ESTIMATE OF ENVIRONMENTS COMPOSED BY n PLANES BY DYNAMIC VISION TECHNIQUES

R. Frezza, M. Cossi, E. Bertuzzi
D.E.I., University of Padova
via Gradenigo 6/a, 35131 Padova, Italy
Tel. +39 49 8277825; fax: +39 49 8277826
e-mail: frezza@dei.unipd.it

A. Vettore
Istituto di Topografia, University of Padova
via Marzolo 9, 35131 Padova, Italy
Tel. +39 49 8275580, fax: +39 49 8275582
e-mail: vettoan@ux1.unipd.it

Commission VI, Working group 3

ABSTRACT

This work concerns the reconstruction of 3-dimensional environments composed by n planes from a video sequence. The goal includes the recovery of precise metric information. The results could be conceivably used for many purposes, for example for the photogrammetric reconstruction of buildings. The algorithm is based on the definition of an a-priori model of the probability distribution of the distance of the features from the walls. An unbiased estimate of a plane approximating a wall is obtained by maximizing the likelihood of the position of the features with respect to the wall. The scene is segmented in n planes applying statistical hypothesis testing techniques.

1. Introduction

The problem of estimating environments composed by n planes from a video sequence has been studied by many researchers see, for example, *[9, 5, 11, 12]*. In most of the previous work the estimates are the planes fitting, in a least square sense, the 3-D position of a number of point features. Most features, however, do not actually belong to the wall, but are either slightly in front of the wall (paintings, hanging coats, etc.) or behind the wall (dents, window frames, etc.). Rarely the distribution of the features is symmetric with respect to the wall and least square estimation techniques return biased estimates. One original aspect of our approach is that we assume an a-priori *probabilistic* model of the distribution of the point features. We estimate the unknown parameters of the probabilistic distribution by maximum-likelihood techniques, in order to obtain an unbiased estimate of the position and orientation of the wall. This procedure is typical of digital photogrammetry and is fundamental to obtain precise estimates in an automatic fashion. Using standard supervised photogrammetric techniques to compute the "ground-truth" we verify the precision of the estimates and compare it with those obtained by least square estimators.

Based on the a-priori model of the distribution of the features, it is also possible to apply statistical hypothesis testing techniques to segment a scene composed by n walls in n planes. Two techniques for segmentation are presented: the first is a statistical test on the hypothesis that the parameters of the distribution of the features are constant, the second is a test on the whiteness of the innovation of a Kalman filter which recursively updates the estimate of the parameters.

2. Motion reconstruction and estimate of a single plane

Many algorithms have been proposed for feature based motion and structure estimation from a monocular sequence of images, see *[10]* for a survey and a unifying perspective on the most successful algorithms presented by different researchers. For the application presented in this paper, we used a scheme specifically derived for features that are distributed on planes originally proposed by J. Weng, T. S. Huang and N. Ahuja *[14, 16]* and then, slightly modified, by O. Faugeras *[5]*. The output of the algorithm are the parameters of motion that describe the trajectory of the videocamera with respect to a reference frame fixed with its initial position and the 3-D position of the features points $P_k(x_k, y_k, z_k)$, $k = 1, \dots, N$ always w.r. t. the same frame. The next step in the interpretation of the images consists in a highest level analysis in which feature points are grouped together when belonging to the same plane. Associating more points to single objects allows for better estimates of their position. Clearly, this kind of processing needs a set of a-priori models describing those objects in the scene that one wants to identify and locate. H. Maitre *[9]*, for example, assumes that the feature points may belong to planes or quadratic surfaces such as

$$z = Ax + By + C$$

$$z = Ax^2 + By^2 + Cxy + Dx + Ey + F$$

He estimates with a least squares algorithms the parameters $\{A, B, C, D, E, F\}$ in order to locally describe the structure of the environment. Here, we propose a probabilistic model of the distribution of the features first applied by Cossi et al. in *[4]*. Details may also be found in *[3]*. The technique is based on the assumption that the features appearing in one image may:

1. belong to a wall;
2. be close to a wall, but not actually belong to it (they may, for example, belong to objects hanged to the wall);
3. be far from the wall (they may, for example, belong to another wall).

For each of these three classes one may introduce a probability density describing the position of the feature points relative to the wall. Let the wall be described by the plane of equation

$$ax + by + cz - d = 0 \qquad (1)$$

The distance r_k of a point of coordinates $P_k = (x_k, y_k, z_k)$ from the plane is given by

$$r_k = ax_k + by_k + cz_k - d.$$

Then, given the parameters $[a, b, c, d]$ of plane *(1)*, one can assume that the distance of the features belonging to class 1 from the wall can be described by the following probability density

$$f_1(r|[a, b, c, d]) = \delta(r)$$

while for the features belonging to class 2 the distance may be described by the density

$$f_2(r|[a, b, c, d]) = \frac{1}{\lambda} e^{\frac{r}{\lambda}}$$

and for those belonging to class 3, since they can be anywhere, the distance can be described by a uniform probability density

$$f_3(r|[a, b, c, d]) = 1/r', \qquad -r' \le r \le 0.$$

Choosing $r' > d$ one can also model dents in the wall.
The model must also take into account that the estimates of the position of the feature points P_k are noisy. If r_t, is the true distance from the wall of a feature point, we assume that its measured distance r is distributed normally according to

$$f(r|r_t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{r-r_t}{\sigma}\right)^2}.$$

We obtain, therefore, the following probabilistic model for the distribution of the feature points conditioned to the parameters of the plane *(1)*

$$\begin{aligned} f(r|[a, b, c, d]) = {} & k_1 f_{1,n}(r|[a, b, c, d]) + \\ & + k_2 f_{2,n}(r|[a, b, c, d]) + \\ & + k_3 f_{3,n}(r|[a, b, c, d]) \end{aligned} \qquad (2)$$

with $k_1 + k_2 + k_3 = 1$ and where, since the density of the sum of two independent r.v.'s is the convolution of the densities, the observations of the features belonging to class 1 are distributed according to

$$f_{1,n}(r|[a, b, c, d]) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2}$$

those belonging to class 2 according to

$$f_{2,n}(r|[a, b, c, d]) = \frac{1}{\lambda} \cdot e^{\frac{\sigma^2}{2\lambda^2}} \cdot e^{\frac{r}{\lambda}} \cdot Q\left(\frac{r}{\sigma} + \frac{\sigma}{\lambda}\right)$$

where Q is the complement of the Gaussian

$$Q(y) = \frac{1}{\sqrt{2\pi}} \int_y^{+\infty} e^{-\frac{y^2}{2}} dy = 1 - \Phi(y)$$

and those of class 3 according to

$$f_{3,n}(r|[a, b, c, d]) = \frac{\Phi\left(\frac{r}{\sigma}\right) - \Phi\left(\frac{r + r'}{\sigma}\right)}{r'}.$$

The unknown parameters of this probabilistic model, including those describing the plane, are $[a, b, c, d]$, λ, σ and r'. We estimate them by maximizing the likelihood *(2)* of the observations.
The estimates, denoted as $\phi_{M.L.}$, are smoothed by a Kalman filter integrating them in time. Let T and R be the translation and the rotation of the videocamera from one time instant t to the next $t+1$. The relationship between the coordinates P of one feature point w.r. t a

reference frame fixed with the camera at the two consequent time instant is

$$P_{t+1} = R \cdot (P_t - T).$$

If at time t the wall is described by equation $n_t^T P_t = d_t$, at the next time instant, w.r. t the camera frame, it will be described by

$$n_t^T R^T P_{t+1} + n_t^T T = d_t \rightarrow n_{t+1}^T P_{t+1} = d_{t+1}$$

where $n_{t+1}^T = R \cdot n_t$ and $d_{t+1} = d_t - n_t^T \cdot T$. It is, then, possible to apply a K.F. to the dynamic model

$$\begin{bmatrix} d(t+1) \\ n(t+1) \end{bmatrix} = \begin{bmatrix} 1 & -\hat{T} \\ 0 & \hat{R} \end{bmatrix} \cdot \begin{bmatrix} d(t) \\ n(t) \end{bmatrix} + \nu \tag{3}$$

$$\phi_{M.L.}(t+1) = \begin{bmatrix} d(t+1) \\ n(t+1) \end{bmatrix} + \eta \tag{4}$$

where the process noise ν is related to the noise in the estimates $\hat{T}$ and $\hat{R}$ of the motion parameters T and R while η is the error in the maximum likelihood estimates of the plane parameters whose variance may be approximated with the Cramèr-Rao bound. Some results of simulations on real image sequences are presented in the figures shown in the last page.

8. Scene segmentation

We propose two techniques to segment the scene in n planes. The first one consists in the application of a statistical test on the hypothesis that the parameters estimated up to time are likely to describe also the probability distribution of the observations taken at the next time step $t+1$.

A statistical hypothesis test *[7]* consists, in practice, in comparing the likelihood of the observations $\mathbf{x}$ with respect to two alternative hypothesis, $\mathbf{H}_0$ and $\mathbf{H}_1$, on the parameters of the a-posteriori probability distribution of $\mathbf{x}$. The test is designed on the values of type I and type II errors which consist, respectively, in rejecting $\mathbf{H}_0$ when it is true and accepting it when it is false. If the hypothesis $\mathbf{H}_0$ is simple *[7]*, then the probability of committing the error of type I can be computed as a function of the ratio between the likelihood of the observations given $\mathbf{H}_0$ and the maximum of the likelihood of the observations given H formulate the hypothesis $\mathbf{H}_1$ as follows

$$\mathbf{H}_0 \ : \ \phi_{M.L.}(t+1) = \phi_{M.L.}(t).$$

The test is then simply a comparison of the likelihood of the observations at time $t+1$ given $\phi_{M.L.}(t)$ with an appropriate threshold γ_α of type I is less than α.

The second technique proposed for segmentation of the scene is based on the theory of the detection of abrupt changes *[1]*. The test checks whiteness of the innovation of the Kalman filter updating the maximum likelihood estimates of the parameters based on the dynamic model *(4)*.

The performance of the two techniques is comparable and some results on real image sequences are shown in the figures.

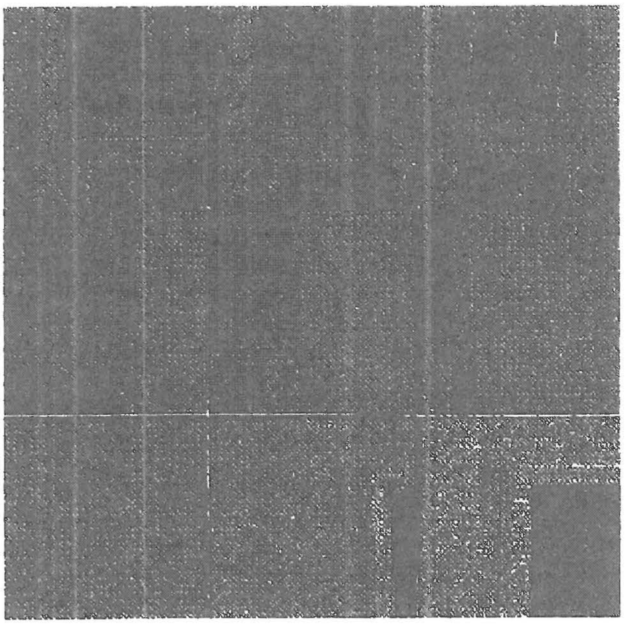

Figure 1: Sequence of images used to test the algorithms.

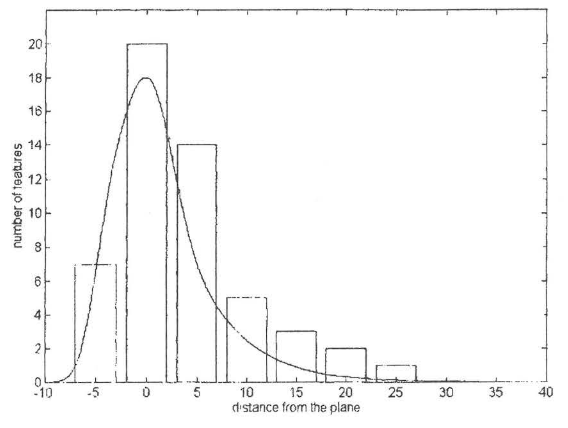

Figure 2: Comparison between the estimated probability density of the distance of the point features from the wall and the experimental data.

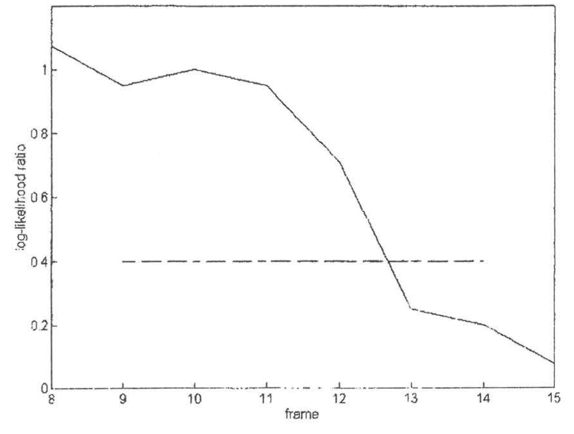

Figure 4: Value of the likelihood function on the image sequence used in the test and shown in figure *(1)*. In frame *13* it is decided that a new wall appeared in the scene.

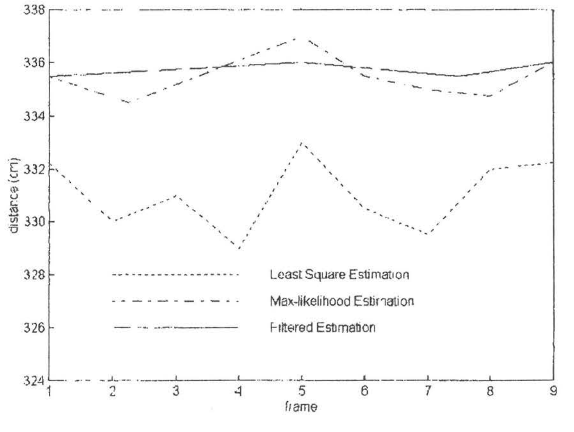

Figure 3: The least square estimate of the position of the wall shows an unacceptable bias. The Kalman filter smoothes the maximum likelihood estimate and returns an unbiased precise estimate.

Figure 5: Blow-up of the first frame in the sequence in which the segmentation algorithm decides that a new wall appeared in the scene.

REFERENCES

[1] M. Basseville and L V. Nikiforov. Detection of Ambient Changes: Theory and Application. Prentice-Hall, Englewood Cliffs, New Jersey, 1993.

[2] E. Bertuzzi. Misura automatica di ambienti composti da piu piani da sequenze di immagini. Tesi di Laurea, DEI Università, di Padova, Italy, 1997 (in Italian).

[3] M. Cossi. Stima ricorsiva di ambienti planari a tratti da sequenze di immagini. Tesi di Laurea,, DEI Università di Padova, Italy, 1994 (in Italian).

[4] M. Cossi, G. Cortelazzo and R. Frezza. Reconstruction of structure and texture of planar environments by dynamic vision techniques. *In Proc. Eusipco 96,* Trieste, Italy, Sep. 1996.
[5] O.D. Faugeras and D. Lustman. Motion and structure from motion in a piecewise planar environment. *Int. J. of Pattern Recognition and Artificial Intelligence,* 2(3):485 – 508, 1988.
[6] K. Kanatani. Computational projective geometry. *CVGIP: Image Understanding,* 54(3): 333-348, 1991.
[7] Kendall and Stuart. The Advanced Theory of Statistics, vol. 2, Charles Griffin & Co., fourth edition, 1979.
[8] S. Mann, R. W. Picard. 'Video Orbits': Characterizing the coordinate transformation between two images using the projective group. *Proc. ICCV 95,* 1995.
[9] H. Maitre and W. Luo. Using models to improve stereo reconstruction. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* PAMI-14(2):269 – 277, 1992.
[10] S. Soatto, P. Perona. Motion estimation from point features; unified view. *Proc. of the IEEE Int. Conf. on Image Processing* ICIP 95,0ct. 1995.
[11] M. Straforini, C. Coelho, and M. Campani. Extraction of vanishing points from images of indoor and outdoor scenes. Image and Vision Computing, 11(2):91 – 99, 1993.
[12] M. Straforini, C. Coelho, M. Campani, and V. Torre. The recovery and understanding of a line drawing from indoor scenes. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* PAMI-14(2):298 – 303, 1992.
[13] R. Szeliski. Image mosaicing for tele-reality applications. *Tech. Rep. CRL 94/2,* DEC, May 1994.
[14] J. Weng, N. Ahuja, T. S. Huang. Motion and structure from point correspondences with error estimation: planar surfaces. *IEEE Trans. on Signal Processing,* SP-39(12):2691 – 2717,1991.
[15] J. Weng, N. Ahuja, T. S. Huang. Motion and structure from two perspective views: algorithms, error. *IEEE Trans. on Pat. Anal. Mach. Intel.,* PAMI-11(5):451 – 476, 1993.
[16] J. Weng, N. Ahuja, T. S. Huang. Optimal motion and structure estimation. *IEEE Trans. on Pat. Anal. Mach. Intel.,* PAMI-15(9):864 – 884, 1993.