Optimal Surface Fusion

Peter R.J. North
School of Cognitive and Computing Sciences
University of Sussex
Janet: petern@cogs.susx.ac.uk

Abstract

This paper presents a general method for combining stereo surfaces using a Kalman filter. A measure of error in surface representation is suggested, and the work shows how a set of surfaces may be combined to give a single surface which minimises this measure. The analysis shows how a stochastic surface may be generated using stereo, and how errors in surface-to-surface registration may be modeled. The cases of multiple, mutually-occluding surfaces and unknown three-dimensional camera motion are considered. Performance is analysed using semi-artificial data. The results are important to multi-sensor fusion and automatic model generation.

The problem of estimating a single optimal surface from noisy measurements occurs in many vision and robotics applications [1, 2, 3, 4]. Here it is considered in the context of building a description of a complex object or environment using stereo reconstruction from many viewpoints [5]. A definition is offered of the optimal surface to represent a set of measured surfaces, and the paper shows how it may be found using the Kalman filter framework. Models of errors in stereo surface reconstruction derived from [6] and [7], and of surface-to-surface registration [8] are presented. Finally performance is analysed using artificial and real data.

Other authors have used Kalman filtering to incrementally combine visual measurements [1, 9]. This paper differs from previous work by representing uncertainty in surface location rather than features such as corners [9] and lines [1] and so relates most closely to [7]. However their work is extended to allow unknown three-dimensional camera motion and model multiple mutually-occluding surfaces. The results presented are of general interest to active vision, sensor fusion and automatic model generation.

1 PROBLEM FORMULATION

Consider N surfaces, $S_1 ... S_N$. We want to find the best representation of these surfaces by a single surface S_0 defined by a regular grid of spline control points relative to an origin O (figure 1).

We define the error in surface fit as follows —

BMVC 1991 doi:10.5244/C.5.21

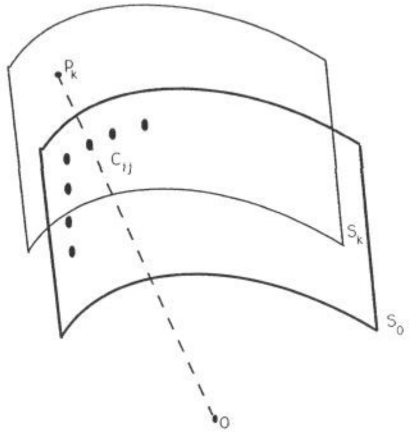

Figure 1: **Surfaces S_k provide measurements of model control points.**

For each control point c_{ij} representing S_0, draw a line through the point location and O. Find the intersection points p_k of this line with each of the surfaces S_k. The error in each control point is given by

$$E_{ij} = \sum_{k=1}^{N} \frac{|c_{ij} - p_k|^2}{\sigma_{p_k}^2}$$

where $\sigma_{p_k}^2$ is the variance of the position of each p_k.

We define the optimal solution as the surface S_0 which minimises this error summed over all the control points defining S_0. Having thus formulated the problem as a least-squares combination of noisy measurements, and assuming we can find values for the p_k and $\sigma_{p_k}^2$, we turn to the Kalman filtering framework to find the optimal solution.

2 OUTLINE OF SOLUTION

A description of the environment is built by incrementally combining surface estimates from multiple viewpoints. Using the Kalman filter framework, we define the *system model* as the best estimate of the surface visible from some given viewpoint. Surface estimates generated from stereo pairs are treated as *measurements* of this model. The transformation from each measurement coordinate frame to the model frame is given by a homogeneous matrix H, determined by surface-to-surface registration.

Both the model surface and measured surfaces are represented by regular grids of spline control points giving the inverse depth $d = \frac{1}{z}$ of the surface from its origin. A control point is free to move on a line passing through the control point and its origin. Each control point is modelled as a Gaussian distribution about its mean value with variance σ_d^2.

Each measured surface is integrated with the system model by intersecting lines through the model control points with the surface, to give a measured value for each model control point. An estimate of the variance of this measurement is calculated from the variance of the original stereo measurements and a sensitivity analysis of the registration process. A decision on whether the measurement relates to the model control point or to some other surface not represented in the model is taken by considering the difference between measured and predicted values relative to their positional uncertainty. Finally the system model is updated using Kalman filtering.

3 STEREO VARIANCE

The stereo algorithm considered here is based on Nishihara's [6]. This performs correlation matching using a coarse to fine strategy on an image pair convolved with a difference of Gaussians filter and thresholded. The autocorrelation surface of the processed image close to the origin approximates a cone [6]. Hence during stereo matching, sections through the cross-correlation surface along epipolar lines are expected to have the form —

$$\Psi(v) = av^2 + bv + c$$

where $\Psi(v)$ is the cross-correlation at disparity v. A parabola can be fitted from three correlation measurements allowing the peak correlation v_p to be determined with sub-pixel accuracy. An estimate [7] for the variance of the location of the peak is given by

$$\sigma_{v_p}^2 = \frac{2\sigma_n^2}{a}$$

where σ_n^2 is the variance of the image noise. Nishihara [6] estimates this as

$$\sigma_n^2 = \frac{w}{4r}$$

where w is the width of the central region of the difference of Gaussians convolution, and r is the radius of the the image patch correlated. Intuitively the confidence in of the peak disparity estimate increases with the "sharpness" of the correlation surface, giving more weight to measurements of highly textured regions parallel to the camera image planes.

Approximating the two cameras as parallel, with normalised focal length and camera separation, the peak disparity v_p, is related to the depth z by

$$v_p = 1/z$$

Hence, following [4, 7] we work with values $d = \frac{1}{z}$, with variance given by

$$\sigma_d^2 = \frac{w}{2ra}$$

4 REGISTRATION ERROR

A measured surface is related to the model by a coordinate transformation given by a rotation R, a translation T, and a scale factor λ. These must be found by a registration process prior to surface fusion. The parameters R, T, and λ are estimated by minimising the error vector

$$E_i = |\lambda R(u_i - T) - v_i|^2$$

for corresponding three-dimensional points u_i and v_i in the two coordinate frames. A fuller discussion is found in [8].

It is necessary to analyse how the errors in calculating the transformation parameters affect the errors in the transformed measurements. We can describe the coordinate transformation between the u_i and v_i by the homogeneous matrix H —

$$v_i = H u_i$$

The errors Δv_i in the v_i, resulting from perturbations $\Delta\lambda$, ΔR, and ΔT in the transformation parameters are approximated during registration by

$$v_i + \Delta v_i \approx H u_i + E_i$$

And so

$$\Delta v_i \approx E_i$$

Hence, in this work, the variances of subsequent transformed measurements due to registration error are approximated by the variance of the error vector E_i found during registration.

$$\sigma_v^2 \approx \sigma_E^2$$

5 SURFACE CLUSTERING

In the general case of multiple, mutually-occluding surfaces, there are two sources of error not modelled in the framework (figure 2). The first is caused by interpolation in the measured surface over depth discontinuities at object boundaries — on projection this results in the surface appearing much closer than the model surface.

The second is the possibility that the surface being integrated is not visible from the model viewpoint. This occurs when the measured surface occludes itself on projection, or when some surface already represented in the model is occluding. The self-occlusion problem can be solved by z-buffering during projection [5] but occlusion by other surfaces is more problematic.

The approach taken here, in common with [1, 2] is to use the covariance information derived previously to cluster the points onto single surfaces. We reject measurements which are far from the existing surface description relative to the certainty in the position of both the predicted and measured model

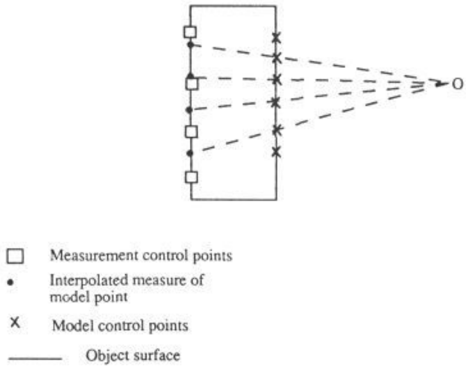

Figure 2: **Surface clustering.**

points.

In the results shown a measurement d of control point c is rejected if

$$|c-d|^2 > 2(\sigma_d^2 + \sigma_c^2)$$

The constant threshold term is somewhat arbitrary. A side effect is to provide further smoothing of the data, eliminating the effects of outlying points caused by, for example, ambiguous stereo matches, which are in any case not modelled well by Gaussian noise.

6 INTEGRATION

The measured surfaces are related to the model by interpolating values for each model control point where its line of positional uncertainty intersects with the surfaces. Interpolation is necessary since control points on the stereo surface will not generally project onto control points on the model. Additionally, the positional uncertainty for the measured and model control points lie on non-intersecting lines, and the uncertainty of the point of intersection of the surface with model control point directions has a bi-modal distribution. The problem is linearised after the fashion of [7] by approximating the uncertainty in interpolated stereo control points as co-linear with the model point uncertainty.

Hence if a measured control point with inverse depth d is transformed to a

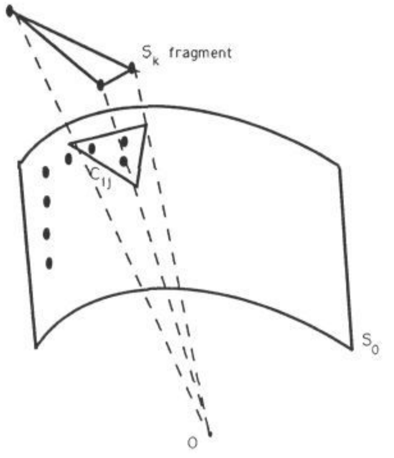

Figure 3: **Interpolation of measurements.**

value of d' in the model coordinate frame,

$$d' = \alpha d$$

then

$$\sigma_{d'}^2 \approx \alpha \sigma_d^2 + \sigma_E^2$$

where σ_E^2 is the variance of the registration error vector.

An *efficient implementation* of the surface fusion is as follows —

- Using a triangular tesselation of the measured surface (figure 3), sets of three points are projected onto the model coordinate frame. Model control points c_i^- intersecting this triangle are found by back-projection, and corresponding measured values for the inverse depth d_i and variance $\sigma_{d_i}^2$ are found using bi-linear interpolation.

- The Kalman filtering framework can now be used to find new estimates for the model control points c_i^+ and associated variances q_i^+ as follows;

  The Kalman gain is calculated as

  $$K_i = \frac{q_i^-}{q_i^- + \sigma_{d_i}^2}$$

  The new model control points are given by

  $$c_i^+ = c_i^- + K_i(d_i - c_i^-)$$

  and their variances given by

  $$q_i^+ = (1 - K_i)q_i^-$$

Figure 4: **Test image mapped onto surfaces.**

7 RESULTS

The method has been tested on three artificially generated sequences of stereo images. Each stereo pair is generated by mapping the poster (figure 4) onto a test surface and rendering it from two simulated viewpoints. Stereo analysis of each image pair gives a depth-map which is incrementally combined with previous measurements as discussed. However the transformation between successive frames is assumed to be known exactly.

The first sequence simulates eight stereo views moving towards a fronto-parallel plane. Graphs of the measured and predicted mean square error in $1/z$ are shown in figure 5(a). The error is seen to fall off particularly sharply since measurements closer to the surface are more accurate.

The second sequence (figure 5(b)) shows operation on the same stereo views moving *away* from the plane. The initial estimates are thus much more accurate. The final mean squared error has, as expected, the same value (approximately 0.02 of the simulated camera separation) for both image sequences.

The third image sequence simulates movement towards the sinusoidal surface shown in figure 6(a). The measured error (figure 6(b)) falls off more slowly in this case, perhaps because of the difficulty in reconstructing steeply sloping surfaces using stereo. The surfaces reconstructed from one (figure 6(c)) and five (figure 6(d)) stereo pairs are shown.

8 CONCLUSION

A method has been presented for incrementally combining stereo surfaces in the context of visual model generation. The surface fusion minimises the measure of error in surface representation proposed. The results on semi-artificial data appear very promising.

Using surfaces rather than features to build a model allows information derived from points matched within any stereo pair to be used, rather than those which can be tracked through a sequence.

The surface representation is appropriate for model generation and some applications, for example visualisation and tracking [10]. Other representations can be derived from it after model-building is complete. For example CAD models may be built by fitting primitives [2, 11] or octree models as outlined in [12]. It is more difficult to retrospectively fit a surface over three-dimensional feature locations since viewpoint occlusion information is lost. Further work is also of interest to combine surface descriptions across modalities.

ACKNOWLEDGEMENTS

I would like to thank David Young, David Hogg, Alistair Bray and Jim Stone for their contribution to this work, and DENI for their support.

References

[1] Ayache, N. and Faugeras, O.D. Building, registrating and fusing noisy visual maps. *First Int. Conf. on Computer Vision*, pages 73–82, 1987.

[2] Durrant-Whyte, H.F. *Integration, Coordination and Control of Multi-Sensor Robot Systems.* Kluwer Academic Publishers, 1988.

[3] Terzopoulos, D. Integrating visual information from multiple sources. In A.P. Pentland, editor, *From Pixels to Predicates.* Ablex Press, 1986.

[4] Grant, P. and Mowforth, P. Economical and cautious approaches to local path planning for a mobile robot. *Proc. of the AVC, Reading*, pages 297–300, 1989.

[5] North, P.R.J. Reconstruction of visual appearance. *Proc. of the BMVC*, pages 205–210, 1990.

[6] Nishihara, H. K. Practical real-time imaging stereo matcher. *Optical Engineering*, 23(5):536–545, 1984.

[7] Matthies, L., Kanade, T., and Szeliski, R. Kalman filter-based algorithms for estimating depth from image sequences. *Int. Journal of Computer Vision IJCV*, 3(3):209–238, 1989.

[8] North, P.R.J. Visual model generation by combining stereo surfaces. *CSRP 192, School of COGS, University of Sussex*, 1991.

[9] Charnley, D. and Blissett, R. Surface reconstruction from outdoor image sequences. *Proc. of the AVC, Manchester*, pages 153–158, 1988.

[10] Bray, A.J. Tracking curved objects by perspective inversion. *Proc. of the BMVC*, 1991.

[11] Grossman, P. COMPACT - a surface representation scheme. *Proc. of the AVC, Manchester*, pages 97 – 102, 1988.

[12] Connolly, C. I. Cumulative generation of octree models from range data. *Proc. of the Int. Conf. on Robotics*, pages 25 – 32, 1984.

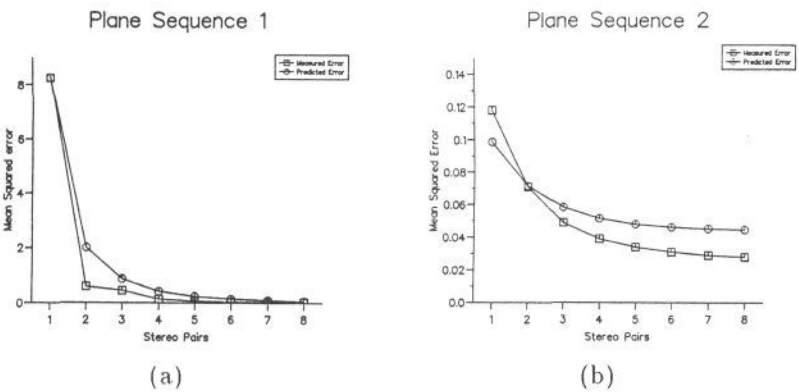

Figure 5: **Results for plane sequence.**

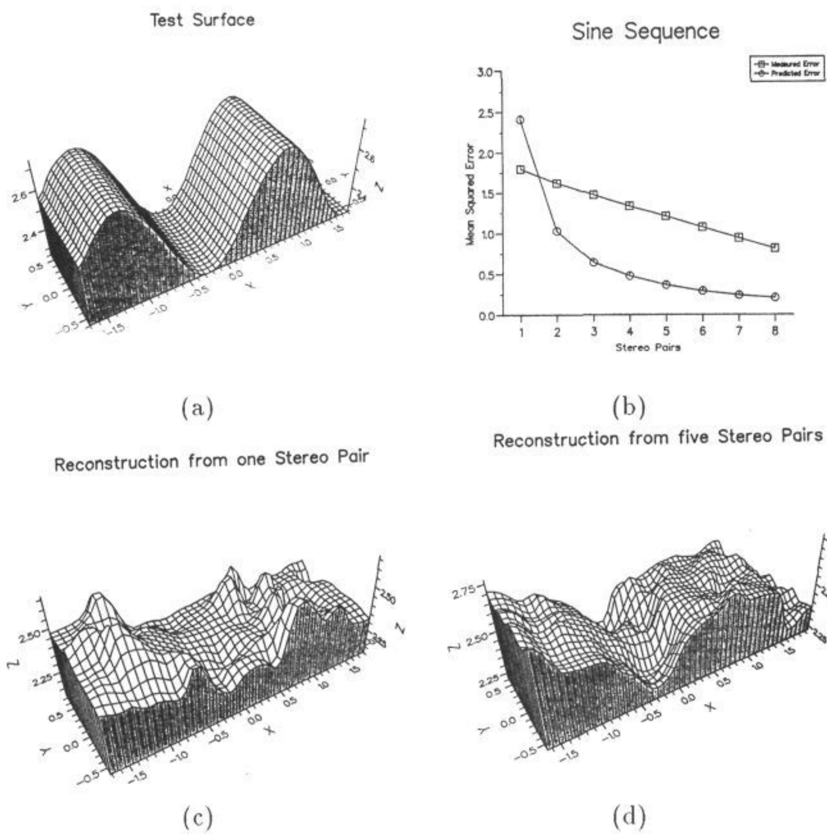

Figure 6: **Results for sine sequence.**