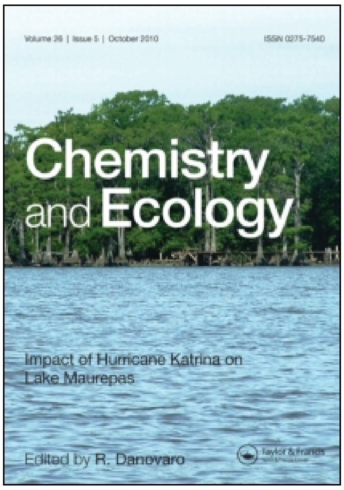

Chemistry and Ecology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713455114

TECSIS: Low-cost methodology to distinguish archaeological findings

R. Ialuna[a]; G. Gagliano[b]; D. Gravili[c]; D. Tegolo[b]
[a] SYNDIAL SPA, Milan, Italy [b] Department of Mathematics and Application, University of Palermo, Palermo, Italy [c] URS Italia SPA—Divisione CEOM, Milan, Italy

To cite this Article Ialuna, R. , Gagliano, G. , Gravili, D. and Tegolo, D.(2006) 'TECSIS: Low-cost methodology to distinguish archaeological findings', Chemistry and Ecology, 22: 4, S403 — S410

To link to this Article: DOI: 10.1080/02757540600738500

URL: http://dx.doi.org/10.1080/02757540600738500

Chemistry and Ecology
August 2006, Vol. 22 (Supplement 1), pp. S403–S410

TECSIS: Low-cost methodology to distinguish archaeological findings

R. IALUNA†, G. GAGLIANO‡, D. GRAVILI§ and D. TEGOLO*‡

†SYNDIAL SPA, Piazza Boldrini 1, 20097 S. Donato Milanese, Milan, Italy
‡University of Palermo, Department of Mathematics and Application,
Via Archirafi 34, 90123 Palermo, Italy
§URS Italia SPA—Divisione CEOM, Via Bertieri 4, 20146 Milan, Italy

(Received 16 March 2005; in final form 24 March 2006)

The automatic or semi-automatic research of archaeological findings includes several methodologies and algorithms of computer vision. The reconstruction of a scene is one of the key steps to meeting that challenge. This paper addresses a methodology for the reconstruction of underwater scenes with mosaicing techniques. The reconstruction of various scenes will be a video mosaic of sea-bottom landscapes starting from single video frames. The methodology is based on the evaluation of optic flow between frames, and its motion estimation has been evaluated on features extracted from the common areas of pairs of consecutive frames. This approach takes the motion model from a geometric projection framework. The estimation of camera movement is a second key point in the mosaicing problem. The methodology used should be robust enough to produce a good performance because of the high level of noise and turbulence involved in sea-bottom video acquisition. For this purpose, geometrical transformations have been used to map each frame into a unique big common reference frame with dimensions similar to that of the union of frames.

Keywords: Mosaicing; Geometrical transformation; Archaeological findings; Motion model

1. Introduction

TECSIS is a project financed by the Italian Ministry of Research (MIUR) to develop ‘Technologies and Intelligent Systems for the development of Archaeological Parks in Southern Italy’. Within this project, CEOM (Oceanographic Mediterranean Centre) and ENEA (Italian National Agency for New Technologies, Energy and the Environment) oversee the task of ‘mapping and recovering marine archaeological sites using geophysical surveys and the development of technologies and methods useful for their characterisation’. The research activity carried out for this task also involves the Sicilian Submarine Archaeological Authority for visual processing, and the University of Palermo for data processing.

The first ‘TECSIS’ campaign took place last year in the sea around the Egadi Islands (off the western coast of Sicily; see figure 1). This site was selected for historical and geographic

---

*Corresponding author. Email: tegolo@math.unipa.it

Chemistry and Ecology
ISSN 0275-7540 print/ISSN 1029-0370 online
http://www.tandf.co.uk/journals
DOI: 10.1080/02757540600738500

Figure 1. Survey area.

reasons, as it was the theatre of several battles between the Romans and the Phoenicians, and moreover was a central point along the trading routes in the Mediterranean Sea. The most important finding was the discovery of a wreck from an 11th-century Arabian ship that probably sank during a storm.

Within the project, a 'low-cost' methodology for characterizing the archaeological findings [1] has been developed. This methodology is based on a system using a mini-submarine remotely operated vehicle (ROV) equipped with two cameras in stereo configuration (figure 2), and a set of software tools for 2D reconstruction (mosaicing) [2] of the archaeological site (see figure 3).

To evaluate the system, we used a number of videos from an archaeological site acquired during a campaign that took place in August 2004. This article describes the TECSIS project, the methodology used, and the results obtained.

Figure 2. Achilles ROV.

Figure 3. Storage vessels (amphorae).

2. Material and methods

2.1 *Video mosaicing*

This section describes the methodology used for creating a mosaic from a sequence of video frames. In a nutshell, a video mosaic is the result of automatically stitching together the images of a video sequence of an object, to obtain a complete view of the object itself. This topic is related to motion estimation problems for exploring underwater sites; moreover, navigation and visualization of submerged wrecks and measuring of objects laid on the sea bottom can be treated in mosaic fashion.

Due to visibility limits (darkness, suspended material, and occlusion problems because of moving objects), the recording of images is often the only way to obtain a large view of submerged sites.

The motion estimation developed is based on the detection of features in common areas of pairs of consecutive images in the video sequence. In this approach, the motion model has been used in a geometric projection framework [3] consisting of:

- cameras oriented in such a way that the image plane is parallel to the sea bottom;
- sharing of the investigated area in longitudinal raw; and
- 0.7–1 m of height from the bottom.

The lack of visibility in the environment means that the videos can only be filmed a short distance from the object, and thus the co-planarity condition cannot be guaranteed. Since a number of methodologies identify points of interest under such conditions, we had to adjust some of these to suit the conditions.

Occlusions and acquisition noise are the major reasons for the complex search for corresponding points. Another aspect is the removal of outliers (i.e. wrong correspondences), as this phase has a great influence on the final results and requires a minimization step. A robust methodology, to minimize the bias produced by such techniques, is based on building a weight matrix which emphasizes the correct correspondence between features. Frames can be rendered and recorded, after having estimated the motion and found the homography parameters (images relationship) to build the final mosaic.

Figure 4. Camera.

2.2 *Acquisition and data analysis*

For practical reasons it is important to use low-cost tools for the exploration of underwater sites. In our case we considered the following set-up:

- an ROV, an underwater vehicle controlled by a remote operator (see figure 2);
- two video cameras for the acquisition of underwater images, mounted in a stereo configuration with a 30 cm baseline and parallel optical axes (i.e. the image planes are parallel) (see figure 4); the images are acquired with the image planes parallel to the sea-bottom; and
- an off-line data analysis software that has been developed under Matlab, producing a mosaicing from an underwater video sequence.

The choice of hosting two cameras in a stereo configuration on the ROV arises for two reasons: the first is to have two distinct mosaics (right and left), for classification and cataloguing; and the second is to identify their correct position in the 3D space, which is useful for measuring submerged objects.

The camera model used is the standard pinhole camera, which permits projective linear mapping of the 3D world in the image frame. The cameras have been calibrated using the Tzai [4] algorithm for the extraction of intrinsic (focal, radial and tangential distortion) and extrinsic (rotation and translation) parameters. The calibration pattern used (shown in figure 5) allows corners (images coordinates) to be extracted to include in the algorithm for calculating parameters.

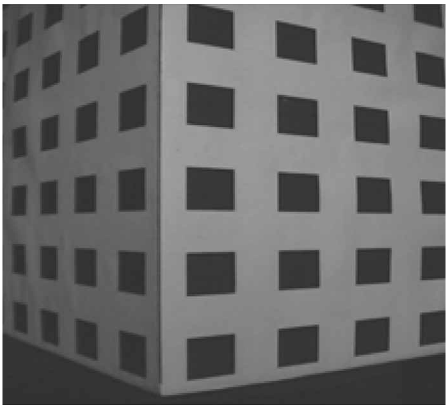

Figure 5. Calibration model.

Data were recorded onto mini-DV digital films at a speed of 25 fps in uncompressed PAL format and non-interlaced form. The videos were then stored on DVDs, maintaining the same format, and then analysed with Videomatch\Gromada Software (version 3.5.2).

3. Results and discussion

3.1 *Geometrical model and visual motion estimation*

The pinhole camera [2] used models linear mapping from the P^3 projective space onto the P^2 projective plane. This can be formally expressed in vector form as $\tilde{m} \doteq P\tilde{M}$, where $\tilde{M}$ is the position of the points in homogeneous coordinates, $\tilde{m}$ is the projection in the retinal plane, P is the matrix of prospective projection, and the symbol $\doteq$ represents the relationship of equality up to a factor scale.

Two different views of the same scene in the 3D space can be linked, under certain conditions, by a homography (linear transformation) of the P^2 space. It is represented by a 3×3 matrix which gives a one-to-one relationship between the corresponding points on the two images. Considering the $\tilde{u}$ and $\tilde{u}'$ coordinates of a couple of points of a scene acquired by two different points, it is then possible to determine a linear transformation given by the following:

$$\tilde{u}' \doteq H_{2\mathrm{D}}\tilde{u},$$

To determine the transformation $H_{2\mathrm{D}}$, four pairs of corresponding points between the two images are needed. Therefore, given a set of corresponding features, a homogeneous system of equations,

$$At = 0,$$

can be set up, where t is the vector of the unknown entries of the matrix $H_{2\mathrm{D}}$, and A is a matrix of size $2n \times 9$. The system can be solved through *singular value decomposition*.

For every linear transformation, with eight independent parameters, it is always possible if the system geometry and the typology of camera motion are known to reduce the number of parameters: table 1 lists the various models. Due to the fluctuations of the sea currents, and to the short distance from the sea bottom, the semi-rigid model is inadequate, even if it is more efficient from a computational point of view. We preferred to use the 'Projective Transformation' model to improve the qualitative performance of the results.

The first step for the creation of a mosaic is the estimation of the homography between two consecutive frames in a video sequence. For every image I_k, a set of features is extracted with the Harris corner detector [6]; then a tracking function, an implementation of the Lucas–Kanade algorithm [7] is performed to evaluate the next image I_{k+1}; then, the SVD matching [8] is used in order to reduce the number of matches used to estimate the best homography $H_{k,k+1}$ which relates the frames I_k and I_{k+1}. A RANSAC sampling method has been used to make the estimation of $H_{k,k+1}$ robust to wrong matches.

3.2 *Creation of the mosaic*

Registration and *rendering* are the main phases for the construction of a mosaic. *Registration* deals with the estimation of the 2D motion between pairs of frames, and the deformation of a single frame towards a global model of the sequence (temporal alignment); rendering builds the mosaic from a set of registered aligned images.

When we obtain the *frame-to-frame motion* parameters, transformations are used for a global model where frames are mapped in a unique, large common reference frame where

Table 1. Description of the models used for *image merging*, ordered for number of independent parameters, p.

Image model	Matrix form	p	Domain
Translation	$H_{2D} = \begin{bmatrix} t_1 & 0 & t_2 \\ 0 & t_1 & t_3 \\ 0 & 0 & t_1 \end{bmatrix}$	2	Image plane is parallel to planar scene; no rotation
Translation and zoom	$H_{2D} = \begin{bmatrix} t_1 & 0 & t_2 \\ 0 & t_1 & t_3 \\ 0 & 0 & t_4 \end{bmatrix}$	3	Same as above but with variable focal length
'Semi-Rigid'	$H_{2D} = \begin{bmatrix} t_1 & t_2 & t_3 \\ -t_2 & t_1 & t_4 \\ 0 & 0 & t_5 \end{bmatrix}$	4	Same as above but with rotation and scale along the image axes.
Affine Transformation	$H_{2D} = \begin{bmatrix} t_1 & t_2 & t_3 \\ t_4 & t_5 & t_6 \\ 0 & 0 & t_7 \end{bmatrix}$	6	Distant scene subtending a small field of view
Projective Transformation	$H_{2D} = \begin{bmatrix} t_1 & t_2 & t_3 \\ t_4 & t_5 & t_6 \\ t_7 & t_8 & t_9 \end{bmatrix}$	8	Most general planar transformation

dimensions are similar to the union of different frames. Let the transformation matrix between the reference frames and the first image frame be $H_{Rif,1}$; the global registration is then defined by a set of matrix transformations:

$$\{H_{\mathrm{Rif},k} : k = 1 \ldots N\} \text{ with } 2 \leq k \leq N$$

and

$$H_{\mathrm{Rif},k} = H_{\mathrm{Rif},1} \prod_{i=1}^{k-1} H_{i,i+1}.$$

A detailed analysis of the overlapped regions must be carried out after all the transformations are obtained. In these regions we can have different intensity values assigned to the same pixel, and this indeterminacy asks for the use of a reasonable policy for assigning a unique intensity value to each pixel. The chosen methodology, called the temporal operator, orders all the intensities by the time and choose the intensity value accordingly to a specific operator. The operator to be used varies: use-first, use-last, mean, and median operator, each making a different contribution to the output image.

Use-first and use-last, respectively, select first or last values in the contributions carrier ordered by time; the mean operator has the advantage of using all the contributions removing the temporal noise inside the acquisition. The median operator removes the temporal noise, too, but also eliminates the objects in motion (transient data) whose brightness is different for every frame; thus, it is extremely useful in such environments where brightness is strongly dependent on parameters that cannot be controlled.

4. Conclusion

The mosaic shown in figure 6 has been created from a video sequence without using any other information. The example shown has been obtained using 10 frames (five inter-frames between

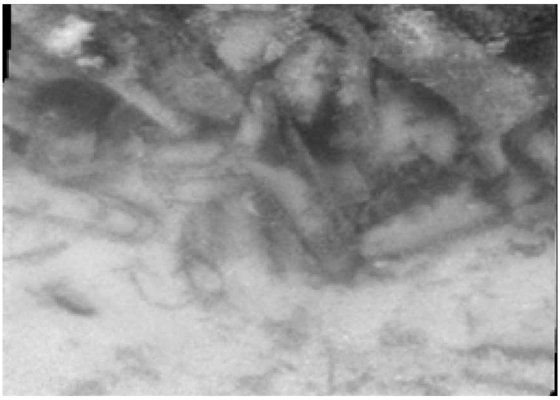

Figure 6. Mosaic based on 10 frames.

one image and the next), recorded using the above written model and using the media-operator. Note that the objects in the scene are not on the same plane in space.

The original images (see figure 7) present a large amount of noise which makes the extraction and matching of features very difficult, thus also reducing the accuracy of the estimation of homographies between frames. Even considering the fact that the objects are not planar, the software can extract motion parameters in a way such that the mosaic created presents a few small artefacts that are only visible to the human eye only by careful inspection.

Future studies should improve the quality of the result by mosaicing independently different parts of the images belonging to coplanar segments of the 3D scene.

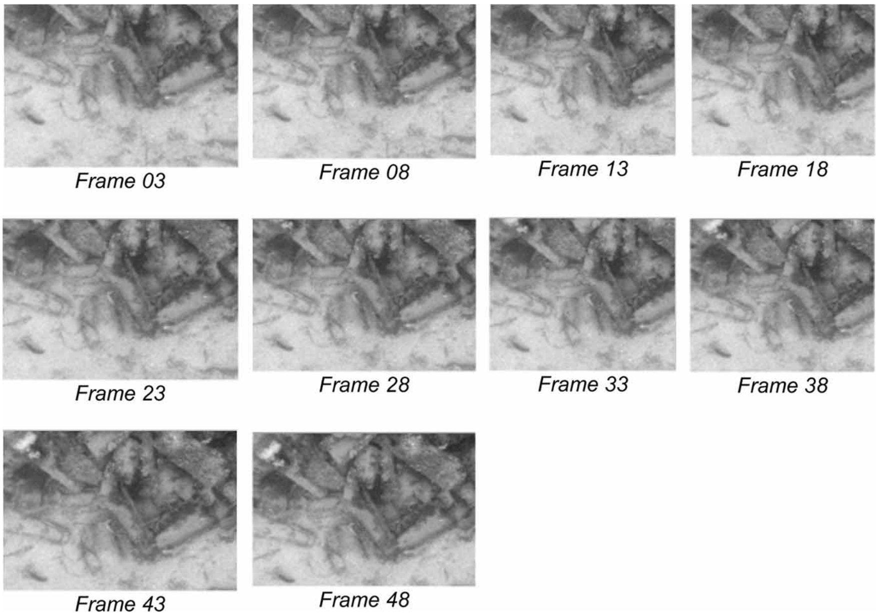

Figure 7. Frames used for the mosaic.

References

[1] R.Y. Tsai. A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV camera and lenses. *IEEE J. Robot. Autom.* **RA-3**, 323–344 (1987).

[2] E. Trucco, A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice-Hall, Upper Saddle River, NJ (1998).

[3] C. Harris, M. Stephens. A combined corner and edge detection. In *Proceedings of the Alvey Conference*, Manchester, UK, August, pp. 189–192 (1988).

[4] B. Lucas, T. Kanade. Optimal navigation by the method of differences. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, Vol. 2, Los Angeles, CA, August, pp. 981–984 (1985).

[5] M. Pilu. A direct method for stereo correspondences based on singular value decomposition. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 261–266 (1997).