Proceedings of the IMProVe 2011
International conference on Innovative Methods in Product Design
June 15th – 17th, 2011, Venice, Italy

Automatic alignment of multi-view range maps by optical stereo-tracking

S. Barone [(a)], A. Paoli [(a)], A. V. Razionale [(a)]

[(a)] University of Pisa, Department of Mechanical, Nuclear and Production Engineering

Article Information

Keywords:
Reverse engineering,
Stereo vision,
Optical tracking.

Corresponding author:
Alessandro Paoli
Tel.: +39 050 2218174
Fax.: +39 050 2218065
e-mail: a.paoli@ing.unipi.it
Address: Largo Lucio Lazzarino 1, - 56100 Pisa - Italy

Abstract

Purpose:
Industrial reverse engineering applications usually require reliable multi-view measurements of complex shapes obtained by full field techniques. One of the most challenging issues in this field is represented by the automatic computation of the best transformation parameters relating the different views into a common reference system.

Method:
A low cost optical tracking system has been developed with the aim at creating an automatic procedure to align 3D point clouds captured by a structured light system. The tracking system uses stereo images and retro-reflective infrared markers rigidly connected to the scanner. Markers are accurately tracked on the basis of automatic intensity-based analyses. Stereo correspondences are established by using epipolar and similarity constraints.

Result:
Experimental tests have been carried out in order to evaluate the capabilities of the developed system in the measurement process of nominal benchmarks and complex shapes.

Discussion & Conclusion:
The accuracy obtained in the automatic alignment of multiple views justifies the use of the proposed approach in the speed up of the measurement process for reverse engineering applications, especially when conventional high-end systems cannot be applied.

1 Introduction

Industrial reverse engineering applications require fast, robust and reliable 3D full measurements of complex shapes. However, most of the acquisition systems are limited to the reconstruction of partial views due to the presence of occlusions and limited working volumes. Many different views must then be collected by moving the acquisition device and/or the target object and aligned with reference to a common coordinate system. The major problem in combining multiple views is the automatic computation of the best transformation parameters (translation and rotation), which relate information in one view with information in other views [1]. The definition of the transformation parameters is usually referred as registration and constitutes the core of the alignment process. This task represents the main bottleneck in the model acquisition pipeline due to the amount of user interaction required.

In this paper, a low-cost optical tracking system has been developed with the aim at creating a fully automatic alignment procedure of 3D point clouds captured by a *full field* technique. Point clouds are acquired by an active stereo vision system based on structured lighting and encoded pattern analysis. The optical tracking system uses binocular images and retro-reflective infrared (IR) markers. Passive markers, made by spheres, are rigidly connected to the scanner and are accurately tracked in real time by the stereo rig on the basis of automatic 2D intensity-based analyses. Stereo correspondences are established by using both epipolar and similarity constraints. 3D marker frames are then created for each pose of the scanner and used to align the point clouds through a dedicated calibration procedure.

Experimental tests have been carried out in order to evaluate the performances of the spheres detection algorithm as well as the stereo correspondence matching. Nominal benchmarks and complex shapes have then been used to assess the feasibility of the developed multi-view registration approach. The proposed methodology demonstrated to speed up the measurement process, making the automatic alignment of single point clouds more precise and robust.

2 Motivation and background

In recent years, technical literature has proposed various approaches, based on fully unmanned methodologies, to carry out the registration process. The most used methodologies can be classified in two main approaches: *shape-based* and *tracking-based* techniques.

Shape-based techniques rely on the automatic detection of features existing on adjacent overlapping views. These features, invariant under rigid transformations, denote geometric and topological properties of 3D shapes and are extracted from each single view. Similar features are then used to compute correspondences between different views in order to match the raw data under some rigidity constraint. A survey of feature-based surface approached can be found in [2, 3]. However, no single algorithm has prevailed in the registration of arbitrary range scans since each approach assumes particular restrictions on the typologies of input data. The main drawback of these methods is due to the requirement of surface features in the overlapped areas.

Flat or uniform curvature geometries are not practically appropriate for shape-based alignments. Moreover, large percentage of computation time is usually required in pre-processing activity, which includes extraction of invariant features and organisation of the extracted primitives.

Tracking-based techniques are rather based on independent devices that track relative motions between target object and acquisition system. *Mechanical tracking* systems (such as turntables, robotic arms or coordinate measurement machines) can produce a good estimate of the roto-translation between consecutive views [4, 5]. These approaches enable fast scans of target objects, even with complex shapes. However, the accuracy of alignments depends on the resolution of tracking device. High precision translation/rotation controlled axes are not always affordable in economical terms and require an additional calibration with reference to the 3D vision system. Moreover, tracking systems can only support limited working volumes.

A less expensive solution certainly involves *non-mechanical tracking* approaches. Magnetic, acoustic, inertial or optical sensing can be used to track the scanning device over time [6]. However, among *non-mechanical tracking* approaches, magnetic and acoustic techniques are not suitable for high accuracy applications in non-controlled environments while inertial approaches are characterized by low tracking autonomy due to drift. Magnetic trackers are based on the transmission and reception of electromagnetic signals and rely on measurements of the local magnetic field vector at the sensor. Even if they are not subjected to line of sight obstructions, ferromagnetic and conductive material in the environment can affect the magnetic field's shape. Acoustic systems use the transmission and sensing of sound waves. The location of the tracked object is usually determined by computing the time-of-flight of a brief ultrasonic pulse. Unlike magnetic waves, acoustic signals require a line of sight between emitters and receivers and their accuracy can be affected by uncertainties in the speed of sound, which significantly depends on temperature, humidity, and air currents. Moreover, walls and objects surrounding the system can be acoustically reflective drastically varying the prediction of the receiver's location. No line of sight requirements are needed for inertial sensing which is based on the use of gyroscopes and accelerometers to measure linear and angular accelerations relative to a fixed reference system. Absolute position and orientation are then computed by integration. Inertial tracking do not suffer from interfering electromagnetic fields or ambient noise. The real weakness that prevents inertial trackers from being the standard for accurate motion tracking is the drift phenomenon. Even if one of the accelerometers has a low bias error, the reported position output can greatly diverge from the true position after a few seconds. For this reason, inertial sensing is characterized by having a low tracking autonomy.

In this context, optical systems has proved to be a valuable alternative for motion tracking because of their low cost, reduced sensitivity to noise and greater flexibility. Optical systems are based on the use of light sources and optical sensors and rely on measurements of reflected or emitted light. Typically, they can be divided into vision based methods and marker based methods [7]. Vision based methods work by matching high contrast passive features (i.e. corners, edges, gray-level patches) using advanced image processing techniques. Feature matching algorithms can be computationally expensive as they extract features or statistical information directly from raster images. Marker based methods rather rely on features artificially added to the scene, considerably simplifying image processing. These methods can be based on passive markers, such as coloured fiducials or retro-reflective targets, or active devices emitting light as light-emitting diodes (LEDs), lasers or simple light bulbs. A further characterization of optical tracking systems is the distinction between *outside-looking-in* (*OLI*) and *inside-looking-out* (*ILO*) systems. When the sensors are fixed in the environment and the object is observed from outside, the system is referred to as *outside-looking-in*. *Inside-looking-out* systems are rather characterized by mounting cameras on the object and calculating its pose by tracking visual landmarks attached in the scene. The primary disadvantage of all optical systems relies on the requirement of a clear line-of-sight between sources and sensors throughout the tracking process. However, this drawback can be readily handled by redundancy of tracked elements and/or optical sensors.

In recent years, various optical tracking methodologies have been proposed as commercial systems [8-10], accommodating stable and constantly improving technologies. Nevertheless, they are very expensive and not free of drawbacks, thus limiting their application. For this reason, the research community has proposed a wide variety of applications based on custom approaches. In Virtual Reality (VR) and Augmented Reality (AR) systems, optical tracking can be used to trace objects in real time [7, 11, 12], or to capture full human body motion [13] for computer animation as well as gait analysis, sports performance evaluation or patient rehabilitation [14]. Moreover, in the surgical field, optical sensing can be used in the development of navigation systems for assisted surgery [15, 16]. Although, optical tracking has been used in several contexts, its application to align 3D point clouds generated by *full field* techniques has not been fully evidenced in technical literature, but only some industrial solutions have commercially been proposed [17].

In this paper, a marker-based optical tracking system has been developed with the aim at creating a low-cost automatic alignment procedure for 3D point clouds captured by a *full-field* technique. The system has been designed to be used in environments where conventional high-end tracking systems cannot be applied. Point clouds are acquired by an optical scanner, based on structured light projection and encoded pattern analysis, and are aligned by tracking retro-reflective markers illuminated by infrared (IR) sources. To achieve reliable results for an industrial context, the markers placement with respect to the optical scanner is accurately calibrated by a semi-automatic image-based methodology which requires a minimal human interaction. Even if the set-up arrangement of the developed system is more time consuming with respect to the more conventional approach, based on the interactive selection of fiducial markers arranged on the measured surfaces, several advantages have been accomplished. Firstly, the automatic registration of any range image is obtained without requiring the presence of overlapping areas between consecutive scans. Thus, a completely free scanning pattern can be followed by the operator during the measurement process without requiring costly and cumbersome mechanical tracking systems. Moreover, the use of an external reference (i.e., the tracking system) ensures that registration errors are not accumulated during the scanning process.

3 Optical tracking system

The proposed methodology relies on the combination of a 3D optical scanner, based on the projection of encoded light stripes, and an automatic multiple-view approach by an optical passive tracking technique.

3.1 Single-view approach

The optical scanner is based on a structured lighting approach, which uses binary patterns in order to capture three-dimensional shapes. The system (Fig. 2) is composed of a DLP projector (1024 × 768 pixels) to generate vertical and horizontal black and white striped light patterns and two monochrome digital CCD cameras (1280 × 960 pixels).

The camera's configuration is calibrated by evaluating the intrinsic (focal distances, co-ordinates of the principal points, radial and tangential distortions) and the extrinsic parameters [18]. The stripe projector is un-calibrated and not directly involved in measurement processes.

In this work, a double binary encoded light stripe approach is used for 3D shape recovery [19]. In particular, a line correlation process is adopted by projecting a sequence of lines defined as crossing zones between white and black parallel stripes with periods progressively halved. Each pixel in the camera images is characterised by a light intensity that can be either bright or dark depending on its location in the respective plane image. A binary code (0,1 with n bit) is assigned to each pixel, where n is the number of stripe patterns, and the values 0 and 1 are associated to the intensity levels, *i.e.*, 0 = *black* and 1 = *white*. This encoding procedure provide $\ell = 2^n - 1$ encoded lines by using n patterns. The stereo correspondence is solved by projecting both horizontally and vertically striped encoded patterns. A double code is assigned to the intersection points through the horizontal and vertical stripes. This yields an automatic and unique correspondence of conjugate points in the camera images. The resolution that can be obtained by the proposed procedure is $\ell_v \times \ell_h$ points, where ℓ_v and ℓ_h are the number of vertical and horizontal lines switched by the projector, respectively. The maximum resolution of the depth measurement is obtained by shifting one of the stripe images four times by the width of one pixel. Taking into account the resolution of the projector integrated to the stereo system, this method provides 784.641 encoded points (ℓ_v = 1023, ℓ_h = 767) by projecting 8 stripe patterns with periods progressively halved and three stripe patterns by shifting the finest stripes (the 8th stripe pattern). This approach allows the resolution of the projector to be exploited avoiding too finer stripes, which may typically produce poor accuracy in line extraction.

The 3D coordinates of the scene points are then recovered by assuming the known image acquisition geometry and using the midpoint triangulation technique for conjugate points [20].

3.2 Automatic multi-view approach

The complete shape recovery of a complex shape generally requires the acquisition of multiple range maps from different views, which have to be transformed into a common coordinate system. The alignment process is carried out by referring each range map to a frame of fiducial markers rigidly connected to the scanner structure and tracked by a stereo vision based approach.

The method is composed of the following steps:

1) spheres detection from grey level images captured by a stereo-rig,
2) stereo correspondences determination by geometrical and similarity constraints,
3) optical scanner pose estimation through a calibration process.

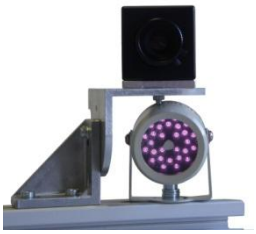
The procedure is automatically and progressively applied to each acquisition, so that the operator is not constrained to follow a predetermined scanning strategy.

3.2.1 The optical tracking system

The proposed tracking methodology is based on the *OLI* approach and uses a calibrated stereo set-up cameras equipped with IR band pass filters. The tracked markers are retro-reflective spheres, rigidly connected to the optical scanner, illuminated by IR light sources. Markers detection on the camera image planes is performed by intensity-based analyses while stereo correspondences are established by exploiting both epipolar and similarity constraints. The 3D marker frames are then created by triangulation and used to align the point clouds through a dedicated marker calibration procedure.

The hardware set-up consists of a stereo rig (Fig. 1-a) assembled using two digital cameras and an adjustable baseline which is fixed with respect to the object to be measured. IR light sources (940 nm wavelength) are mounted below the cameras (Fig. 1-b). The optical configuration is calibrated by evaluating the intrinsic and extrinsic parameters [18]. The two CMOS digital cameras have a resolution of 2048 × 1536 pixels and are sensitive to near infra-red (*NIR*) light. 12 mm focal length lenses, equipped with infrared filters having a cutting off wavelength at about 825 nm, have been used.

a)



b)

Fig. 1 a) Optical tracking set-up, b) IR light source.

Optical passive tracking systems using natural features are usually characterized by high computational costs and lack of robustness. For this reason, retro-reflective spheres (diameter 13 mm) have been used as artificial markers in order to speed up the tracking process. These markers are rigidly connected to the optical scanner (Fig. 2) and are disposed at different relative distances between each pair of them. The adopted distribution, which can be easily varied, is motivated by occlusion problems and simplifies markers identification by the stereo vision arrangement. The use of retro-reflecting spheres and IR lighting, combined with short camera exposure times, solves any illumination problem thus simplifying image processing.

Since the scene is illuminated by IR sources located below the digital cameras, retro-reflective spheres reflect the IR light back into the cameras. By mounting IR filters in front of the lenses, the resulting images display white circles over a dark background, regardless of the viewing direction (Fig. 3-a). In this work, the coarse circle detection is performed by binarising the original image using a threshold value τ (Fig. 3-b). The probability distribution of the grey intensity levels of the original images presents a substantial positive skeweness due to the preponderance of dark over bright pixels. For this reason, the threshold value τ is obtained as the grey intensity value corresponding to 99% of the maximum value of the cumulative distribution function.

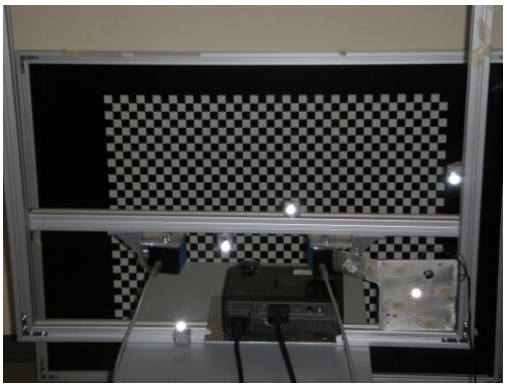


Fig. 2 Optical scanner with retro-reflective markers illuminated by IR sources.

All the connected components are then labelled (Fig. 3-c) and a set of two attributes, area (*A*) and extent (*ext*, specifies the proportion of the pixels in the bounding box that are also within the connected region), are calculated in order to discard possible false positives caused by uneven reflections. Maximum and minimum threshold values for area have been defined on the basis of images resolution while the adopted range for the circle extent, which is theoretically expressed by π, takes into consideration the pixel discretisation and the border effect. 2D coordinates of the circles are preliminarily evaluated as intensity centroids of the connected regions whose properties verify, at the same time, the two constraints above described. A refinement process is then applied over a window centred on each preliminary centroid, thus allowing a sub-pixel accuracy.

In order to calculate the 3D coordinates of the spheres centers, the matching between conjugate points on the two camera image planes must be solved. In [21] a method to solve the stereo correspondence problem for un-calibrated cameras is proposed using the Scott and Longuette-Higgins algorithm [22]. Let *I* and *J* be two images (right and left) containing *m* features I_r ($r = 1 \ldots m$) and *n* features I_l ($l = 1 \ldots n$), respectively, which have to be correlated. The method is based on the definition of a proximity matrix *G* where each element, G_{rl}, represents the Gaussian weighted distance between two features I_r and I_l:

$$G_{rl} = e^{-r_{rl}^2 / 2 \cdot \sigma^2} \qquad (1)$$

where $r_{rl} = \| I_r - I_l \|$ is their Euclidean distance if they were considered on the same image plane.

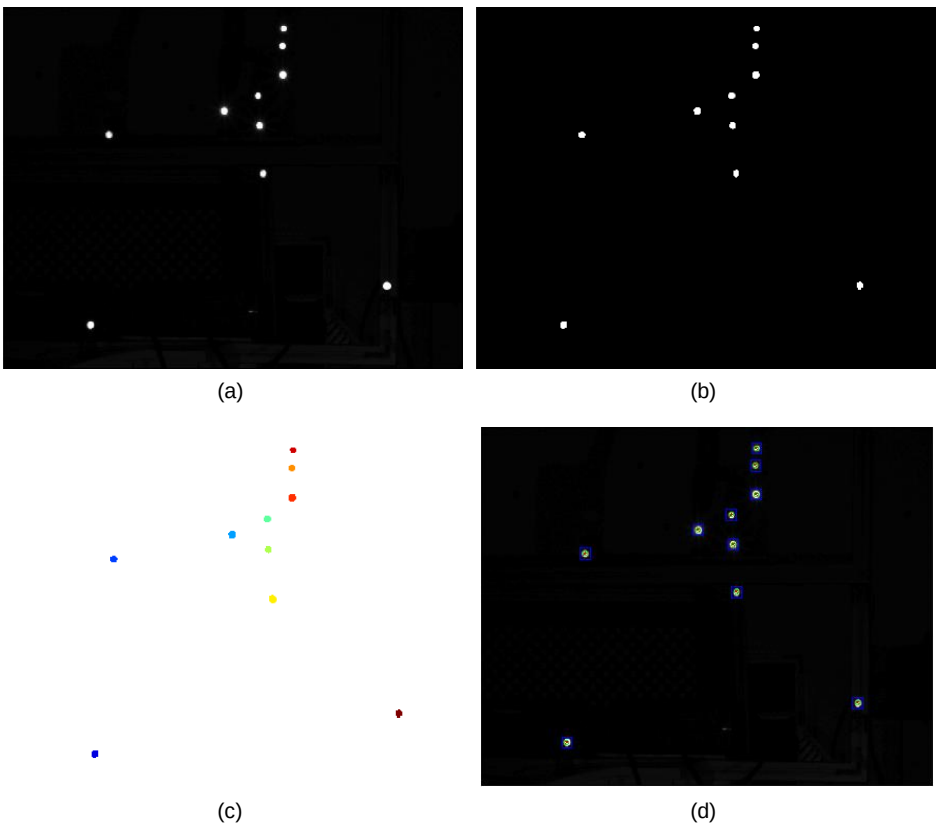


Fig. 3 Spheres detection process: a) original image, b) binarised image, c) labeled image, d) centroids extraction.

The parameter σ controls the degree of interaction between the two sets of features: a small value of σ enforces local interactions, while greater values allow more global interactions. A Singular Value Decomposition (SVD), $G = TDU^T$, is performed and the diagonal matrix D, containing the singular values along its diagonal elements, is converted to a new matrix E obtained by replacing singular values with 1. The product $P = TEU^T$ is then computed. This new matrix P reveals a direct one-to-one mapping between features I_r and I_l when P_{rl} is the greatest element both in its row and in its column.
In this paper, a robust stereo correspondence of markers is proposed by combining *epipolar* and *similarity* constraints.

Once the stereo set-up has been calibrated, the *epipolar* constraint can be easily deduced by:

$$m_r^T F m_l = 0 \qquad (2)$$

where m_r and m_l are the conjugated points on right and left images respectively while the matrix F, referred as the *Fundamental Matrix*, is the algebraic representation of the *epipolar* geometry [23]. Given a point on the right image, m_r, its conjugate m_l on the left image belongs to a line, the *epipolar line*, defined as intersection of the image plane with the plane defined by m_R and the focuses of the two optical devices, *i.e.* the *epipolar plane*. Each candidate point m_R in the right image can only correspond to such point in the left image that lies on the relative *epipolar line*. An algorithm of stereo point-pairs rectification [24] is used to reduce the correspondence search to one dimension (epipolar lines become parallel and conjugate points assume the same vertical coordinate, y). The *epipolar* geometry can then be considered in the proximity matrix G by:

$$G_{rl}^{ep} = e^{-|y_r - y_l|^2 / 2\cdot\sigma_{ep}^2} \qquad (3)$$

where (x_r, y_r) and (x_l, y_l) are the rectified horizontal (x) and vertical (y) image coordinates of points I_r and I_l and σ_{ep} is a tuning parameter which reflect the expected error in the epipolar constraint (and therefore in the calibration of the optical stereo-rig). The use of the only epipolar constraint could cause improper matching when multiple different points are close to the same epipolar line. For this reason, a similarity constraint has been also introduced with the aim at enforcing the epipolar geometry constraint, similarly to [7]. A correlation between correspondent features can be established by analyzing their neighboring points, which should be distributed similarly (*similarity constraint*). This constraint is introduced into the proximity matrix G by:

$$G_{rl}^{sim} = e^{-d_{sim}(r,l)^2 / 2\cdot\sigma_{sim}^2} \qquad (4)$$

where $d_{sim}(r,l)$ gives an estimate of the distribution of the rectified points within a circular region R around the considered features I_r and I_l (Fig. 4):

$$d_{sim}(r,l) = \sqrt{d_x(r,l)^2 + d_y(r,l)^2} \qquad (5)$$

with:

$$d_x(r,l) = \frac{1}{m}\sum_{i=1}^{m}(x_i - x_r) - \frac{1}{n}\sum_{j=1}^{n}(x_j - x_l) \qquad (6)$$

$$d_y(r,l) = \frac{1}{m}\sum_{i=1}^{m}(y_i - y_r) - \frac{1}{n}\sum_{j=1}^{n}(y_j - y_l) \qquad (7)$$

σ_{sim} is a parameter reflecting the expected similarity error. The total proximity matrix G is then given by:

$$G_{rl} = G_{rl}^{ep} \cdot G_{rl}^{sim} \qquad (8)$$

The 3D coordinates of markers are then determined by a triangulation procedure given the conjugated 2D point pairs. The combination of the geometry constraints allows the matching problem to always have a unique solution.

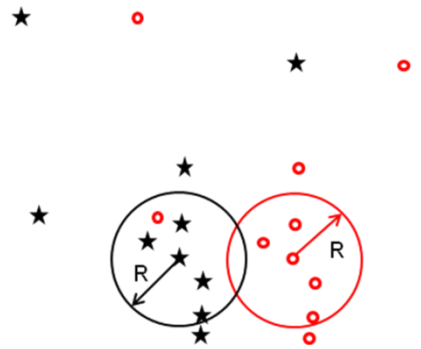


Fig. 4 Similarity constraint.

3.2.2 The tracking calibration process

In this work, the calibration phase can be described as a procedure having two different goals. The first one is the typical intent of z tracking calibration procedure and consists in the determination of the 3D structure of the markers connected to the scanner. In this phase, a calibrated 3D model graph, ***Q***, is created. Each graph vertex represents a 3D marker while each edge represents the distance between two markers. An edge exists if the two markers can be simultaneously observed by the tracking system. The redundancy of the number of spheres has been designed in order to guarantee that, even in cases of partial occlusions, at least three of them where always visible by the stereo set-up arrangement. However, a further goal must be accomplished and relies on relating the markers model graph with respect to the optical scanner. The matching of correspondent 3D markers, relative to distinct scanner poses, must allow the automatic alignment of range maps captured by the optical scanner. This is possible only if each single range map can be referred to the calibrated model graph.

The calibration process requires the use of points having known positions in the world reference system (O_w, X_w, Y_w, Z_w). In this work, a calibration board, composed of a printed pattern attached onto a planar glass surface, has been used. The calibration procedure is performed by placing the board in a position where the targets are visible by both the optical scanner and the stereo tracking system (Fig. 5). Since the two optical configurations are calibrated, triangulation processes allow 3D reconstructions of the chessboard corners in the scanner reference system (O_s, X_s, Y_s, Z_s) and in the tracking reference system (O_t, X_t, Y_t, Z_t). The rigid body transformation (R_{ws}, T_{ws}) between the world reference system, integral to the chessboard, and the scanner reference system can then be determined by applying the SVD method [25] between the respective 3D point structures. Rotation and translation parameters (R_{wt}, T_{wt}) between the world reference system and the tracking reference system can be similarly obtained. By composing the two rigid transformations it is possible to determine rotation and translation parameters (R_{st}, T_{st}) between the scanner and the tracking system. The 3D marker coordinates, measured by the stereo tracking system, can then be expressed in the scanner reference system by applying R_{st} and T_{st}.

The procedure is repeated for different placements of the scanner, in front of the calibration board, in order to handle possible marker occlusions. The mean of the detected markers coordinates, transformed in the scanner

reference system, is then calculated to obtain the final calibrated 3D model graph ***Q***.

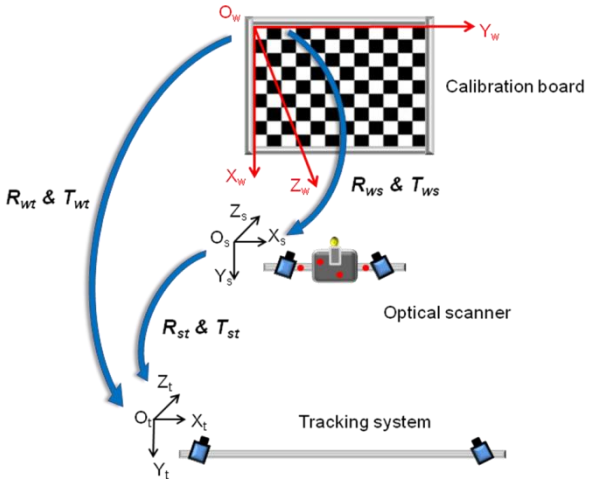


Fig. 5 Calibration process scheme.

3.2.3 The optical scanner pose estimation

The last step of the developed tracking approach consists of recognizing and matching correspondent 3D markers, detected for different range map acquisitions, with the aim at estimating the scanner pose. For each pose, markers are identified by computing the Euclidean distances between marker pairs and comparing these distances with those determined in the calibration process. The model graph ***Q*** encodes the topological relationship between all the spheres integral to the scanner. The pose estimation can then be efficiently reduced to a triangle-based matching process (Fig. 6) associating similar triangles between two data sets: the calibrated 3D model graph ***Q*** and the captured set ***P***, constituted by M_q and N_p markers respectively, with q, $p \geq 3$, can be searched.

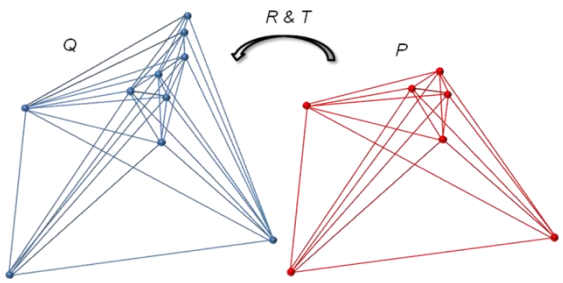


Fig. 6 Graph matching process.

In particular, the algorithm can be structured in the following steps.

1) The $p \times p$ symmetric adjacency matrix $\boldsymbol{D}^p$ is associated to the set $\boldsymbol{P} = \{N_1 \ldots N_p\}$; the elements of the matrix are composed by the Euclidean distances $d^p_{i,j} = \left\| N_i - N_j \right\|$.
2) Analogously, the $q \times q$ symmetric adjacency matrix $\boldsymbol{D}^q$ is associated to the set $\boldsymbol{Q} = \{M_1, \ldots M_q\}$, the elements of the matrix are composed by the Euclidean distances $d^q_{k,l} = \left\| M_k - M_l \right\|$.
3) Each column of the matrix $\boldsymbol{D}^p$ is compared to all the columns of the matrix $\boldsymbol{D}^q$; in particular, the differences of the distance values are computed and compared to a fixed tolerance ε, as:

$$\left\| d^p_{i,j} - d^q_{k,l} \right\| < \varepsilon \text{ with } d^p_{i,j} \neq 0,\ d^q_{k,l} \neq 0 \qquad (9)$$

If relation (9) is satisfied, the couple k,l is stored into an array of pointers representing a list of possible pairs of points (M_k, M_l) corresponding to (N_i, N_j). The search is iterated until relation (9) is satisfied at least twice for the same column: two corresponding pairs and, consequently, three common markers have been identified between the data sets ***Q*** and ***P***. The iterations are carried out varying the tolerance ε from a small value ($\varepsilon = 0$) to a fixed threshold (ε = 1 mm) by incremental steps of 0.05 mm.

The equivalence among all the remaining distances defined by matched marker pairs is then verified, allowing the final detection of similar triangles. Possible geometrical ambiguities, such as collinear configurations, are avoided checking all the angles between the matched markers. Given the set $\boldsymbol{C} = \{O_1 \ldots O_c\}$ of markers matched at the previous step, the following constraint for the dot product:

$$\min_{i=2..c} \left\{ abs\left[\left(\frac{\overline{O_i O_{i-1}}}{\left\| \overline{O_i O_{i-1}} \right\|} \right) \cdot \left(\frac{\overline{O_i O_{i+1}}}{\left\| \overline{O_i O_{i+1}} \right\|} \right) \right] \right\} < \frac{\sqrt{3}}{2} \qquad (10)$$

ensures that exists at least one angle between 30° and 150°. If none of the angles satisfy to this condition, the iteration continues until another pair of corresponding markers, if existent, has been detected.

Once the correspondent pairs of fiducial markers have been identified and geometrically verified, the translation vector **T** and rotation matrix **R**, defining the rigid body transformation between two consecutive scanner poses, are directly determined by applying the singular value decomposition (SVD) method [25].

4 Results

Preliminary tests have been carried out in order to assess the spheres detection algorithm and the stereo correspondence matching. Significant statistical values have been obtained by processing different images captured varying both the position of the scanner with respect to the tracking system and the distribution of the spheres with respect to the scanner. The high sphere's detection rate obtained (95% over 400 visible markers) is owing to the use of IR lighting and relative filters. The stereo correspondence matching is subjected to three different tuning parameters: the expected error in the epipolar geometry, σ_{ep}, the expected similarity error, σ_{sim}, and the radius R of the region used to search for similarity between points. While σ_{ep} can be easily estimated since it depends on the calibration of the optical devices, σ_{sim} and R are rather affected by the placement of the spheres in the workspace. Their optimal values are then estimated through experimental tests accordingly to the application.

The epipolar constraint can generate false matches when multiple points are close to the same epipolar line. This condition is hard to be simulated since it relies on the geometrical placement of the spheres with respect to the optical devices of the tracking system. For this reason, a methodology to rather assess the strength of the similarity constraint in the retrieval of false matches has been developed. For each pair of images, false matches deriving from the application of the epipolar constraint have been simulated by using pseudorandom scalar values, uniformly distributed on the unit interval, in the definition of the proximity matrix G^{ep} (Fig. 7-a). Correct stereo correspondences have been grouped so to fill the main diagonal.

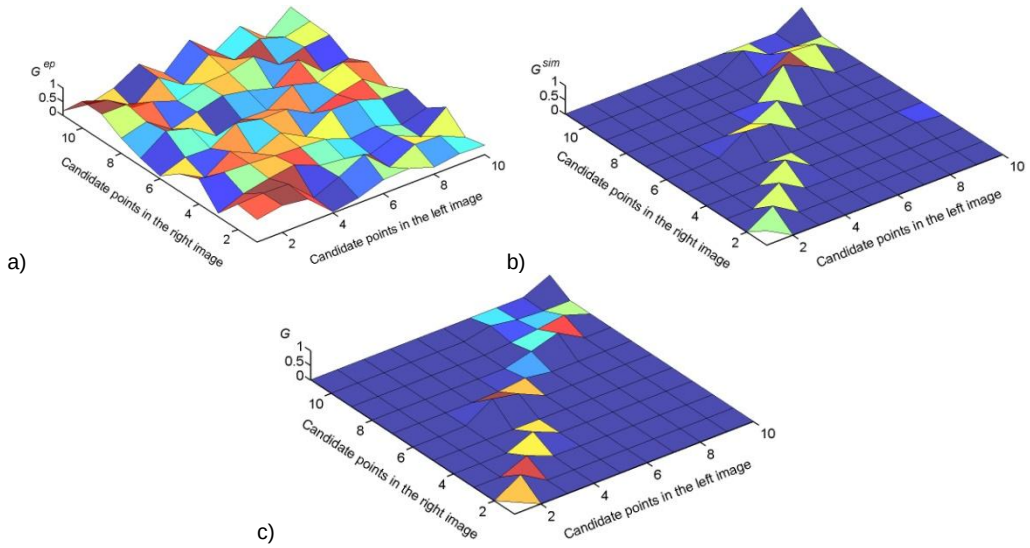


Fig. 7 a) Contribution of the random epipolar constraint, G^{ep}, and b) of the similarity constraint, G^{sim}, to the total proximity matrix G (c) defined by their element-wise product.

The contribution of the similarity constraint, G^{sim} (Fig. 7-b), to the total proximity matrix G (Fig. 7-c) allows the retrieval of some of the false matches randomly created. The percentage of this retrieval gives an estimate of the strength of the similarity constraint as conceived.

Fig. 8-a reports the influence of the parameter σ_{sim} on the percentage of stereo matches detected by using the only similarity constraint. Fig. 8-b outlines the capability of the similarity constraint (continuous red line) in the retrieval of false matches created by simulating random epipolar correspondences (dashed black line). The last plot evidences the influence of the radius R. In general, this parameter is closely related to the number of spheres that are used and, above all, to their disposition in the workspace. However, low values demonstrated to be not appropriate to identify similarities between points by analyzing their neighbouring regions.

The developed stereo matching approach demonstrated a significant robustness in the determination of the three-dimensional spheres coordinates with respect to the tracking reference system. However, the accuracy of the final point cloud alignment process also depends on the tracking calibration process. For this reason, some experimental tests, based on nominal samples and complex shapes, have been carried out to assess the accuracy of the final alignment phase. The optical scanner has been configured for a working distance of 1000 mm with a measuring area of 300 × 200 mm and accuracy, for a single acquisition, of 0.03 mm. A planar surface, represented by the calibration board shown in Fig. 2, has been measured by aligning different point clouds obtained moving the optical scanner through translations and rotations. The surface, measuring 1000 × 800 mm, needs at least twelve different scans to be fully acquired. Three different tests have been carried out varying the number (6÷10) and the disposition of the spheres. For each test, a dedicated tracking calibration process has been performed in order to create the three-dimensional model graph to be used in the alignment phase. Fig. 9 shows an example of the point cloud alignment as obtained for the planar surface by the proposed tracking methodology. Tab.1 reports the mean, μ_{plane}, and the standard deviation, σ_{plane}, of the residual errors of the plane best fitting the points acquired in three different measurements of the planar surface.

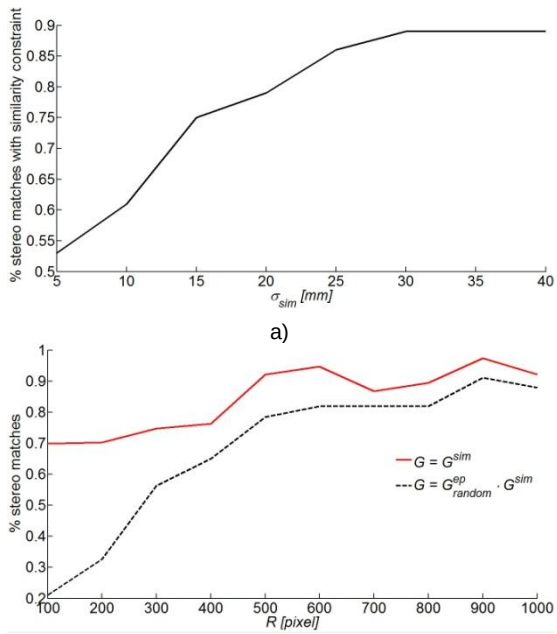


Fig. 8 a) Stereo matches rate by using the similarity constraint with different σ_{sim} values (a) and R values (b). The latter case also reports the influence of simulated false epipolar correspondences.

When the acquisition of real free-form complex shapes is involved, the quality assessment of multi-view approaches represents a difficult task since ground truth data are not always available. Fig. 10 shows the results obtained in the acquisition of a motorcycle component (Fig. 10-a) by aligning eight different scans using the proposed tracking methodology (Fig. 10-b). Even if a metric to evaluate the relative placement between overlapping areas should be used, a simple visual inspection can be sufficient to state the overall accuracy in the point cloud alignment. In general, the results obtained with the developed multi-view approach, also in the worst cases, have demonstrated to be accurate enough to represent a good starting point for global optimisation techniques based on the Iterative Closest Point (ICP)

algorithm. In this context, a good coarse matching is an essential prerequisite for a fast and successful ICP convergence, which involves an iterative solution aimed at minimising the mismatch between corresponding points in overlapping regions of adjacent point clouds.

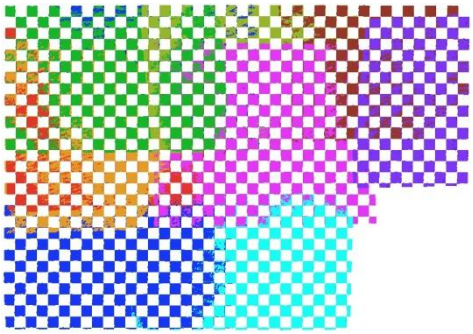
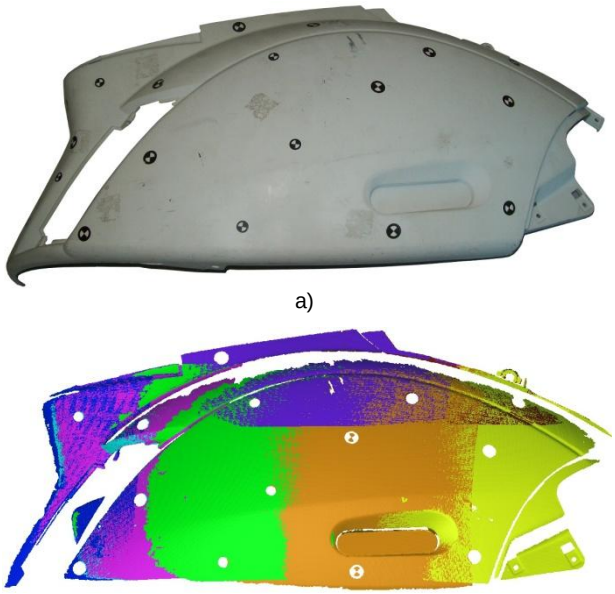


Fig. 9 Point clouds alignment of the target planar surface.

	μ_{plane} [mm]	σ_{plane} [mm]
1	-0.032	0.224
2	-0.029	0.358
3	0.239	0.946

Tab. 1 Mean and standard deviation of the residual errors of the plane best fitting the points acquired in three different measurements of the planar surface.



a)

b)

Fig. 10 a) Motorcycle component and b) eight different point clouds aligned by the proposed tracking methodology.

5 Conclusion

In this paper, a low-cost optical tracking system has been developed with the aim at measuring free form surfaces by automatically aligning 3D point clouds captured by an optical scanner. The system is based on the detection of passive infrared markers on stereo-pair images. 3D marker frames are then created establishing accurate stereo correspondences, by integrating epipolar and similarity constraints, and used to align the point clouds. An image based calibration phase, requiring a minimal human intervention, has been defined to determine the location of the 3D marker frames with respect to the optical scanner.

Results have demonstrated a significant robustness of the system with respect to occlusions and illumination conditions. Moreover, the accuracy obtained in the automatic alignment of multiple point clouds legitimizes its use in the speed up of the measurement process of complex shapes for reverse engineering applications, especially when conventional high-end systems cannot be applied.

References

[1] F. Bernardini and H. Rushmeier. *The 3D Model Acquisition Pipeline*. Computer Graphics Forum 21, 2 (2002) pp. 149-172.

[2] B. M. Planitz, A. J. Maeder, J. A. Williams. *The correspondence framework for 3D surface matching algorithms*. Computer Vision and Image Understanding 97, 3 (2005) pp. 347-383.

[3] J. Salvi, C. Matabosch, D. Fofi, J. Forest. *A review of recent range image registration methods with accuracy evalutation*. Image and Vision Computing 25, 5 (2007) pp. 578-596.

[4] V. Carbone, M. Carocci, E. Savio, G. Sansoni, L. De Chiffre. *Combination of a Vision System and a Coordinate Measuring Machine for Reverse Engineering of Freeform Surfaces*. The International Journal of Advanced Manufacturing Technology 17, 4 (2001) pp. 263-271.

[5] M. Callieri, A. Fasano, G. Impoco, P. Cignoni, R. Scopigno, G. Parrini, G. Biagini. *RoboScan: An Automatic System for Accurate and Unattended 3D Scanning*. Proceedings of 3DPVT'04, September 6 – 9, 2004, Thessaloniki, pp. 805-812.

[6] G. Welch and E. Foxlin. *Motion tracking: No silver bullet, but a respectable arsenal*. IEEE Computer Graphics & Applications 22, 6 (2002) pp. 24–38.

[7] A. van Rhijn, J. D. Mulder. *Optical Tracking and Calibration of Tangible Interaction Devices*. Proceedings of IPT EGVE Workshop (2005), October 6 - 7, 2005, Aalborg, pp. 41-50.

[8] VICON website. http://www.vicon.com/ accessed 30 Nov 2010

[9] Advanced Real Tracking GmbH website. http://www.ar-tracking.de/ accessed 30 Nov 2010.

[10] Atracsys website. http://www.atracsys.com/ accessed 1 Dec 2010

[11] M. Ribo, A. Pinz, A. L. Fuhrmann. *A new Optical Tracking System for Virtual and Augmented Reality Applications*. Proceedings of IMTC 2001, May 21 – 23, 2001, Budapest, pp. 1932-1936.

[12] K. Dorfmüller. Robust tracking for augmented reality using retroreflective markers. Computers & Graphics 23, 6 (1999) pp. 795-800.

[13] A. Hornung, S. Sar-Dessai, L. Kobbelt. Self-Calibrating Optical Motion Tracking for Articulated Bodies. Proceedings of VR'05, March 12 – 16, 2005, Bonn, pp. 75-82.

[14] H. Zhou and H. Hu. *Human motion tracking for rehabilitation: A survey*. Biomedical Signal Processing and Control 3, 1 (2008) pp. 1-18.

[15] A. A. Farag, M. S. Hassouna, A. El-Baz, S. Hushek. *Real-time vision-based approach for probe tracking in operating room*. International Congress Series 1268, (2004) pp. 467-472.

[16] A. Lorsakul, J. Suthakorn. Toward Robot-Assisted Dental Surgery: Path Generation and Navigation System Using Optical Tracking Approach. Proceedings of ROBIO 2008, February 22 – 25, 2009, Bangkok, pp. 1212-1217.

[17] Breuckmann GmbH website. http://www.breuckmann.com/en/industry-technology/products/naviscan.html accessed 2 Dec 2010.
[18] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE Journal of Robotics and Automation RA-3, 4 (1987) pp. 323-344.
[19] S. Barone and A. V. Razionale. *A reverse engineering methodology to capture complex shapes*. Proceedings of XVI INGEGRAF 2004, June 2 – 4, 2004, Zaragoza.
[20] R. I. Hartley and P. Sturm. *Triangulation*. Computer Vision and Image Understanding 68, 2 (1997) pp. 146-157.
[21] M. Pilu. *A Direct Method for Stereo Correspondence base on Singular Value Decomposition*. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 17 – 19, 1997, San Juan, pp. 261-266.
[22] G. Scott and H. Longuette-Higgins. *An Algorithm for Associating the Features of Two Images*. Proceedings of the Royal Society of London, Series B: Biological Sciences 244, (1991) pp. 21-26.
[23] R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, 2000.
[24] A. Fusiello, E. Trucco, A. Verri. *A compact algorithm for rectification of stereo pairs*. Machine Vision and Applications 12, 1 (2000) pp. 16-22.
[25] D. W. Eggert, A. Lorusso, R. B. Fischer. *Estimating 3-D rigid body transformations: a comparison of four major algorithms*. Machine Vision and Applications 9, (1997) pp. 272-290.