2.5D Feature Tracking and 3D Motion Modeling

Mozhdeh Shahbazi
Department of Applied Geomatics, University of Sherbrooke
2500, Boul. de l'Universite
Sherbrooke, Canada

ABSTRACT

Image-based tracking of objects is becoming an important area of research within computer vision and image processing community. However, there are still challenges with regard to robustness of the algorithms. This paper explains an algorithm to track the pre-defined objects within stereo videos (image sequences) in a condition where cameras are fixed and objects are moving. The tracking technique used in this research, applies the intensity-based least squares matching (LSM) to find the correspondent targets in successive frames. Unlike ordinary correlation-based registration methods, LSM takes both geometric and radiometric variations of images into account, succeeding at sub-pixel scale feature tracking. The proposed algorithm combines three dimensional updated object constraints with adaptive two dimensional LSM to ensure the robustness and convergence to optimum solution. While tracking the features in stereo images, photogrammetric techniques are applied to extract the coordinates of the features in object space which result in detecting the 3D trajectory of the features. The average tracking error is about 0.11 pixel at x-direction and 0.15 pixel at y-direction. The 3D motion vectors are modeled by mean magnitude precision of 0.65 millimeter and orientation precision of 0.27 degree.

General Terms

Algorithm, photogrammetry, tracking

Keywords

Motion modeling, feature, stereo-vision, least squares matching, calibration

1. INTRODUCTION

Image-based tracking of objects is becoming an important area of research within computer vision and image processing community [1]. Different studies have provided tracking algorithms to fulfill various applications such as vehicle and pedestrian monitoring [2], medical image registration [3], mobile mapping [4], lip tracking for speech processing [5], body motion detection like facial motion analysis [6]. The features can be defined as different structures in the image itself such as points and edges or as more complex structures defined based on an object. Feature tracking is one of the most fundamental operations in computer vision - it is probably the most popular way of extracting motion information from image sequences, namely videos. It is also applicable in the field of videogrammetry which is extracting and following the three dimensional models from video frames. Besides, it is the main key in iterative image matching and co-registration [7].

As the object or the camera moves and rotates within the imaging scene, the patterns of image intensities change in a complex way. Therefore, the features are, to some extent, deformed both geometrically and radiometrically [8].

Least squares matching is one of the most popular techniques applied, so far, for both multi-image matching, target locating, 3D surface matching and tracking (e.g. [3], [4], [9], [10], [11]); since it considers the geometric and radiometric variations of correspondence targets in a flexibly adoptive way. However, as a non-linear problem, LSM is tightly dependent to initialization procedures such as determining the approximate values of the variables, matching window size and observation weights [12,13,14]. Otherwise, the adjustment might completely fail or converge to a false solution. Besides, in most applications, image tracking is not the only matter of concern while three dimensional motion modeling of the moving objects is interested as well.

This paper improves the conventional least squares matching aimed to feature tracking on image sequences by adding object-space constraints to initialization procedure, providing a two and a half robust tracking algorithm. While tracking the features in stereo images, photogrammetric techniques are applied to extract the coordinates of the features in object space which result in simultaneously detecting the 3D trajectory of the features. It is assumed that applying this adoptive LSM technique along with stereo-vision facilitates tracking the 3D moving features in a video sequence.

2. IMAGING GEOMETRY

As mentioned in Section 1, the problem is to track the determined objects within stereo videos in a condition where cameras are fixed. The cameras, applied in this study, are two identical Canon HD camcorders (VIXIA HF R30) with 3.28 megapixels, 1/4.85-inch CMOS sensors.

The cameras are separately calibrated by photogrammetric self-calibration technique, explanation of which is out the scope of this paper; readers are referred to [15, 16]. By the way, their lens and sensor parameters including principal distance, principal point coordinates, lens distortion parameters and sensor electrical biases are determined via calibration. These parameters are applied to compute the systematic errors of image observations (subsubsection 3.2.2). The results of self-calibration procedure are listed in Table 1.

The objects to be tracked on the stereo videos are designed as white circles at black background. Eight white circles with two centimeters diameters are plotted on paper and pasted on a wooden surface so that its displacing is made easy. The features are designed as circles so that their automatic localization on images would also be possible by the circle detection method of this paper and, thus, the results of LSM tracking can be evaluated.

The cameras are installed on two tripods on a stable surface and are rotated so that their view fields are well overlapped on the scene where the targets are moved. The target plane is depicted in Figure 1.

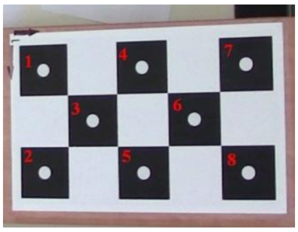

Fig 1. Targets installed on wooden surface

As there has been no electrical instrument to trigger the cameras at precisely the same time, the old-fashioned method of filming clapboards is used. When the cameras are well installed and the recording buttons are pressed, two boards are clapped in front of the cameras and to verify the accuracy of synchronization, they are clapped once more at the end of the test. Then at the programs, the instants, in which the boards are just clapped, are distinguished.

The difference between the frame numbers in two cameras is a constant time. For example, in this test, the clapping instant in the first video takes place at frame number 858 and the same instant in the second video happens on frame number 1042; i.e. the time difference between two cameras is 184 frames. To verify that this difference is constant through the whole test period, the same measurement is done for the ending clapboards. It is realized that the ending frame on the first video is the 5981th and on the second one is the 6165th; i.e. the difference has been constant through the test. Thereafter, each frame from the first video is correspondent to its unique stereo pair in the second video and vice versa by the determined time difference.

In order to determine the features 3D trajectory, an arbitrary object-space coordinate frame requires to be defined. The origin of the object coordinate system is established on the upper left corner of the target plane "at the beginning of the video", the wooden plane of targets is considered as XY plane with Z axis pointing outwards and the true distance between the circles determines the scale to ensure that motions vectors are modeled at the real ground scale.

3. THEORECTICAL PRINCIPLES

3.1 Least Squares Adjustment

Herein, the basics of the least squares adjustment of the second form are explained. Assume a set of equations in which observations are nonlinear function of some parameters. The problem is to estimate those parameters. If the number of equations (observations) are more than the number of unknowns (parameters), then there won't be a unique solution for each unknown. Here the least squares solution is one the possible solutions by which sum of squared residuals between real observations and estimated observations as functions of parameters would be minimized [17]. The mathematical model of observation equations expresses the relationship between observation vector $\vec{L}$ and unknown vector $\overrightarrow{X}$.

$$\vec{L} = F(\overrightarrow{X}) \tag{1}$$

The vector function $\boldsymbol{F}$ represents equations of observations. Since the equations are nonlinear regarding the unknowns, linearization is accomplished by replacing the nonlinear functions by their Taylor series approximation. That is

$$\vec{L} - F(\vec{X}^0) + \frac{\partial F}{\partial \vec{X}} \delta \vec{X} = 0 \tag{2}$$

Renaming the above matrices leads to

$$\overrightarrow{W} + A.\delta\vec{X} = 0 \tag{3}$$

where W is called miss-closure vector and A is design matrix. The least squares estimation of unknowns is obtained by the following equation.

$$\delta\hat{\vec{X}} = (A^T PA)^{-1} A^T P\overrightarrow{W} \tag{4}$$

where P is the weight matrix of observations.

This solution, $\delta\hat{\vec{X}}$, must be added to the initial approximations of unknowns, $\vec{X}^0$, to improve the solutions for next iteration. The Iterations are repeated recursively until reaching a convergent solution.

The covariance matrix of solutions is computed by equation 5. The elements on the diagonal of the covariance matrix are the squared standard deviations (variances) of the resolved variables which are a measure of the solution precision.

$$C_X = (A^T PA)^{-1} \tag{5}$$

3.2 Photogrammetric Issues

3.2.1 Collinearity Condition

Regarding the optical and photogrammetric point of view, each ground point is related to its image observation by the collinearity equations as follows:

$$\begin{aligned} x + \Delta x + f\frac{U}{W} = 0 \\ y + \Delta y + f\frac{V}{W} = 0 \end{aligned} \tag{6}$$

where:

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = R_3(\kappa).R_2(\varphi).R_1(\omega)\begin{bmatrix} X - X^C \\ Y - Y^C \\ Z - Z^C \end{bmatrix} \tag{7}$$

In the above equations, (x, y) are the image coordinates of the observation on the camera, (X,Y,Z) are the object-space coordinates of the point, $(\omega,\varphi,\kappa, X^C, Y^C, Z^C)$ are exterior orientation parameters of the camera in which (ω,φ,κ) are the rotation angles of the imaging coordinate system of the camera with respect to object-space coordinate system around X, Y and Z axes respectively and (X^C,Y^C,Z^C) are the coordinates of the camera perspective center according to the object space coordinate system. The parameter f is principal distance of the camera which is accurately determined in calibration process; R_1,R_2,R_3 are the fundamental rotation matrices and $(\Delta x, \Delta y)$ are systematic errors on the image point which are calculated from camera calibration parameters as in equation 8:

$$\begin{aligned}\Delta x &= -x_p + (x - x_p).(K_1 r^2 + K_2 r^4 + K_3 r^6) + P_1(r^2 + 2(x - x_p)^2 \\ &\quad + 2P_2(x - x_p).(y - y_p) + A_1(x - x_p) + B_1(y - y_p) \\ \Delta y &= -y_p + (y - y_p).(K_1 r_i^2 + K_2 r^4 + K_3 r^6) + P_2(r^2 + 2(y - y_p) \\ &\quad + 2P_1(x - x_p).(y - y_p)\end{aligned} \qquad (8)$$

where (K_1, K_2, K_3) are the radial lens distortion coefficients, (P_1, P_2) are the decentring les distortion terms, (A_1, B_1) are the electronic biases, namely affinity and shear, and (x_p, y_p) are principal point coordinates of the sensor. The radius, *r*, is defined as the distance between the image point and the principal point.

The photogrammetric space resection using collinearity equations means determining the exterior orientation parameters of the camera, $(\omega, \varphi, \kappa, X^C, Y^C, Z^C)$, from equation 6. Since digital photogrammetry is not the main subject of this paper, further information can be found on [15, 18].

On the other hand, the space intersection by collinearity condition, means determining the object coordinates, *(X,Y,Z)*, given the exterior orientation parameters. The intersection procedure is briefly explained in the following subsubsection as it plays an important role in the presented algorithm.

3.2.2 Space Intersection

Applying space intersection for two images, whose exterior orientation parameters are specified, makes it possible to calculate the object coordinates for points that lie in the stereo overlap area.

To calculate the object point coordinates, the collinearity equations of subsubsection 3.2.1 can be reformed to the following format:

$$\begin{aligned} X - X^C - (Z - Z^C)\frac{D}{F} = 0 \\ Y - Y^C - (Z - Z^C)\frac{E}{F} = 0 \end{aligned} \qquad (9)$$

where:

$$\begin{bmatrix} D \\ E \\ F \end{bmatrix} = R_1(\omega).R_2(\varphi).R_3(\kappa) \begin{bmatrix} x + \Delta x \\ y + \Delta y \\ f \end{bmatrix} \qquad (10)$$

As implied by equation 9, for projection of each object point at one image, two equations can be formed. There are three unknown parameters *(X,Y,Z)* while there are only two equations. That's when the same two collinearity equations from the stereo pair are added to the observation equations; i.e. there are four observation equations from two images to solve three unknowns. The optimum solution is acquired by applying least squares adjustment. Readers are referred to [15,18] for further details.

3.3 Least Squares Matching

Least squares matching provides several advantages over simple normalized cross correlation (NCC). First of all, the sub-pixel accuracy, in NCC, is obtained virtually by interpolating the correlation values on discrete pixels to estimate the fractional peak. However, LSM directly results in sub-pixel accuracy. Although the normalized cross correlation is invariant against mean gray level, it is not to local dissimilarities and rotations. Since the object moves and rotates in different directions, its radiometric and geometric characteristics may differ from one image to another so much that the correlation coefficient would not be an ideal measure of similarity and translation (shift) would not be an effective transformation between two image patches [19]. LSM allows simultaneous radiometric correction and local geometrical transformation, whereby the system parameters are automatically assessed, corrected, and thus optimized during the least squares iterations.

As any other matching problem, given a coordinate at left image, the problem is to find its corresponding point at right one. Assume two image windows are given as discrete two-dimensional functions *f(x,y)*, and *g(x,y)*. *f(x,y)* and *g(x,y)* can be defined as conjugate patches of a stereo pair in the ‘left’ and the ‘right’ images respectively. Typically, *f(x,y)* is named as the template and *g(x,y)* as the target. If the centers of these two windows are correspondent and completely correlated, then the following equation is true.

$$f(x, y) = g(x, y) \qquad (11)$$

The difference values, *f(x,y)-g(x,y)*, are actually the distances between gray levels of the pixels in template and target. Finding the match point means determining the location of the function values *g(x,y)*. This is achieved by minimizing a goal function which measures the distances between grey levels in template and target. The goal function to be minimized in LSM approach is the L_2-norm of the residuals of least squares estimation.

In the least squares context, equation 11 can be considered as an observation equation which models the vector of observations *f(x, y)* with a function *g(x, y)*, whose location in the right image needs to be determined. The location can be described by shift parameters. However, to account for a variety of systematic image deformations and to obtain a better match, image shaping parameters and radiometric corrections are introduced in addition to shift parameters. In this study, the image shaping is considered as an affine transformation and the radiometric parameters are considered as simple radiometric gain and drift.

According to the aforementioned hypotheses, equation 11 can be re-formulated.

$$f(x, y) = G_0 + G_1 \times g(T_1(x), T_2(y)) \qquad (12)$$

G_0 and G_1 are the radiometric correction parameters, respectively gain and drift. T_1 and T_2 are the affine transformations which move each pixel of the template from left image to its correspondent on the right image as follows.

$$\begin{aligned} T_1(x) = a_0 + a_1 x + a_2 y \\ T_2(y) = b_0 + b_1 x + b_2 y \end{aligned} \qquad (13)$$

In equation 12, for each template there are eight unknown parameters. Therefore, the template should contain at least eight pixels. However, more pixels are needed to ensure that the template window reflects the spectral properties of the region around its center point, the point that should be matched to a point on right image. Later, the approach used in this paper to determine the window size adoptively is explained.

Assume the template is an *N* by *N* window from the left image and the problem is to find the corresponding point to its center at the right image using equation 12. For each pixel in the template with coordinate *(x_i, y_i)*, one observation equation can be formed by equation 12. Therefore, there are N^2 observation equations and eight unknowns. To solve this problem, the least squares approach is utilized.

In order to use the least square solution of subsection 3.1, the observation and unknown vectors are first formed.

$$\vec{L} = \begin{bmatrix} f(x_1, y_1) \\ f(x_2, y_2) \\ \vdots \\ f(x_{N^2}, y_{N^2}) \end{bmatrix}, \qquad \vec{X} = \begin{bmatrix} G_0 & G_1 & a_0 & b_0 & a_1 & b_1 & a_2 & b_2 \end{bmatrix}^T \tag{14}$$

The function at the right hand of equation 12 should be replaced by its Taylor series approximation:

$$f(x,y) = G_0^0 + G_1^0 \times g^0 + [1 \quad g^0 \quad G_1^0 \frac{\partial g^0}{\partial x} \quad G_1^0 \frac{\partial g^0}{\partial y} \quad G_1^0 \frac{\partial g^0}{\partial x} x \quad G_1^0 \frac{\partial g^0}{\partial y} x \quad G_1^0 \frac{\partial g^0}{\partial x} y \quad G_1^0 \frac{\partial g^0}{\partial y}] \delta \vec{X} \tag{15}$$

where:

$$g^0 = g(a_0^0 + a_1^0 x + a_2^0 y, b_0^0 + b_1^0 x + b_2^0 y) \tag{16}$$

in which any variable with superscript "*0*" represents the initial value of that variable which will be updated at any iteration of adjustment by the estimated $\delta \hat{\hat{X}}$ (equation 4) and image gradients $(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y})$ are calculated by convolutions of target window by Sobel kernels.

As indicated by equation 15, the integer pixel coordinates from left image are transformed to decimal coordinates at the right images. Therefore, it would be necessary to estimate the gray level of the points at non-integer coordinates. The method used in this study is a bilinear interpolation. Bilinear interpolation is an extension of linear interpolation for interpolating functions of two variables on a regular 2D grid (like an image). Assume the problem is to determine the gray value at location (x^0, y^0) on the right image. The coordinates (x^0, y^0) are decomposed to their integer and fractional parts:

$$\begin{aligned} P &= \mathrm{int}(x^0) \rightarrow x^0 = P + p \\ Q &= \mathrm{int}(y^0) \rightarrow y^0 = Q + q \end{aligned} \tag{17}$$

The final gray value $g(x^0, y^0)$ is computed using equation 18.

$$g(x^0, y^0) = A \times p + B \times q + C \times p \times q + D \tag{18}$$

where:

$$\begin{aligned} D &= g(P,Q) \\ C &= g(P+1,Q+1) + g(P,Q) - g(P+1,Q) - g(P,Q+1) \\ B &= g(P,Q+1) - g(P,Q) \\ A &= g(P+1,Q) - g(P,Q) \end{aligned} \tag{19}$$

As a non-linear problem, LSM accuracy is highly dependent to the initial values fed to the adjustment. The solution to handle this sensitivity, applied in this paper, is to perform LSM in three steps. In the first step, the equation 12 is reduced to a simpler problem in which the only shaping transformation are the shift parameters; i.e. only a_0 and b_0 are considered as unknown variables while (G_0, a_2, b_1) are set to zero and (G_1, a_1, b_2) are set to one. In the second step, the original problem is reduced so that radiometric gain and offset (G_0, G_1) are the only unknown parameters while a_0 and b_0 are set fixed to their values updated by the first step and (a_1, b_2) are set to one and (a_2, b_1) are set to zero. Finally, in the last step, the entire original problem with all eight parameters has to be solved. However, the approximate values for (G_0, G_1, a_0, b_0) are this time closer to the reality in comparison with their very initial values. This, greatly, helps the adjustment iterations to converge to the globally optimum solutions.

However, the important issues to ensure robustness of LSM are yet remained as defining the initial shift values (a_0, b_0), defining the template window size (*N*) and giving appropriate weights to observations (defining matrix *P* in equation 4). They are all addressed in Section 4.

4. TRACKING METHODOLOGY

Back to the tracking problem, in order to follow the features trajectory in true scale, the exterior orientation of the cameras should be determined in the same scale. As described in Section 2, the arbitrary object-space coordinate is defined on the first frames of the test. Therefore, the object-space position of the features can be coordinated.

If the features can be located precisely on the first frames, the exterior orientation parameters of the cameras can be determined via space resection. In this paper, a semi-automatic circle detection technique is developed. It is used for locating the features on the first and the second frames of the cameras as well as evaluating the accuracy of the proposed tracking algorithm.

4.1 Circular Target Detection

As the object moves in different directions and orientations, the circular targets are no longer projected as exact circles on the image. Their image is transformed to an ellipse which might be oriented at any direction.

As explained before, the exterior orientation parameters of the cameras will be extracted using the targets of the first frames. It means that precision of determining the centers of the circular features on the first frames plays an important role on the accuracy of exterior orientation parameters. Consequently, the accuracy of space intersection for further tracked features would highly rely on the accuracy of exterior orientation parameters of the cameras too. According to the sensor pixel size, lens focal length and the initial distance of the cameras from the targets, one pixel error in detecting the targets would lead to five millimeters error on camera positioning, applying the rule of error propagation. This clarifies the importance of a proper target detection method which is able to localize the targets by sub-pixel accuracy.

To determine the accurate center coordinates of the circular targets, the approximate place of the circle is given to the algorithm manually. According to the diameter of the circles (2 centimeters), the approximate distance of the cameras from the targets (1500 millimeters), the lens focal length and sensor pixel size, it can be seen that each circle is imaged on an area of 25 pixels width. Therefore, an area of 50 pixels around the approximate position of the circle is considered as the processing window.

The processing window is converted to binary system applying Otsu's method of gray level thresholding in which the first and zero order cumulative moments of the image histogram are used to determine the global threshold [20]. The binary window is searched for closed-connected components by defining eight-neighbor connectivity [21]. As a result, the parent regions, ignoring the child holes, are segmented from the background. Putting the constraints on the upper and lower limits of candidate circular element's primitive and area, the closed components are reduced to the circular target of interest whose center should be determined.

Then, an ellipse is fit to the pixels of the inner border of the closed region by least squares fitting, according to the following equation:

$$(x-x_0)^2 + K.(y-y_0)^2 = R^2 \qquad (20)$$

where (*x,y*) are the coordinates of the pixels on the inner border of the component, (x_0,y_0,*K*,*R*) are the unknown parameters of ellipse, namely, center coordinates in x and y direction, scale factor and radius. Putting equation 20 into least squares adjustment iterations, the terminating condition is set to ensure that the estimated parameters are converged better than 0.01 pixel. This means that the targets centers are determined by 0.01 pixel accuracy.

4.2 Tracking Algorithm

Starting by the first frames, the exterior orientations of the cameras are determined via space resection using collinearity condition (equation 7), where object coordinates are known, the image features (circles centers) are measured accurately applying the circular target detection technique. The exterior orientation parameters of the cameras are hold fixed through the test as the cameras are not moving.

Going to the second frames, the object is moved and positions of features both on object space and image plane are varied. In the second frames, again, features are detected by the circular target detection method. Applying photogrammetric space intersection leads to the 3D coordinates of the targets in object-space frame.

Thereupon, the tracking is completely automated. Assume that we have a feature, to be tracked from the *t^th^* frame to the *t+1^th^* frame. The following procedure is developed to solve this problem.

1- The object coordinates of the feature from the t^{th} and $t\text{-}1^{th}$ frames are already determined via space intersection of stereo pairs.

2- As the frame interval is so short in time, the movements of the object between two successive frames can be considered linear with constant speed. Therefore, the approximate object coordinates of the feature in the $t+1^{th}$ frame can be estimated linearly from the coordinates in the previous frames by:

$$\vec{X}^{t+1} = \vec{V}.\Delta t + \vec{X}^{t} \qquad (21)$$

where $\vec{V}$ is the velocity vector estimated from two previous frames as follows.

$$\vec{V} = \frac{\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}^{t} - \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}^{t-1}\right)}{\Delta t} \qquad (22)$$

Substituting equation 22 into equation 21 yields the final formula to approximate coordinates of the object in the $t+1^{th}$ frame.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}^{t+1} = 2\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}^{t} - \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}^{t-1} \qquad (23)$$

3- The object coordinates of the $t+1^{th}$ frame are fed into equation 6 to estimate the image coordinates (x^{t+1}, y^{t+1}) where the object point is imaged.

$$\begin{aligned} x^{t+1} &= -(\Delta x^{t+1} + f\,\frac{U^{t+1}}{W^{t+1}}) \\ y^{t+1} &= -(\Delta y^{t+1} + f\,\frac{V^{t+1}}{W^{t+1}}) \end{aligned} \qquad (24)$$

4- The approximate values for shift parameters (a_0 and b_0) to transfer the feature from the t^{th} frame to the $t+1^{th}$ frame can now be estimated as:

$$\begin{aligned} a_0 &= x^{t+1} - x^{t} \\ b_0 &= y^{t+1} - y^{t} \end{aligned} \qquad (25)$$

5- The approximate imaging scale of the feature is determined by the following equation.

$$\lambda^{t+1} = \frac{\sqrt{(x^{t+1}-x_p)^2 + (y^{t+1}-y_p)^2 + f^2}}{\sqrt{(X^{t+1}-X^C)^2 + (Y^{t+1}-Y^C)^2 + (Z^{t+1}-Z^C)^2}} \qquad (26)$$

Multiplying the target diameter (two centimeters) by the imaging scale, gives an approximation of feature size on the $t+1^{th}$ frame. Therefore, the LSM template window size (*N*) is adoptively defined as twice the approximate feature size on the $t+1^{th}$ frame.

6- The least squares matching technique is now performed on the $t+1^{th}$ frame as "right" image and the t^{th} frame as "left" image. LSM is initiated by the values of a_0 and b_0 from equation 25. The exact image coordinates (x^{t+1}, y^{t+1}) of the feature at the $t+1^{th}$ frame along with their estimated variances are calculated at LSM procedure.

If *t* is equal to 2, then the observations weights are determined by the accuracy output of circular target detection. Elsewhere, the least squares matching errors from the t^{th} frame determine the observations weights.

7- The same procedure from the third step is performed for the other stereo pair as well.

8- Applying space intersection on stereo images, the accurate 3D object-space coordinates of the feature at the $t+1^{th}$ frame are calculated.

9- Moving to the next frame, the whole procedure from the first step is repeated and the feature is sequentially tracked through the frames while its object-space coordinates are calculated as well.

5. EXPERIMENTAL RESULTS

The results of camera calibration and exterior orientation form the first frames are summarized in Tables 1 and 2. Figure 2 illustrates the cameras orientations with regard to the defined object-space coordinate system (Section 2).

Table 1. Calibration parameters of the cameras

Parameter	Camera1	Camera2
K_1	0.007962	0.008288
K_2	0.000075	0.000032
K_3	-0.000062	-0.000056
P_1	0.000503	-0.000053
P_2	0.000171	0.000332
x_p (mm)	-0.085548	-0.004828
y_p (mm)	-0.024051	-0.001796
f (mm)	4.820844	4.829939
A_1	0.000335	-0.000967
B_1	-0.000562	0.000181

Table 2. Exterior orientation parameters of the cameras

Parameter	Camera1	Camera2
ω (deg)	4.069±0.102	-8.653±0.109
φ	42.626±0.075	48.112±0.074
κ	79.710±0.066	90.781±0.055
X^C (cm)	137.05±0.11	147.41±0.13
Y^C	2.81±0.17	62.33±0.18
Z^C	127.08±0.15	127.38±0.16

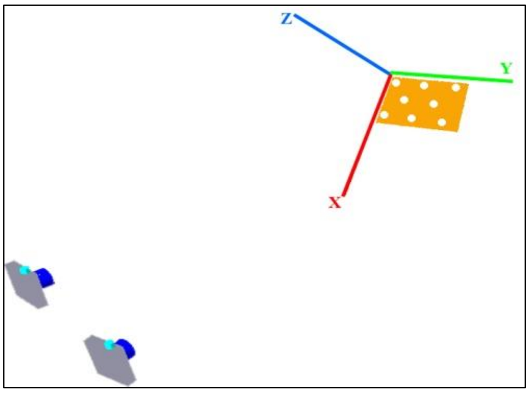

Fig 2. Cameras, initial position of targets, object-space coordinate system

The systematic error of image observations based on equation 8 and the parameters of Table 1 are plotted against the radial distance from the principal point. The results are compared for two cameras in figure 3.

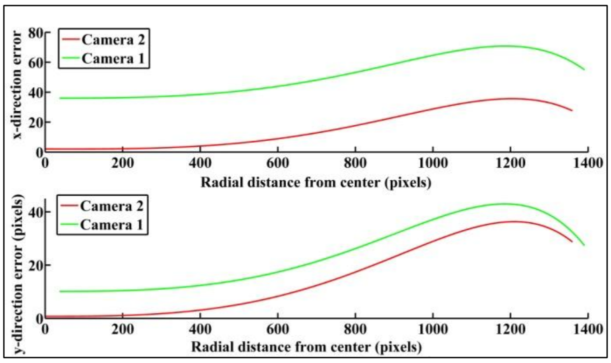

Fig 3. Systematic errors of cameras vs. radial distance

Figure 3 explains how much camera calibration affects the accuracy of the results and why separate calibration of both cameras should be performed while the cameras are nominally identical.

In order to assess the accuracy of the feature tracking algorithm, the image coordinates of the features are measured by semi-automatic circular target detection (subsection 4.1) on some random frames and the results are compared by those of LSM tracking. The mean residuals at eight features are illustrated in Figure 4. The average tracking error is about 0.11 pixel at x-direction and 0.15 pixel at y-direction.

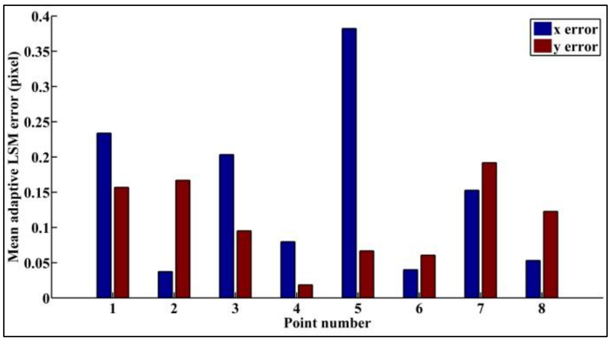

Fig 4. Average tracking error of test features

In order to evaluate the accuracy of motion modeling, in terms of magnitude, the true distances are compared with those of computed 3D coordinates. As the targets are originally printed at predefined coordinates on the paper, it is known that the distances between targets 1 and 2, targets 3 and 6, targets 4 and 5 as well as targets 7 and 8, are the same and equal to 160 millimeters. Therefore, the distances between these targets from their computed coordinates at each epoch are calculated and compared to the true 160 millimeters length. The distance residual between targets *i* and *j*, at frame *t*, r_{i-j}^t, is computed as:

$$r_{i-j}^t = \left|160 - \sqrt{\left(X_t^i - X_t^j\right)^2 + \left(Y_t^i - Y_t^j\right)^2 + \left(Z_t^i - Z_t^j\right)^2}\right| \qquad (27)$$
$$for\ \ (i,j) \in \{(1,2),\ (3,6),\ (4,5),\ (7,8)\}$$

in which, (X_t^i, Y_t^i, Z_t^i) are the 3D coordinates of the target *i* computed at frame *t*. r_{i-j}^t is the error of calculating the distance between targets *i* and *j* at moment *t*.

Entirely, the mean and variance of the distance residuals are 0.65 and 0.34 millimeter respectively.

In order to evaluate the accuracy of positioning in terms of orientation, the true right angles are compared with those of computed coordinates as follows:

$$\theta_{i-j-k}^t = \left|90^\circ - \cos^{-1}\left(\frac{\overrightarrow{\Delta X}_t^{i-j} \bullet \overrightarrow{\Delta X}_t^{k-j}}{\left|\overrightarrow{\Delta X}_t^{i-j}\right| \times \left|\overrightarrow{\Delta X}_t^{k-j}\right|}\right)\right|$$
$$for\ \ (i,j,k) \in \{\ (4,1,2),\ (3,5,6),\ (5,8,7)\ \} \qquad (28)$$

where:

$$\overrightarrow{\Delta X}_t^{i-j} = (X_t^i - X_t^j, Y_t^i - Y_t^j, Z_t^i - Z_t^j)$$
$$\overrightarrow{\Delta X}_t^{k-j} = (X_t^k - X_t^j, Y_t^k - Y_t^j, Z_t^k - Z_t^j)$$

In equation 28, θ_{i-j-k}^t is called orientation residual which is the error of calculating the right angle between three targets *i*, *j* and *k* whose vertex is target *j*.

Totally, the mean and variance of the orientation residuals are 0.27 and 0.12 degree respectively.

This means that, on average, the features trajectory is determined in 3D object-space by magnitude precision of 0.65 millimeter and orientation precision of 0.27 degree.

In Figures 5-6, samples of tracking results for two frames is depicted where the upper image shows the 3D coordinates of the objects, the lower left image is the frame of the first camera on which the features are tracked and demonstrated by colored plus signs, and the lower right image is the correspondent frame from the second camera on which the targets are shown by respective colors.

6. CONCLUSION

The least squares matching technique is improved in this study to track the image features. As a flexible registration technique, LSM takes both geometric and radiometric transformations of two images into account and enables a precise sub-pixel registration. As a non-linear solution, however, LSM is strongly sensitive to initial values of variables. Therefore, a three step adjustment procedure is performed which reduces this sensitivity by predicting more suitable approximations at each step.

Howbeit, the most important initial values to be determined are the shift parameters which transfer the feature from one frame to its subsequent one. Unlike conventional studies, a ground-based adaptive methodology is developed in this paper to approximate the shift initial values, determine the window size and define observations weights. The proposed methodology is tested on a set of pre-defined features while they are moved and rotated irregularly. The tracking technique succeeds in following the features frame-by-frame precisely. Experiments show that having a fixed window size and estimating the initial shift parameters based on cross correlation make the LSM fails because of rapid object deformations. The proposed algorithm tracks the features more accurately than 0.15 pixel in both directions.

At the same time as extracting the features, their 3D coordinates are calculated from stereo pairs and motion vectors are visualized with magnitude and orientation precisions better than 0.7 millimeter and 0.3 degree.

The circular targets are chosen in this study to make the evaluation of tracking procedure possible. In the future, the proposed algorithm will be evaluated on body motion detection which will be highly applicable in actor animation, speech recognition and relevant applications.

7. REFERENCES

[1] Tissainayagam, P., and Suter, D. 2005. Object tracking in image sequences using point features. Pattern Recognition, 38(1), 105-113.

[2] Koller, D., Weber, J., and Malik, J. 1994. Robust multiple car trackingwith occlusion reasoning. ECCV 94, 189–196.

[3] Hsu, L. Y., and Loew, M. H. 2001. Fully automatic 3D feature-based registration of multi-modality medical images. Image and Vision Computing, 19(1), 75-85.

[4] Tao, C. V., Chapman, M. A., and Chaplin, B. A. 2001. Automated processing of mobile mapping image sequences. ISPRS journal of Photogrammetry and Remote Sensing, 55(5), 330-346.

[5] Blake, A, and Isard, M. 1998. Active Contours, Springer, Berlin.

[6] Pateraki, M., Baltzakis, H., Kondaxakis, P., and Trahanias, P. 2009. Tracking of facial features to support human-robot interaction. In Proceedings of IEEE International Conference on Robotics and Automation.

[7] Jiang, N. 2009. The extraction, restoration and tracking of image features. Doctoral dissertation, Arizona State University, Arizona.

[8] Smith, S. M. and Brady, J. M. 1995. Real-time motion segmentation and shape tracking. Transactions of IEEE on Pattern Matching and Machine Intelligence, 17(8), 814-820.

[9] Previtali, M., Barazzetti, L., Scaioni, M., and Tian, Y. 2011. An automatic multi-image procedure for accurate 3D object reconstruction. In Proceedings of IEEE Congress on Image and Signal Processing.

[10] Shahbazi, M., and Motagh, M. 2012. Improved Interferometric Synthetic Aperture Radar processing via advanced co-registration and phase correction techniques. In Proceedings of IEEE Conference on Intelligent Data Understanding.

[11] Akca, D. 2007. Matching of 3D surfaces and their intensities. ISPRS Journal of Photogrammetry and Remote Sensing, 62(2), 112-121.

[12] Shin, D., and Muller, J. P. 2012. Progressively weighted affine adaptive correlation matching for quasi-dense 3D reconstruction. Pattern Recognition. 45(1), 3795-3809.

[13] Rosenholm, D. 1987. Least squares matching method:some experimental results. The Photogrammetric Record, 12(70), 493-512.

[14] Rosenholm, D. 1987. Empirical investigation of optimal window size using the least squares image matching method. Photogrammetria, 42(3), 113-125.

[15] Luhmann, T., Robson, S., Kyle, S. and Harley, I. 2007. Close range photogrammetry: principles, techniques and applications. John Wiley & Sons, UK.

[16] Shahbazi, M., Homayouni, S., Saadatseresht, M., and Sattari, M. 2011. Range camera self-calibration based on integrated bundle adjustment via joint setup with a 2D digital camera. Sensors, 11(9), 8721-8740.

[17] Michaiel, E. M. 1976. Observations and least squares. IEP-A Dun-Donelley, USA.

[18] Wolf, P. R., and Dewitt, B. A. 2000. Elements of Photogrammetry: with applications in GIS. McGraw-Hill, USA.

[19] Gruen, A. 1985. Adaptive least squares correlation: a powerful image matching technique. South African Journal of Photogrammetry, Remote Sensing and Cartography, 14(3), 175-187.

[20] Otsu, N. 1979. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62-66.

[21] Gonzalez, R. C., Woods, R. E., and Eddins, S. L. 2004. Digital image processing using MATLAB. Pearson Education, India.

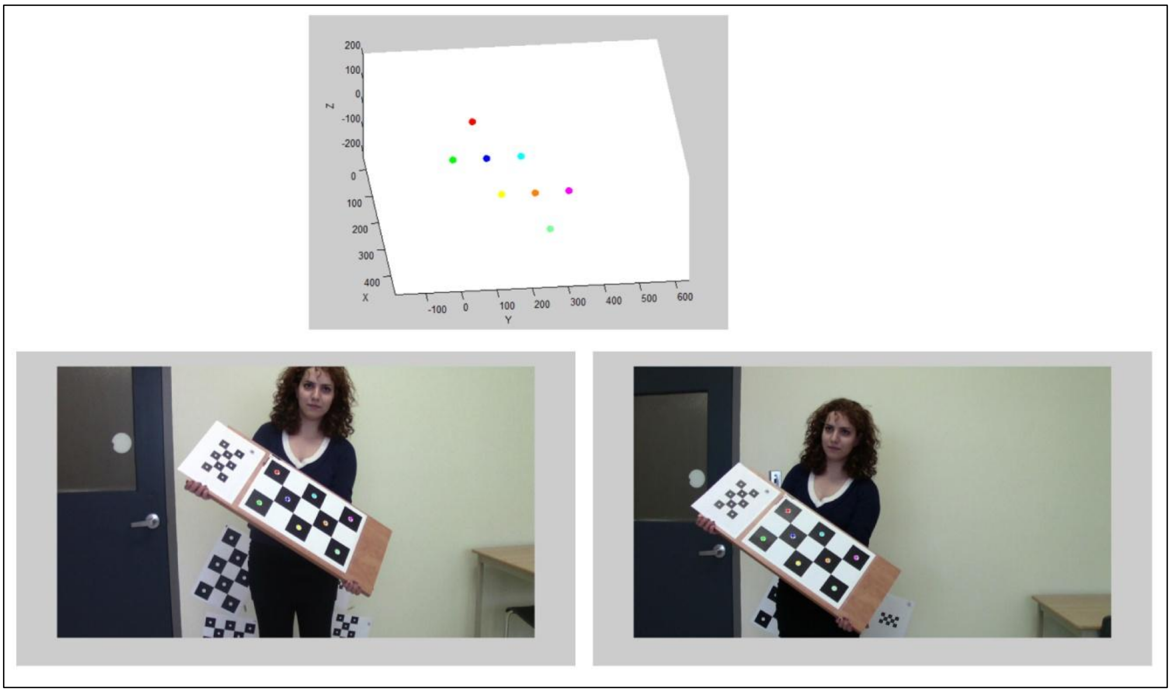

Fig 5. Sample of feature tracking results

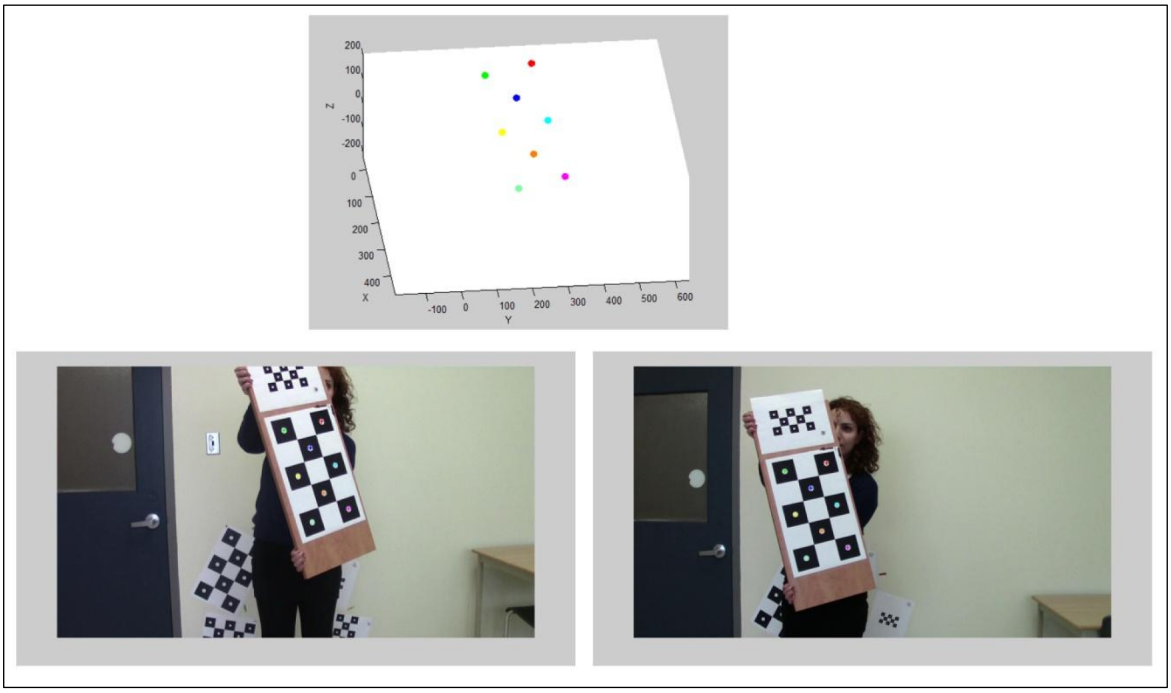

Fig 6. Sample of feature tracking results