Camera Self-localization Using Uncalibrated Images to Observe Prehistoric Paints in a Cave*

Tommaso Gramegna, Grazia Cicirelli, Giovanni Attolico, and Arcangelo Distante

Institute of Intelligent Systems for Automation – C.N.R.
Via Amendola, 122 D-I – 70126 Bari, Italy
{gramegna, grace, attolico, distante}@ba.issia.cnr.it

Abstract. The fruition of archaeological caves, hardly accessible by visitors, can benefit from a mobile vehicle which transmits to users located outside a continuous stream of images of the cave that can be visually integrated with information and data to increase the fruition and understanding of the site. This application requires self-positioning the vehicle with respect to a target. Preserving the cave imposes the use of natural landmarks as reference points, possibly using uncalibrated techniques. We have applied the modified POSIT algorithm (camera pose estimation method using uncalibrated images) to self-position the robot. To account for the difficulty of evaluating natural landmarks in the cave the tests have been made using a photograph of the prehistoric wall paintings of the archeological cave "Grotta dei Cervi". The modified version of the POSIT has been compared with the original formulation using a suitably designed grid target. Therefore the performance of the modified POSIT has been evaluated by computing the position of the robot with respect to the target on the base of feature points automatically identified on the picture of a real painting. The results obtained using the experimental tests in our laboratory are very encouraging for the experimentation in the real environment.

1 Introduction

Earth meanders often contain artistic treasures hardly accessible by visitors for the adverse structural characteristics of the zone and the exigency to preserve these treasures from deterioration and damages. In the south of Italy the walls of the archaeological cave named "Grotta dei Cervi" ("Stag's Cave" *n.d.r*) are rich of red and black paints of hunting, stags and men realized with red ochre and guano of bats. They are among the most remarkable paintings of the European prehistory. In Fig. 1 is illustrated a map of a corridor of the cave hosting four zones, indicated on the map, particularly interesting for the presence of prehistoric wall paintings.

The access to the cave is now denied to unauthorized persons since the exploration of the cave is very hard and a visitor could bring polluting elements in this singular environment. So, the application of a technological solution can allow the remote fruition of the archaeological site with its artistic and historic treasures and, at the same time, can preserve the cave and the safety of the users.

[*] This work has been developed under project TECSIS supported by grant MIUR n°12905, decreto MIUR 1188, 02.08.2002.

F. Roli and S. Vitulano (Eds.): ICIAP 2005, LNCS 3617, pp. 107 – 114, 2005.

Among several solution, it is currently under development a rover with a sophisticated equipment able to autonomously navigate in the cave and to control the instruments needed for vision, measurements and characterization of the site.

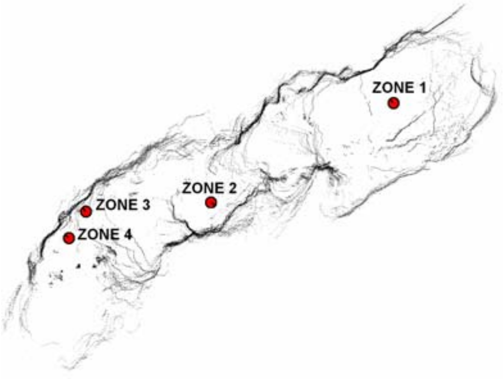


Fig. 1. Map of a corridor of Grotta dei Cervi. Four zones particularly interesting for the presence of prehistoric pictures are indicated

The exigency to preserve the site, avoiding modification of the environment with the installation of external elements, makes necessary to use natural landmarks, in this case wall paintings, as reference points for the rover navigation. A relevant problem in the autonomous navigation of a vehicle is self-localization. This paper deals with the automatic self-positioning of the camera mounted on the vehicle with respect to a predefined target, having the whole target image in the camera field of view.

Several methods to estimate the camera pose can be found in literature [1, 2]. Most of them use landmarks for self-localization [3]. Depending on the knowledge of the 3D target model, model-based [4] or model-free [5] approaches can be used. The POSIT algorithm (Pose from Orthography and Scaling with Iterations), proposed by DeMenthon and Davis [4], is based on geometric considerations that combine perspective and scaled orthographic projections applied to feature points. Being based on a linearized camera model, it is immediate, efficient, works for scenes with a few number of features and has a very low computational cost. It is described in Section 2.

Since the accuracy of 2D features strongly influences the convergence of the method, the pin-hole camera model can not be used for the image representation of the 3D reality. For significant lens distortion, it would be necessary a calibration phase. To avoid this phase, we have modified the original version of POSIT algorithm to use it in real-time applications even with uncalibrated images [6, 7]. The new method, described in Section 3, performs better than the original one (Section 4).

After the validation of the method using a grid target, we have tested the applicability of this method in a complex natural environment such as the archaeological site of Grotta dei Cervi. The only information about this environment come from some photographs of the cave. For this reason, to obtain a 3D target we have created a 3D structure projecting on two orthogonal planes the photograph reproducing some wall paintings that supply a set of feature points.

For the validation of the final results, we have acquired a set of images to evaluate the camera pose for each point of view. The correspondence between the reference 2D image and the current 2D image is obtained by the Sum of the Absolute Differences (SAD) matching metric introduced in Section 5, through which it is possible to find unambiguous and trackable landmarks in an image. Experimental results relative to the application of the modified POSIT algorithm using a target reproducing prehistoric wall paintings are presented and discussed in Section 6.

2 POSIT Algorithm

The original version of the POSIT algorithm finds the pose of an object from a single image, on the base of the 3D target model in the object coordinate frame of reference. Necessary conditions are the extraction of at least four no-coplanar points and the matching of the extracted features with the corresponding model points.

In a pin-hole camera model, as shown in Fig. 2, a target object, with feature points M_i, is positioned in the camera field of view. The focal length f, the M_i coordinates in the object coordinate frame of reference point M_0 and the image points m_i (x_i,y_i) are all known. The goal is to compute the rotation matrix **R** and the translation vector **T** of the object coordinate frame with respect to the camera coordinate frame.

R is the matrix whose rows are the coordinates of the unit vectors **i**, **j** and **k** of the camera coordinate frame in the object coordinate frame. **T** is the vector $\mathbf{OM_0}$. If M_0 has been chosen to be a visible feature point for which the image is a point m_0, **T** is aligned with vector $\mathbf{Om_0}$ and is equal to $Z_0\mathbf{Om_0}/f$.

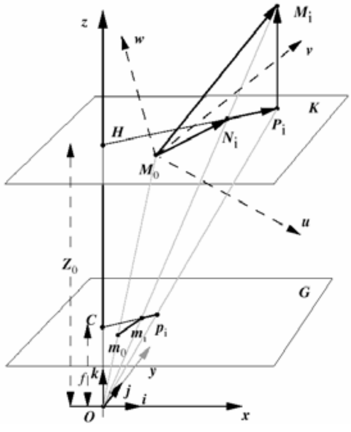


Fig. 2. SOP (p_i) and PP (m_i) of an object point M_i and a reference point M_0

The Scaled Orthographic Projection (SOP) is an approximation of the true Perspective Projection (PP) where all M_i points have the same depth Z_0 as the reference point M_0. If we define the scaling factor of the SOP as the ratio $s=f/Z_0$, the full translation vector $\mathbf{OM_0}$ can be expressed in the following way:

$$\mathbf{OM_0} = \mathbf{Om_0} / \mathrm{s} \tag{1}$$

The vector $\mathbf{M_0M_i}$ can be expressed in the following way:

$$\mathbf{M_0M_i} \cdot \mathbf{I} = x_i(1+\varepsilon_i) - x_0 \qquad (2)$$
$$\mathbf{M_0M_i} \cdot \mathbf{J} = y_i(1+\varepsilon_i) - y_0$$

where ε_i, **I** and **J** are defined as $\varepsilon_i = (\mathbf{M_0M_i} \cdot \mathbf{k})/Z_0$, **I**=f **i**/$Z_0$ and **J**=f **j**/Z_0.

If ε_i is known, equations (2) provide a linear system of equations in which the only unknowns are **I** and **J**. Once **I** and **J** have been computed, the scaling factors $s_1 = (\mathbf{I} \cdot \mathbf{I})^{1/2}$, $s_2 = (\mathbf{J} \cdot \mathbf{J})^{1/2}$ and $s = (s_1 + s_2)/2$ are obtained and the unit vectors **i** and **j** derive from the normalization of **I** and **J**.

The POS algorithm finds the pose for which the point M_i has, as SOP, the image point p_i. (ε_i equal to zero). This solution is an approximation because ε_i is not exact. Once **i** and **j** have been computed, a more exact ε_i can be computed in the POSIT algorithm and the equations can be solved again with these better values. By iterating these steps, the method converges to an accurate SOP image and an accurate pose.

3 Pose Recovery with Uncalibrated Images

Equation (1) is valid for a camera perspective approximation, on the basis of the theoretical hypotheses of the POSIT algorithm, but, using uncalibrated real images, this assumption is not valid. The use of a single scaling factor in the definition of **T** is valid only if the image plane has no distortion. The idea on the base of the new version of the method is to use different scaling factors in the definition of **T**.

The three scaling factors s_1, s_2, and s all converge to a single value using the pin hole camera model. This does not happen in the real case for the presence of lens distortion. Since the scaling factors s_1 and s_2 are derived from the normalization of the unit vectors **i** and **j** of the image plane, it is possible to use these scaling factors obtaining a "scaling factor vector" $\mathbf{s_V}$. In this case, **T** can be expressed as follows:

$$\mathbf{T} = \lfloor x_0/s_x \quad y_0/s_y \quad f/s_z \rfloor \qquad \mathbf{(3)}$$

where $\mathbf{s}_\mathrm{V} = \lfloor s_x \quad s_y \quad s_z \rfloor = \lfloor s_1 \quad s_2 \quad s \rfloor$.

4 Comparison Between POSIT and Modified POSIT Algorithm

The two versions of the POSIT algorithm have been compared using a 3D grid target, shown in Fig. 3.a, and a Sony DFW-SX900 camera. The extraction of the feature points (black points) has been performed using the Harris corner detection method [8]. For the validation of the final results, nine images have been acquired moving the camera of 10 cm for each acquisition over a grid of size 20x20 [cm] at 155 cm of height. The camera orientation has been set so that the whole target was in the camera field of view. The target was contained at the distance of 98cm from the centre of the observation grid. In the modified version only the x and the y components of **T** have a new formulation. To obtain the reference values for each position, the method of Tsai with an iterative optimization of Lenvenberg-Marquardt [9] has been used.

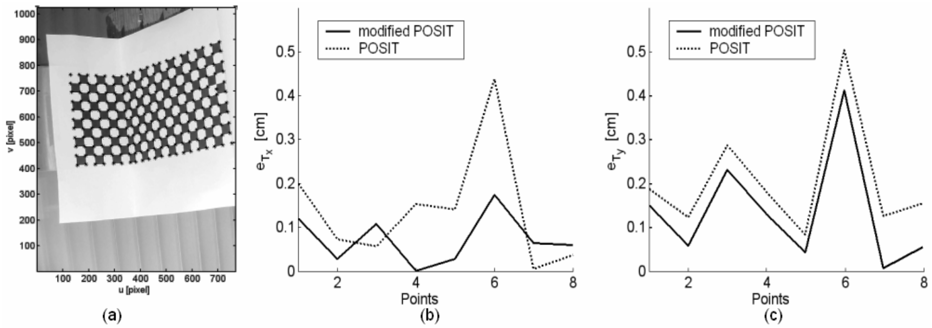


Fig. 3. Grid target (a) and comparison between e_{Tx} (b) and e_{Ty} (c) for the POSIT algorithm and its modified version

Table 1. Pose parameter errors, for the modified version of POSIT, when N is decreased

N	e_x [mm]	e_y [mm]	e_z [mm]	e_θ [°]	e_φ [°]	e_ϕ [°]
231	1.202	1.509	6.415	0.751	0.472	0.488
207	1.359	1.579	6.440	0.895	0.475	0.509
183	1.590	1.615	6.786	1.014	0.518	0.498
159	1.743	1.512	7.608	1.055	0.507	0.490
135	1.943	1.324	8.935	1.100	0.478	0.463
111	1.821	1.174	9.366	1.042	0.450	0.478
89	1.755	1.021	9.932	0.938	0.380	0.444
67	1.577	0.780	10.654	0.907	0.364	0.466
45	1.600	0.543	11.807	0.867	0.315	0.417
22	1.048	0.162	13.825	0.369	0.064	0.284
4	1.100	0.757	16.677	0.132	0.343	0.793

Figures 3.b and 3.c shows the comparison between the x and the y component errors of **T**, indicated with e_{Tx} and e_{Ty}, for the POSIT algorithm and its modified version. It can be observed the improvement of the modified version with respect to the original one.

To avoid calibration patterns, since the convergence of the algorithm requires the extraction of at least four no-coplanar feature points, we have analyzed the pose errors in the central point of the measure grid for the modified version of POSIT (Table 1), when the number of the considered feature points N decreases.

In this case it is possible to use a target with less geometrical constraints. The errors e_i on each pose parameter increase when N decreases. Nonetheless, even in the case of only four no-coplanar points, the difference with respect to the use of a larger N is less than 2 cm.

5 Correspondences Between 2D Images

The application of the modified POSIT requires the knowledge of the 3D target model and the extraction of at least four no-coplanar feature points. On the other hand, the only available information about Grotta dei Cervi are some photographs, starting from which we have obtained a 3D target object by projecting on two orthogonal planes the photograph showed in Fig. 4.

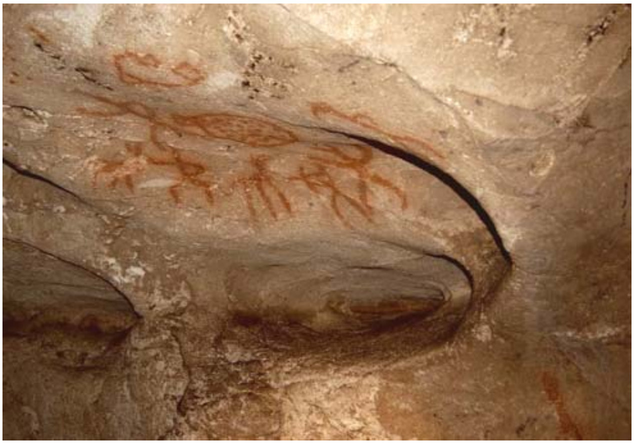
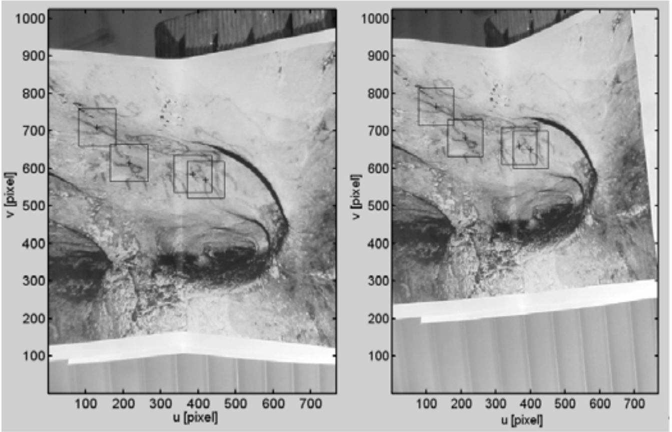


Fig. 4. Photograph of some wall paintings in the Grotta dei Cervi

This photograph shows a wall rich of prehistoric paintings, corresponding to zone 4 indicated in Fig. 1, suitable for the extraction of the necessary features.



(a) (b)

Fig. 5. Example of feature matching between two images

The first image, acquired from the central point of the measure grid described in the previous section, has been used as reference image and the coordinates of the image points corresponding to the 3D model points have been evaluated on this image plane. To extract the right feature points from the other 2D images, corresponding to the feature points of the 3D model, we have used the SAD matching metric that uses the similarity of windows around the candidate points [10]. With this technique, it is possible to find feature points in an image which are trackable and, at the same time, may not be confused with other possible points.

Fig. 5 shows an example of the matching between the reference image (Fig. 5.a) and another image acquired from another point of the observation grid (Fig. 5.b). The cross markers indicate the matched feature points applying the SAD technique and the squares indicate the windows around each point. The size of the window directly influences the algorithm performance, in terms of results accuracy and computing time, since it delimit the search area for the matchings. We have experimentally set the fitting value for the square size to 101 pixels. As can be observed from the figure, the feature points are extracted from the wall paintings of the cave.

6 Experimental Results

After the validation of the method we have tested the applicability of the modified POSIT for the camera self-localization in a prehistoric cave. Even for this test, the method of Tsai has been used to obtain the reference value for each pose. Once the reference pose values have been estimated, the 3D grid target has been substituted with the 3D target obtained by deforming the photograph of the cave. In this way, it is possible to compare the pose values obtained with the modified POSIT with the reference values obtained with the Tsai method. The experimental results are shown in Table 2. In Table 3 the average μ and the standard deviation σ of the error on each component of the camera pose computed over the set of 9 points of view are shown.

Table 2. Pose parameter errors for each measure point

Point	e_x [mm]	e_y [mm]	e_z [mm]	e_θ [°]	e_φ [°]	e_ϕ [°]
1	0.15	0.81	2.79	1.83	5.57	4.86
2	1.84	0.51	1.44	1.80	5.87	6.01
3	0.25	2.18	10.02	3.43	4.29	0.39
4	0.38	0.57	5.62	2.38	5.83	1.70
5	0.49	0.10	9.13	4.97	0.50	2.53
6	0.01	0.38	7.61	2.26	5.13	5.04
7	0.19	1.06	2.69	1.75	4.76	1.76
8	1.59	0.41	9.36	2.04	4.86	0.69
9	1.68	1.96	14.77	0.05	11.81	1.46

Table 3. Average and standard deviation of the error on each component of the camera pose

	x [mm]	y [mm]	z [mm]	θ [°]	φ [°]	ϕ [°]
μ±σ	0.70±0.71	0.83±0.64	5.56±3.56	2.43±1.11	4.25±1.94	2.71±2.05

These results are encouraging since the accuracy decreases with larger values of the ratio between the distance camera-target and the object size, as evaluated in a previous validation phase of the POSIT algorithm [6]. For the application in our laboratory, which uses a picture of the target, this ratio is big (it is equal to 25) but, in the real application it is expected to be lower due to the larger dimension of the target (the real paintings of the cave's walls).

7 Conclusions

The use of the modified POSIT for positioning a robot moving in the prehistoric cave of Grotta dei Cervi has been validated. The system provides a stream of images to enable the site fruition by users located outside. This images, and the vehicle position, allow the superposition of visual data to improve fruition and understanding of the site. The modified POSIT improves its performance by using a new formulation of the scaling factor and can work on uncalibrated images and on natural landmarks, automatically identified on the wall paintings to preserve the cave and avoid the installation of external elements. The extraction of the feature points from 2D images, corresponding to feature points on the 3D model of the wall hosting the paintings, has been accomplished using the SAD matching metric.

Experimental tests have been done to both assess the better performance of the modified POSIT with respect to the original formulation and to verify the possibility of reach a satisfactory positioning using natural features identified on the wall paintings. The results obtained in the experimental test in our laboratory are of good omen for the application of this method in this challenging real environment.

References

1. Malis, E.: Survey of vision-based robot control. ENSIETA European Naval Ship Design Short Course, Brest, France (2002)
2. Sugihara, K.: Some location problems for robot navigation using a single camera. Computer Vision Graphics and Image Processing, Vol. 42. (1988) 112-129
3. Betke, M., Gurvits, L.: Mobile Robot Localization Using Landmarks. IEEE Transactions On Robotics And Automation, Vol. 13 (2). (1997) 251-263
4. DeMenthon, D.F., Davis, L.S.: Model-based object pose in 25 lines of code. International Journal of Computer Vision, Vol. 15 (2). (1995) 123-141
5. Basri, R., Rivlin, E., Shimshoni, I.: Visual homing: surfing on the epipole. International Journal of Computer Vision, Vol. 33(2). (1999) 117-137
6. Gramegna, T., Venturino, L., Cicirelli, G., Attolico, G.: Visual servoing based positioning of a mobile robot. Proceedings of the 35th International Symposium on Robotics (2004)
7. Gramegna, T., Venturino, L., Cicirelli, G., Attolico, G., Distante, A.: Optimization of the POSIT algorithm for indoor autonomous navigation. Robotics and Autonomous Systems, Vol. 48(2-3). (2004) 145-162
8. Harris, C.G., Stephens, M.J.: A Combined Corner and Edge Detector. Proceedings of Fourth Alvey Vision Conference, (1988) 147-151
9. Salvi, J., Armangué, X., Batlle, J.: A Comparative Review of Camera Calibrating Methods with Accuracy Evaluation Pattern Recognition, Vol. 35 (7), (2002) 1617-1635
10. Konecny, G., Pape, D.: Correlation Techniques and Devices. Photogrammetric Engineering and Remote Sensing, (1981) 323-333