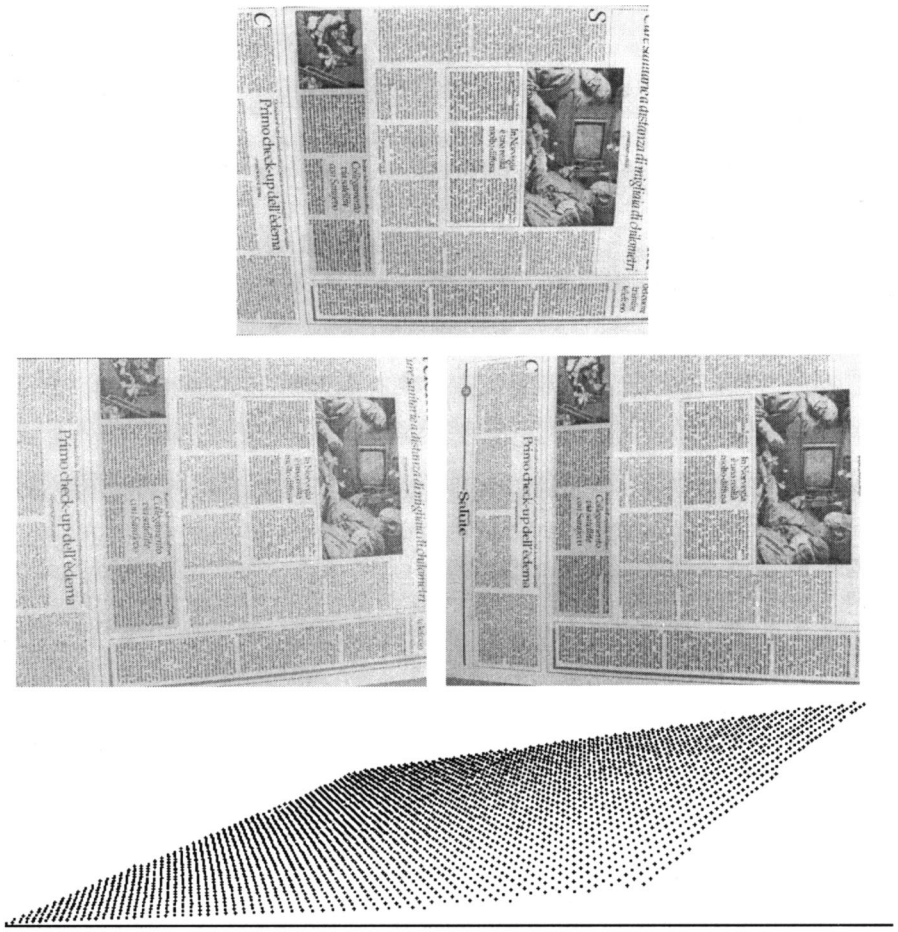

Figure 3. Original views of the newspaper's page (glued to a planar surface) and 3D points reconstructed through area matching.

selecting all those that do not violate some rules of congruence based on a set of geometrical constraints. By doing so, we drastically reduce the search space over which to test for matching correctness. At this point we can proceed with an exhaustive search through the above reduced set of hypotheses and select the match that maximizes an appropriate measurement of the matching quality.

Once the matching process is complete, the egomotion estimation can be performed rather easily by searching for the rigid motion that minimizes an appropriate *merging cost* function between two sets of 3D lines that pertain to two different partial reconstructions. Notice that 3D contours are generally reconstructed as chains of segments whose length and fragmentation may vary quite drastically from multi-view to multi-view. We thus proceed by first determining the 3D line portions that best fit (through linear regression) the chains of fragments of edges that have been recognized as straight. Then instead of measuring the distances between extremal points of two segments, we measure the distance between the extremal points of one segment and the line that the other segment lies upon (see *figure 5*). Such distances are used for defining the merging cost as follows:

$$C_s = \sum_{i=1}^{N} [(d_i^b)^2 + (d_i^e)^2] \quad (1)$$

In fact, the orientation of edges is usually less sensitive to fragmentation problems than their location in the 3D space [1, 18].

4.2. Egomotion refinement from curved contours

As already said above, curved contours are used to improve the accuracy of the egomotion's

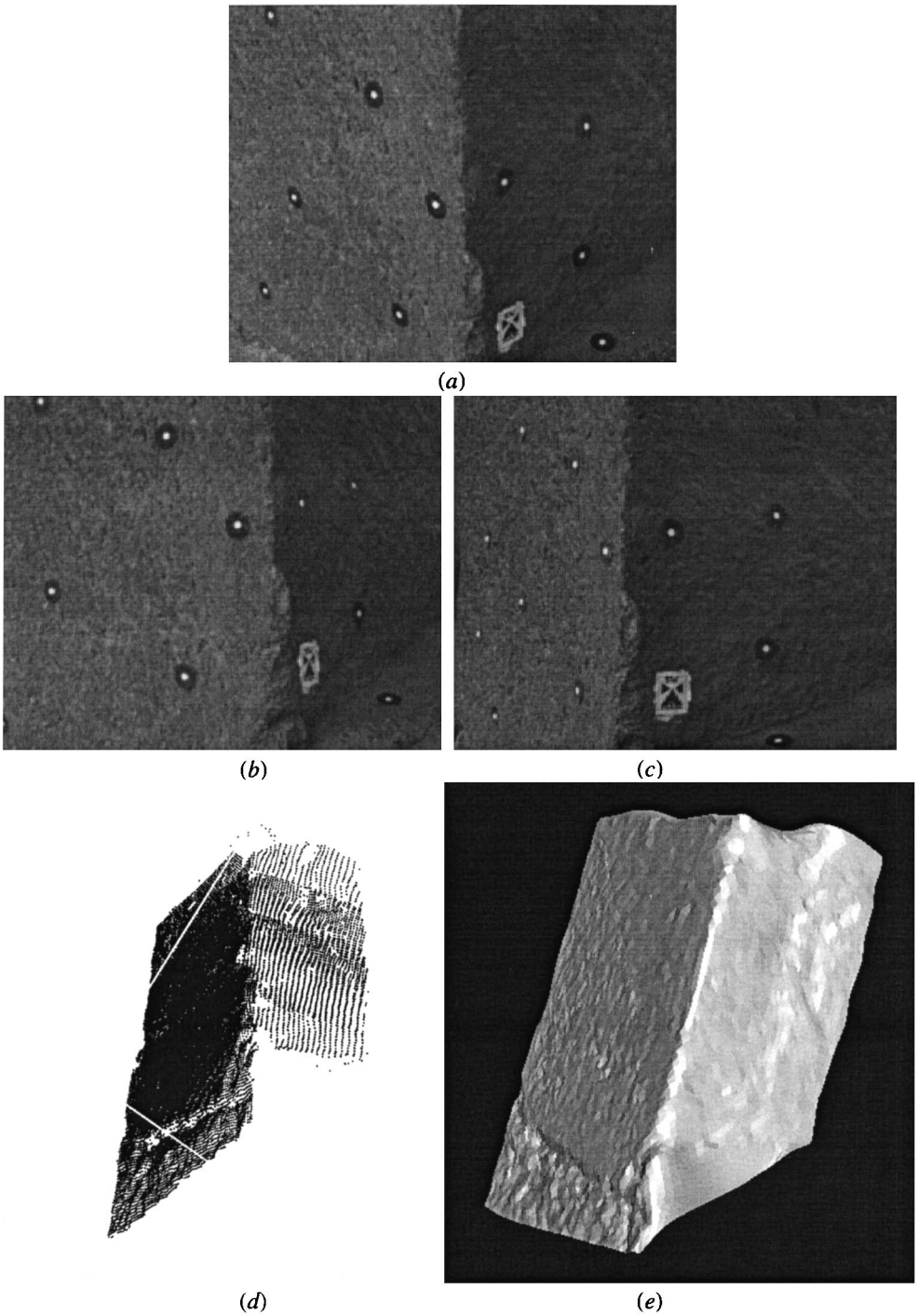

Figure 4. (a–c) One of the available triplets of views of a large stone of the Roman amphitheater of Aosta. The added targets (circular stickers) are used for photogrammetric comparison and for the estimation of the egomotion through 3D point matching. (d) Perspective view of the global merging of all 3D points reconstructed through 3D area matching. (e) Perspective view of the reconstructed global surface.

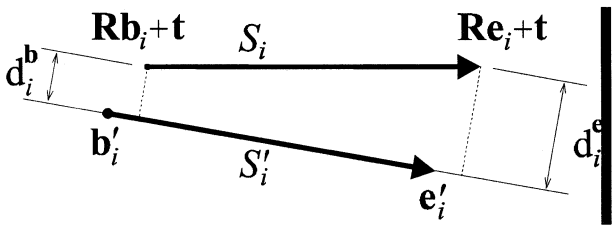

Figure 5. Evaluation of the merging cost of two straight 3D contours.

estimate. Although a matching process is required in this case too, this step is now simplified by the knowledge of a first approximation of the camera motion, determined from straight edges. In fact, by applying the pre-determined rigid motion to the set of curved edges, we can decide whether two curved edges are matched, depending on their global distance, which can be measured, with reference to *figure 6*, as:

$$d_g = \frac{1}{2}[d(C,C') + d(C',C)] \tag{2}$$

where

$$d(C,C') = \frac{1}{N}\sum_i d(E_i,C') = \frac{1}{N}\sum_i \|\overline{E_i E'_i}\| \tag{3}$$

The global cost function for motion refinement is of the form $C = C_s + kC_c$, where C_s and C_c are the merging costs associated to straight and curved contours, respectively, and k is weight for balancing the two contributes.

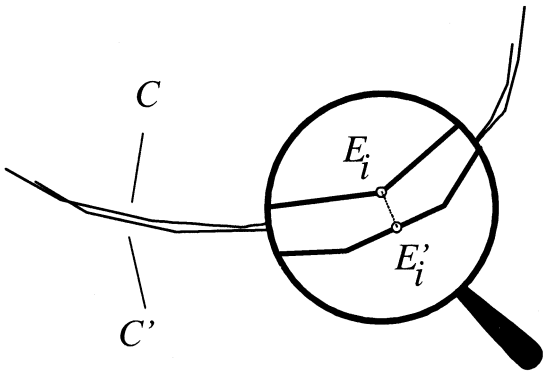

Figure 6. 3D curve matching: evaluation of the distance between two polylines.

4.3. Examples of application

The method has been extensively tested against convergence problems and has been applied to a series of trinocular acquisitions of real images in order to evaluate qualitatively and quantitatively the accuracy of the results and the speed of convergence. Furthermore, the performance of the proposed method has been compared with that of a previously studied method [2, 8] based on point correspondences between artificially added markers. Quantitative results have been obtained by measuring the maximum thickness of the bundles of edges when superimposing different sets of them with the estimated motion parameters. The performance of the proposed method has been proven to be equal to or better than that of the point-based approach, resulting in a maximum bundle size of about 100 ppm in all tests (after merging all 3D edges coming from 20 multi-views).

In *figure 7*, the results on 3D data merging are reported for an object of complex shape, in both cases of egomotion estimated through point and line correspondences. In the first case the cost function is a rigidity constraint based on the distance between reconstructed 3D points of different 3D data sets. Such points are markers that have been artificially added to the scene (white dots placed on the object's support). In the second case the egomotion is computed with the method proposed in this paper. Even though no artificially added markers have been used for the estimation, the accuracy of the estimate is comparable with that obtained through point-matching.

5. Conclusions

In this paper we presented our global approach to accurate 3D reconstruction with a calibrated multi-camera system. In particular, we presented a simple and effective technique for calibrating CCD-based multi-camera acquisition systems. The proposed method was proven to be capable of highly-accurate results even when using very simple calibration target-sets (with little or no *a-prioria* information on it) and low-cost imaging devices, such as standard TV-resolution cameras connected to commercial frame-grabbers. We also showed our approach to adaptive calibration, which proved effective for keeping track of camera parameter drift through natural feature tracking.

We also proposed and illustrated a general and robust approach to the problem of close-range par-

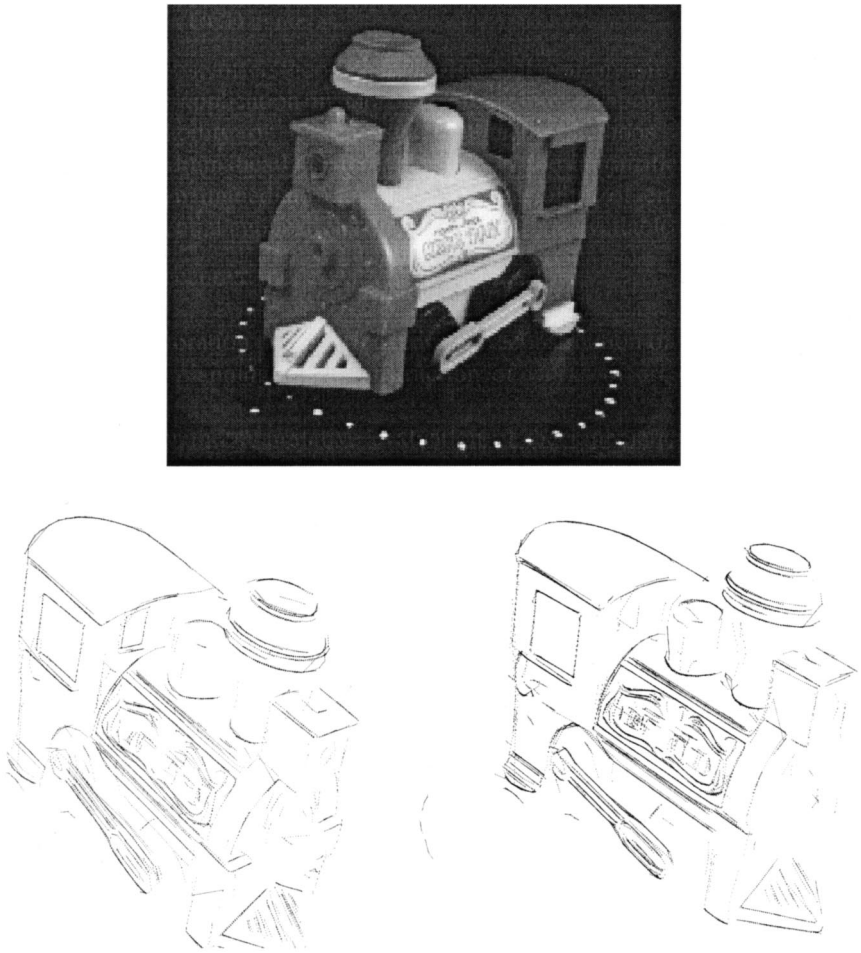

Figure 7. One of the original views of the object, fusion of all 3D edge sets through 3D point correspondences (marks added on the object's support), fusion of all 3D edge sets through 3D contour matching (natural edge features).

tial 3D reconstruction of objects from stereo-correspondences. The method is independent of the geometry of the acquisition system which can be a set of *n* cameras with strongly converging optical axes. The robustness of the approach can be mainly attributed to the physicality of the matching process, which is virtually performed in the 3D space. In fact, both 3D location and local orientation of the surface patches are estimated, so that the geometric distortion can be accounted for. The method takes into account the viewer-dependent radiometric distortion as well.

Finally, we presented a method for performing an accurate patchworking of the partial reconstructions, through 3D feature matching. The method, based on the best fusion of 3D curves, provides very accurate results even when using standard TV-resolution CCD cameras.

The global approach that we propose offers good characteristics of non-invasiveness, flexibility and accuracy that make it suitable for a variety of applications in the field of the preservation and restoration of the Cultural Heritage. In general, the availability of non-invasive automatic methods for the 3D reconstruction of 'unstructured' objects may dramatically reduce the costs of the automatic surveying of monuments and works of art. This will have an immediate impact on the feasibility of effective protection plans for the cultural heritage. Some of the potential applications that would become available at low costs are:

- the creation of a database of accurate models of 3D works of art
- restoration planning
- fast documentation of the restoration process (in-

expensive reconstruction of portions of large objects)
- accurate 3D registration for planning the reconstruction of works of art in case of accidental damage, building copies, etc.

In addition, effective methods for the quantitative evaluation of the kinematics of the environmental impact (erosion, damage, etc.) or the prediction of fractures or structural failures would help us prevent the damage from taking place at all.

References

[1] Ayache N., Artificial vision for Mobile Robots, MIT Press, 1991.

[2] Pedersini F., Sarti A., Tubaro S., A Multi-view Trinocular System for Automatic 3D Object Modelling and Rendering, XVIII International Congress for Photogrammetry and Remote Sensing, 1996, Vienna, Austria.

[3] Otha Y., Kanade T., Stereo by intra- and inter-scanline search using dynamic programming, IEEE Trans. on PAMI 7 (2) (1985) pp. 139–154.

[4] Pigazzini P., Pedersini F., Sarti A., Tubaro S., 3D area matching with arbitrary multiview geometry, EURASIP Signal Processing: Image Communications, Special issue on 3D *video technology*, early issues of 1998.

[5] Pedersini F., Sarti A., Tubaro S., Robust area matching, IEEE International Conference on Image Processing, 26–29 October, 1997, Santa Barbara, CA, USA.

[6] Pedersini F., Sarti A., Tubaro S., Egomotion estimation of a multicamera system through line correspondence, IEEE International Conference on Image Processing, 26–29 October, 1997, Santa Barbara, CA, USA.

[7] Pedersini F., Sarti A., Tubaro S., Automatic surface reconstruction of 3D works of art, International Conference on Electronic Imaging and the Visual Arts (EVA '97), March 19–25, 1997, Florence, Italy.

[8] Pedersini F., Sarti A., Tubaro S., 3D motion estimation of a trinocular system for a full-3D object reconstruction, IEEE Int. Conf. on Image Processing, September 1996, Lausanne, Switzerland.

[9] Tsai R.Y., A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, IEEE J. on Robotics and Automation, RA-3, (4) (1987) pp. 323–344.

[10] Weng J., Cohen P., Herniou M., Camera calibration with distortion model and accuracy evaluation, IEEE Trans. on PAMI, Oct 14 (10) (1992) 965–980.

[11] Pedersini F., Pele D., Sarti A., Tubaro S., Calibration and self-calibration of multimulti-ocular camera systems, Int. Workshop on Synthetic-Natural Hybrid Coding and Three-Dimensional (3D) Imaging (IWSNHC3DI '97), September 5–9, 1997, Rhodes, Greece.

[12] Pedersini F., Tubaro S., Accurate 3D reconstruction from trinocular views through integration of improved edge-matching and area-matching techniques, VIII European Signal Processing Conference, September 10–13, 1996, Trieste, Italy.

[13] Kitchen L., Rosenfeld A., Gray-level corner detection, Pattern Recognition Lett. 1 (1982) 95–102.

[14] Giraudon G., Deriche R., On corner and vertex detection, Proc Int. Conf. on Computer Vision and Pattern Recognition, June 1991, Maui, Hawaii, pp. 650–655.

[15] Rohr K., Recognizing corners by fitting parametric models, Intl. J. of Computer Vision 9 (3) (1992) 213–230.

[16] Pedersini F., Sarti A., Tubaro S., Tracking camera calibration in multi-camera sequences through automatic feature detection and matching, IX European Signal Processing Conference, September 8–11, 1998, Rhodes, Greece.

[17] Mallet J.L., Discrete smooth interpolation, ACM Transactions on Graphics 8 (2) (1989) 121–144.

[18] Zhang Z., Faugeras O.D., 3D Dynamic Scene Analysis: A Stereo Based Approach, Springer-Verlag, 1992.