

Segmenting Correlation Stereo Range Images using Surface Elements

Don Murray and James J. Little
Department of Computer Science
University of British Columbia
Vancouver, British Columbia, Canada V6T 1Z4
donm@cs.ubc.ca little@cs.ubc.ca

Abstract

This paper describes methods for segmenting planar surfaces from noisy 3D data obtained from correlation stereo vision. We make use of local planar surface elements called patchlets. Patchlets have 3D position, orientation and size parameters. As well, they have positional confidence measures based on the stereo sensor model. Patchlet orientations (i.e., surface normals) provide important additional dimensionality that reduces the ambiguity of segmentation-by-clustering. Patchlet size allows the use of continuity or coverage constraints when segmenting bounded surfaces from depth images.

We use a region-growing approach to identify the number of surfaces that exist in a stereo image and obtain an initial estimate of the surface parameters. We refine segmentation using a maximum likelihood clustering approach that is optimised with Expectation-Maximisation. Confidence measures on the patchlet parameters allow proper weighting of patchlet contributions to the solution. We provide experimental results of the segmentation on complex outdoor scenes.

1. Introduction

The objective of this research is to segment surfaces from stereo images. The extraction of structure from point data is a topic that has many applications in robotics, graphics and computer vision. This work was motivated by our research in robot mapping and navigation [9]. Previously our visually guided robot has used stereo-vision-based voxel maps. Our goal is to segment high-level surfaces from the robot acquired stereo images in order to create more compact and accurate maps that can be used not only for robot navigation but also environment modeling and intelligent environment and object recognition.

Stereo vision is a method of range scanning that is becoming increasingly practical with the recent availability of inexpensive commercial systems. Stereo vision has many advantages, not the least of which is that with the steady increase in computational power of desktop systems, stereo

vision produces faster and better quality data with each passing year. However, stereo vision range data contains considerably more noise than data from structured light systems and laser range finders.

When we approached the task of segmenting surfaces from stereo vision, we made two important decisions. First, since stereo data is noisy, it was essential to model the range sensing errors appropriately and to use probabilistic methods that could include the sensor-based uncertainty. Secondly, we decided to use surface elements as the fundamental sensor primitive rather than 3D points.

Methods that incorporate surface normal information have a strong advantage over points-only approaches – namely that at any data element in the representation, surface properties such as orientation and surface coverage are known without the requirement of accessing and analysing other elements in the local neighbourhood.

We have presented a surface element data primitive based on stereo error modeling that we call the *patchlet* [8]. The patchlet is similar to the particles discussed in Section 2, however it is more tightly coupled to the sensor error model and includes uncertainty measures on its position and orientation parameters.

In this paper we develop a probabilistic approach to surface segmentation based on the patchlet data primitive and clustering in the plane parameter space. This method uses region-growing initially to determine the number of surfaces in the scene and surface parameters estimates. It then uses Expectation-Maximisation (EM) to refine the model to its maximum likelihood solution.

This paper is organised as follows. Section 2 presents a brief overview of the use of oriented particles for surface modeling. In Section 3 we provide an overview of the patchlet model. In Section 4 we layout our patchlet/surface comparison metrics. The model-selection and initial-guess solutions are presented in Section 5 and the EM refinement is given in Section 6. Results are given in Section 7, and finally in Section 8 we cover some of our plans for this research in the future.

2. Related work

The combination of position and orientation data together for higher level surface analysis has been a well accepted concept for many years. Sander proposed methods for inferring points along curves on a surface and parameterizing the local surface properties at these points using Darboux frames in [12]. These methods were later extended by Ferrie, Lagarde and Whaitte by aggregating the frames into continuous surface curves [2].

Szeliski and Tonnesen developed oriented particles for surface modeling and manipulation in [15]. Particles applied thin-plate constraints on their neighbours in a manner similar to active contour control points. This ensured smooth and continuous shapes even though they were based on a point representation.

In [5], Fua developed methods to combine stereo vision data from multiple view points and to use this data to generate oriented particles very similar to Szeliski's. Fua organised his stereo point data into voxels and generated the orientation of his particles by local plane fits to each sufficiently populated voxel. The particles also had a size parameter based on the size of the voxel.

In work inspired by Fua, Sara and Bajcsy developed their *fish scale* paradigm that is also based on interpreted stereo data [13]. In their methods, rather than fitting planes to the voxelated data they create a fuzzy set that represents the data. This set has a position and a covariance matrix. By considering the covariance matrix they classify a fuzzy set into a class of shape such as ball, plate or line.

Use of oriented particles with size for surface representation has also gained support in the graphics community. Pfister et al. use *surfels* as rendering primitives especially for very complex shapes that would require a large number of triangles to represent [10]. With QSplat, Rusinkiewicz and Levoy developed a multiresolution point-based rendering system that could adapt the rendering quality to the speed of the processor [11].

3. Patchlet model

Patchlets differ from the previous work in a few important ways. Correlation stereo identifies matching local surface patches, not points. Therefore, it is natural that the results of stereo should be reported in terms of surface patches rather than point clouds. Patchlets are created on an image-by-image basis rather than using point data combined from multiple points of view. As well, the uncertainty from the stereo-based 3D points propagates to uncertainty measures of the orientation and position parameters of the patchlets. This facilitates the use of probabilistic methods when interpreting patchlets; they can be properly weighted based on their parameter confidence. Finally, patchlet have a size parameter that is based on the projection of the disparity

pixel onto the sensed surface, rather than uniform division of space. The patchlet parameters are illustrated in Figure 1.

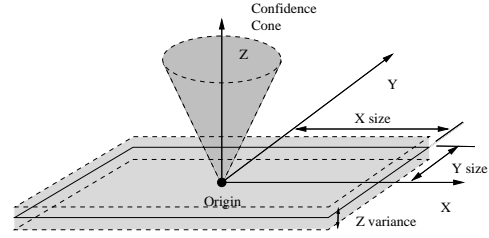


Figure 1: The Patchlet model

The patchlet parameters are:

X_P the 3D position of the patchlet centre in the world coordinate frame;

ψ_P the orientation or normal vector;

T_W^P the 4×4 homogenous transform that defines the patchlet's local coordinate system;

H_P height, or size parameter in the local coordinate Y axis;

W_P width, or size parameter in the local coordinate X axis;

Λ_P the positional variance in the direction of the normal;

κ_P the spherical confidence in the normal direction.

The positional confidence, Λ_P , is represented with a Gaussian variance in the direction of the normal vector, ψ_P . The spherical confidence κ_P is based on the Fisher spherical probability distribution [4]. This distribution is similar to a Gaussian but designed for a spherical space. The probability density function (PDF) of the Fisher model is

$$P(x|\mu, \kappa) = \frac{\kappa e^{\kappa \cos(\phi)}}{4\pi \sinh(\kappa)} \quad (1)$$

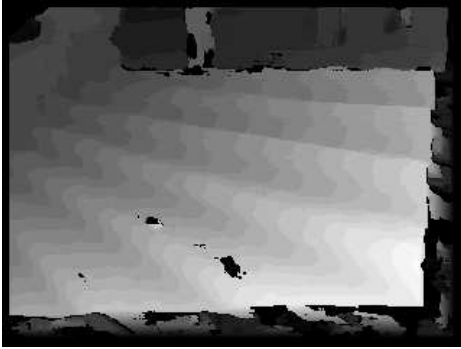
In this equation, ϕ is the angular difference between the mean direction and a sample orientation. κ is the *concentration* parameter. The higher the concentration parameter, the more tightly clustered the distribution. One can think of the concentration parameter as the inverse of standard deviation.

Patchlet orientation is determined by a local best-fit plane to a neighbourhood of disparity pixels in the disparity image. The patchlet size is the disparity pixel projected onto this planar estimate. This means the size depends greatly on the distance from the camera and the obliqueness of the detected surface. However, it does ensure that patchlets "tile" a surface without gaps. For a complete description of patchlets and the details of patchlet construction, see [8].

Figure 2(a) shows the reference image of an outdoor stereo image obtained with a Point Grey Research Digidlops

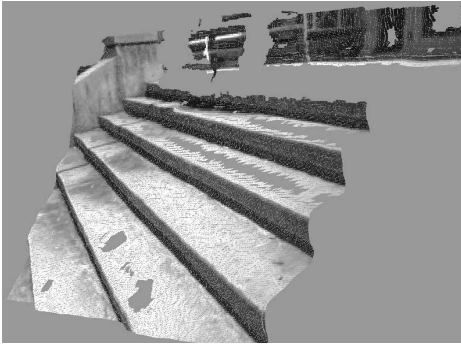


(a) Reference image

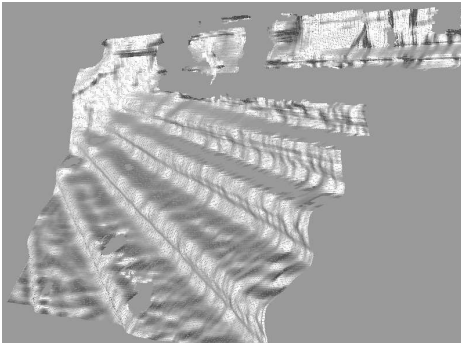


(b) Disparity image

Figure 2: Stereo scene and disparity image of outdoor steps



(a) Colored with reference image intensity



(b) White shaded

Figure 3: Patchlets generated from scene in Figure 2 and rendered from a novel viewpoint

trinocular stereo rig. Figure 2(b) shows the disparity image generated with their commercial stereo matching software. In the disparity image, the brighter a pixel is in the image the closer it is to the camera.

In Figure 3(a) we show a patchlet cloud generated from the disparity image in Figure 2. Each patchlet is coloured by the greyscale value of the reference image. Figure 3(b) shows the same patchlet cloud only this time coloured white with Gouraud shading. While the overall structure is present, clearly there is a fair amount of error between the patchlets and the true surfaces of the scene due to stereo errors.

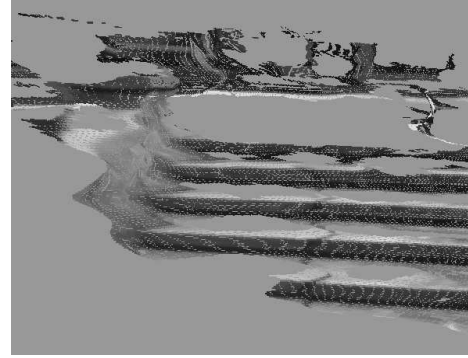


Figure 4: Another viewpoint on the patchlet cloud. This region is 7m away from the camera. Considerable 3D smoothing is observed due to the high distance from the camera.

Figure 4 illustrates even more how inaccurate the data can be. Here we see a view of the far balustrade of the staircase. This balustrade is over 7m from the camera. Since accuracy diminishes with distance, the disparity values are much less certain at this range. In addition, the stereo mask projects to a larger physical area. The smoothing from the stereo mask is therefore more prominent. We can see that the vertical surface of the balustrade is smoothed considerably with the horizontal surfaces of the steps.

Additionally, we can see in Figure 3(b) that there is a systematic noise pattern in the 3D data similar to a sine wave. This noise source has already been corrected as suggested by Shimizu and Okutomi [14]. However, although their approach reduces the phenomenon, it does not remove it completely.

4. Surface segmentation

The contribution of this paper is the development of a probabilistic framework for segmenting planar surfaces from patchlet clouds. Our approach to this problem is to find surfaces by clustering patchlets that are tightly grouped in plane parameter space where each surface in the scene will be identified as a single cluster. Since we have a combina-

tion of surface models to fit in an optimal manner, we represent this system as a mixture model, where each submodel in the mixture is a single surface.

The three main tasks in the segmentation-by-clustering problem are:

- Model selection – determining how many surfaces there are.
- Data association – determining for each point, the surface from which it was sensed.
- Model parameter estimation – determining the surface parameters

If the model selection and data association problem are solved, the parameter estimation is simply a best-fit of each model to its associated data. Therefore, in practice it is the model selection and data association problems that are difficult.

In our approach, we divide the segmentation process into two steps. The first step is a region-growing method based on RANSAC. It estimates the number of surfaces in the scene (thereby solving the model selection problem) and obtains reasonable initial estimates of their positions. The second step is to refine these surfaces to their optimal values using a maximum likelihood formulation and Expectation-Maximisation (EM). EM is well-suited for probability maximisation for mixture models as it solves the parameter estimation and data association problems simultaneously. This method does not require any *a priori* organization of the patchlets and consequently can work with combined data sets from multiple registered stereo views.

4.1. Surface model

Our surface model is identical to that of the patchlet (see Section 3) except in terms of scale. The surface parameters are identified with the S subscript to differentiate them from patchlet parameters. They are, therefore: X_S – position; ψ_S – orientation; T_W^S – coordinate system transform; (H_S, W_S) – height and width; Λ_S – positional variance; and κ_S – orientation confidence.

In order to use a probabilistic framework we need the relationship between the surface and patchlet structures. We construct the probability of a patchlet given a generative surface as:

$$P(X|S) = P_x(X|S)P_\psi(X|S)P_B(X|S) \quad (2)$$

where P_x is the Gaussian positional probability model, P_ψ is the orientation or Fisher probability model and P_B is the bounded surface probability model. These partial probabilities are given below.

$$P_x(X|S) = \frac{1}{\sqrt{2\pi\Lambda}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\Lambda}} \quad (3)$$

where

$$x = X_P \cdot \psi_S \quad (4)$$

$$\mu = X_S \cdot \psi_S \quad (5)$$

$$\Lambda = \frac{\Lambda_S \Lambda_P}{\Lambda_S + \Lambda_P} \quad (6)$$

And

$$P_\psi(X|S) = \frac{\kappa e^{\kappa \psi_S \cdot \psi_P}}{4\pi \sinh(\kappa)}$$

where

$$\kappa = \frac{\kappa_S \kappa_P}{\sqrt{\kappa_S^2 + \kappa_P^2}} \quad (7)$$

As shown in (6), Λ is the combined uncertainty of both the surface and the patchlet. Similarly κ is the combined spherical uncertainty of the surface and patchlet orientations. $(X_P \cdot \psi_S - X_S \cdot \psi_S)$ is the distance of the patchlet off the surface plane and $(\psi_S \cdot \psi_P)$ is the cosine of the angular difference between the normal vectors.

$P_B(X|S)$ models the effect of the surface boundaries. In early experiments we used unbounded planes to model our surfaces and encountered considerable corruption of model estimates due to outliers. With unbounded planes, it is common that some of the planes will pass close to some of the outliers. Simply, any slice through the 3D data may pass near some spurious points. This problem encouraged us to use bounded planes for our surface models. With a bounded surface model, if a point does not project near or within the surface boundaries, it will not be associated with the surface and cannot corrupt the surface estimation.

Our heuristic for the bounded surface probability is

$$P_s(X|S) = \begin{cases} 1 & \text{if } \hat{X}_P \text{ inside } S \\ 1 - \frac{d(\hat{X}_P, S)}{B} & \text{if } 0 < d(\hat{X}_P, S) \leq B \\ 0 & \text{if } d(\hat{X}_P, S) > B \end{cases}$$

where \hat{X}_P is the perpendicular projection of X_P onto S , $d(\hat{X}_P, S)$ is the distance of \hat{X}_P outside of the bounded area of S and B is a user specified threshold. With this model, if a patchlet projects to within the surface, there is no penalty applied to the overall probability score. If it falls within a threshold outside of the surface boundary, a penalty is applied that is linearly proportional to the distance away from the surface, and if the patchlet is beyond the boundary by more than the threshold, it is completely unrelated to the surface. This encourages locality of patchlets assigned to surfaces. For our experiments, the threshold B was 30 cm.

5. RANSAC-based initialisation

Although EM is an excellent technique for maximising a mixture model probability, it is still local. It must be initialised close enough to the global optimum so that, by iteratively improving the probability, it will not be caught in

a local maximum. For this reason we developed a method that obtains fairly good initial estimates for surface positions and also solves the model selection problem.

The method is a greedy algorithm. It works by finding the “best”, or most supported, surface in the data set first and removing the patchlets that support this surface from the set. It iterates until it cannot find a good surface in the remaining patchlet set and then terminates. More explicitly, the algorithm proceeds as follows:

0. set $i = 1$
1. find surface S_i with most support from data set \mathbf{X}
2. find subset of data, \mathbf{X}_i associated with S_i
3. **if** \mathbf{X}_i is smaller than threshold
 end
4. remove \mathbf{X}_i from \mathbf{X}
5. save S_i
6. $i = i + 1$
7. **goto** 1.

While this algorithm appears simple and straightforward, the difficulty lies in step 1, finding the surface with most support from the data set. This problem is not trivial. To solve it we adapted the Random Sample Consensus algorithm (RANSAC) [3]. The RANSAC algorithm concept is to randomly sample the minimum data necessary to determine a solution to a problem, and then evaluate how much “consensus” this solution generates with the remaining data. Many random samples are selected, giving a high probability that one of the random samples will generate a solution with a high level of agreement from the other data.

Our modification of this algorithm is as follows. Our minimum sample set is a single patchlet. A single patchlet gives an immediate estimate of a planar position and orientation. We then use region growing within the patchlet cloud. Each patchlet that abuts the current candidate patchlet set is tested to see if it agrees sufficiently to the initial sample. If its Mahalanobis distance from the candidate plane is within two standard deviations, then it is accepted into the candidate patchlet set. This is done several times (in our experiments, 100 times) and the candidate surface with the most agreeing patchlets is ranked the best surface. This algorithm avoids the problem of explicitly estimating planar boundaries, as the connectivity requirement of the region growing enforces contiguity of the patchlet set.

The Mahalanobis distance used is

$$D(S, X) = D_x(S, X) + D_\psi(S, X) \quad (8)$$

$$D_x(S, X) = \frac{(X_S \cdot \psi_S - X_P \cdot \psi_S)^2}{\Lambda_P} \quad (9)$$

$$D_\psi(S, X) = \frac{(\kappa_P \cos^{-1}(\psi_S \cdot \psi_P))^2}{2\pi} \quad (10)$$

$D(S, X)$ is the Mahalanobis distance (in standard deviations) between the patchlet X and surface S . In (10) we

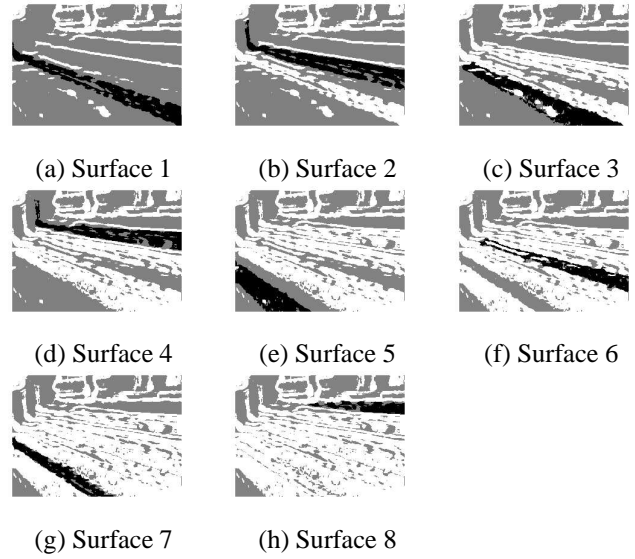


Figure 5: RANSAC-based greedy surface segmentation - black indicates pixels associated with the surface, grey are valid pixels, white are invalid pixels or pixels have already been assigned to a different surface

have converted the standard Mahalanobis equation to use the Fisher concentration parameter. This conversion uses the approximate relation between κ and σ , $\kappa = \frac{\sqrt{2\pi}}{\sigma}$. The angular distance between ψ_S and ψ_P is $\cos^{-1}(\psi_S \cdot \psi_P)$.

This algorithm is illustrated in Figure 5 using patchlets generated from the stereo image given in Figure 2. The grey regions indicate the patchlets in the unassigned patchlet set. White regions are where there is no data, either due to stereo failure or removal of data from a previous iteration. The black area indicates the best surface for this iteration. We can see that as the method progresses, smaller and smaller surfaces are selected and more and more of the image becomes white as the allocated patchlets are removed from the set. We used a threshold of 800 patchlets as the minimum requirement for sufficient support for a surface. The algorithm terminates in this case after 8 surfaces are obtained.

Figure 6 shows the result of this initial segmentation algorithm. The surface boundaries are aligned in the direction of maximum variation of the associated patchlet set, and the height and width are set to the minimum required to completely enclose the set. In Figure 6 we have allocated pixels to surfaces based on the surface that obtains the highest probability, $P(X|S)$ as given in (2), for each patchlet. Patchlets that have extremely low scores for all surfaces are assigned the black color which is the “no surface” association.

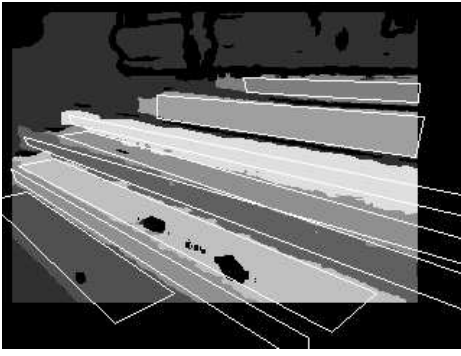


Figure 6: Results for RANSAC-based segmentation

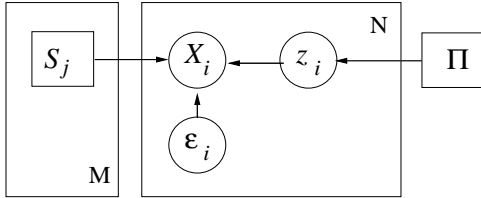


Figure 7: Graphical model of problem

6. Surface refinement with EM

The RANSAC surface extraction method described in the previous section provides a reasonable method for both model selection and model initialisation. However, to ensure an optimal surface estimation we have developed a surface refinement method using EM. EM is an iterative hill-climbing approach to probability maximisation that works in two succeeding stages, the “E” or “Expectation” stage and the “M” or “Maximisation” stage. The “E” determines the probability of data association for each data element in the system. It does not solve precisely which model is associated with the data element, instead it determines for each model the probability that the data is associated with that model. The “M” optimises the model parameter estimation based on the data weighted by their association probabilities. A full development of theory behind EM is outside the scope of this paper. However, there are many excellent descriptions such as [7, 1].

The graphical model in Figure 7 shows the probabilistic relationships between the different parts in our model. In this representation we have N patchlet elements generated from M surface models. A data association variable z_i is generated from the model likelihood π_j . The patchlet data elements X_i are generated from the data association variables z_i and the associated surface model parameters S_{z_i} , and are corrupted by sensor noise ϵ_i . z_i is a single value variable that identifies the model from which the patchlet was sensed (i.e., $z_i = 1$ means X_i was sensed from S_1).

To find the most likely set of surfaces we maximise

the joint probability of the set of surfaces and patchlet, or $P(\mathbf{SX})$. In the EM framework this is the same as maximising the complete log-likelihood function which, in this case, is

$$\langle l_c \rangle = \sum_{i=1}^N \sum_{j=1}^M \langle \xi_{ij} \rangle [\log P(X_i | S_j) + \log \pi_j] \quad (11)$$

where $P(X_i | S_j)$ is given in (2). In this equation we have re-defined the data association parameter z_i from a single value variable to an array of probability variables $\{\xi_{i1} \dots \xi_{iM}\}$ where $\sum_j \xi_{ij} = 1$. ξ_{ij} thus represents the probability that patchlet X_i is associated with surface S_j . $\langle \xi_{ij} \rangle$ is the expected value of ξ_{ij} .

Following D’Souza [1], the “E” step updates the expected values of $\langle \xi_{ij} \rangle$ and is determined by

$$\begin{aligned} \langle \xi_{ij} \rangle &= \frac{P(z_i = j | X_i, \mathbf{S})}{\sum_{k=1}^M P(z_i = k | X_i, \mathbf{S})} \\ &= \frac{P(x_i | z_i = j, \mathbf{S}) P(z_i = j)}{\sum_{k=1}^M P(x_i | z_i = k, \mathbf{S}) P(z_i = k)} \\ &= \frac{P(x_i | z_i = j, \mathbf{S}) \pi_j}{\sum_{k=1}^M P(x_i | z_i = k, \mathbf{S}) \pi_k} \end{aligned} \quad (12)$$

In the “M” step we adjust the model parameter estimates based on the expected data association. To do this, we take the derivative of the complete log-likelihood function (11) with respect to the model parameters, set them equal to zero and solve. Since the surface position uses a Gaussian model, we can obtain the closed form solution to the update equation for X_S which is

$$X_{S_j} = \frac{\sum_{i=1}^N \xi_{ij} \psi_{S_j} \cdot X_{P_i}}{\sum_{i=1}^N \xi_{ij}} \quad (13)$$

Unfortunately for the surface orientation parameter ψ_S , there is not a closed form solution to the maximisation of $\langle l_c \rangle$ with respect to ψ_S . Instead we compute the solution numerically by holding all other parameters constant and solving using Downhill Simplex numerical optimisation.

After the surface parameters (X_S, ψ_S) are updated, we adjust the surface boundaries. In the above formulation, a surface that increases in size can only increase in likelihood. There is never a penalty for increased size. To keep surface boundaries limited, we chose the maximum area of a surface to be the sum of the areas of the associated patchlets. That is, the area for S_j is set to $A_j = \sum_{i=1}^N \xi_{ij} a_i$ where a_i is the area of patchlet i determined by $H_P \times W_P$. In this respect, the total area of all surfaces in the mixture model is constrained to be equal or less than the combined patchlet area.

After the maximum size of a surface is determined, a gradient descent search is performed in the surface in-plane orientation and height and width parameters. The function that is maximised is the area of patchlets that fall within

the surface boundaries weighted by their data association probability ξ_{ij} .

Finally, it should be noted that we employed a “garbage class” similar to that used by Liu et al. in [6] to reduce the effect of unclassified patchlets. As can be seen in Figure 6, there are many patchlets within the system that do not fit any of the surface models. The above EM methods assume that for each and every patchlet in the system there is a surface model that is responsible for it. Rather than leaving outliers to be erroneously associated with one surface or the other, and corrupt their parameter estimation, we allow them to be collected by the garbage class.

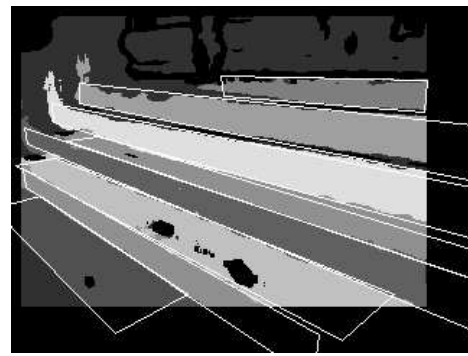
The garbage class is a uniform distribution that covers the entire planar parameter space with a constant but small likelihood value. That is, every patchlet in the system is deemed to fit this class equally, but with a very weak strength. The effect is that, if a patchlet fits any surface model even remotely well, the garbage class contribution is negligible. However, if a patchlet fits all of the surface models in the mixture poorly, it will be most closely associated with the garbage class. This prunes poorly fitting data and prevents them from distorting the final fit of the surfaces. In our experiments the garbage class that had a model likelihood, π_j , of 5% and a quality of fit to each patchlet of also 5%, making the overall attraction of the class 0.0025 (0.05×0.05).

7. Results

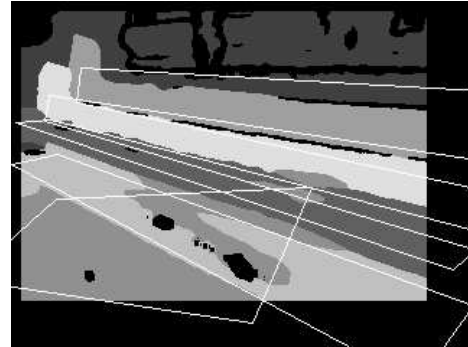
Figure 8(a) shows the surface segmentation of the image in Figure 2 after it has been refined by the EM process. This segmentation was done using priors on the uncertainties of the surface models of $\sigma_S = 5cm$ and the equivalent standard deviation for the Fisher model of 5° .

When compared with Figure 6 we can see that improvements are made in the boundaries of the surfaces. At the far end of the stairs some of the surfaces bled up the vertical surface of the balustrade. However, as shown in Figure 4 there is considerable rounding between the balustrade and the steps caused by the stereo process. As well, the uncertainty measures on the balustrade patchlets are very high. Given that the certainty of the patchlet parameters is weak, they are more susceptible to agree with models that do not match their parameters closely. The segmentation results of RANSAC is similar in quality to EM refinement. EM does generate larger classes with less residual error in the planar fit. However, the main advantage of EM is that it is amenable to combining data from multiple images, which the RANSAC method is not.

To investigate how important the patchlet parameters are to the stability of the solution, we performed the same EM refinement without the patchlet normal information. The result of this given in Figure 8(b). This refinement was ini-



(a) Segmentation with full patchlet model



(b) Segmentation with points only

Figure 8: EM-refined segmentation : patchlets versus points

tialised with the estimates established from the RANSAC process, so the starting-point estimates were reasonably close to an optimal value. The results with point-only information were very unsatisfactory. Three of the eight initial surfaces collapsed to a 0% likelihood ($\pi_j = 0$) and so had no associated patchlet. The upper surfaces were diagonal combinations of vertical and horizontal steps while the lowest step was both diagonal and spanning several surfaces.

Figure 9 shows the results of the segmentation on a variety of scenes of different complexity. The reference image is shown in the top row with the segmentation results shown below. These results give the reader an impression of how the methods work on a wider selection of real images. Figure 9(b) is interesting in that this scene is too far from the camera for the surfaces to be distinguishable. The upper steps have been so heavily smoothed by stereo that they truly appear as a ramp in the disparity image, and so their clustering together is expected and supported by the data. Figure 9(d), however, shows that even at long range the methods work, as long as the surfaces are of large enough scale to occupy sufficient pixels in the stereo image. Figure 9(e) and (f) are included to give the reader some idea of how the methods perform on natural, non-planar shapes.

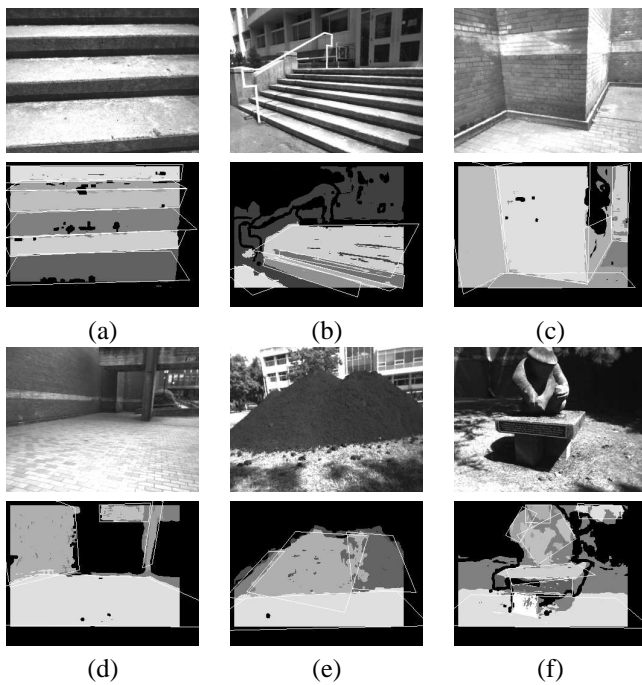


Figure 9: Segmentation results on various scenes

8. Conclusions

In this paper we have presented a method for segmenting noisy correlation stereo data into planar surfaces automatically. The method uses patchlets, an augmented surface element structure that incorporates position, orientation, and size. As well, patchlets have uncertainty measures on their position and orientation that are based on the propagation of the stereo error model from the disparity image to the patchlet parameters. The segmentation method is based on clustering and uses Expectation-Maximisation to obtain a maximum likelihood surface segmentation. We have demonstrated its effectiveness on a variety of images.

Patchlets are not restricted to planar surface representation. They can be used for modeling complex surfaces as well. The EM-based surface extraction method is not restricted to data from a single image and is well suited for use with larger patchlet clouds that are combined from multiple viewpoints. We would like to investigate the methods effectiveness on such combined data sets, and also investigate adapting the method to allow incremental additions of new patchlet data sets, as is likely to occur in a practical robot mapping system. As well, we are interested in pursuing patchlet-based multi-resolution modeling. Finally, it would of interest to use EM to learn the variance of each surface model fit and so identify the natural accuracies, as supported by the data, of the surfaces in the scene.

References

- [1] A. A. D'Souza. Using EM to estimate a probability density with a mixture of gaussians. Technical report, University of Southern California, Department of Computer Science, 1999.
- [2] F. P. Ferrie, J. Lagarde, and P. Whaithe. Darboux frames, snakes, and super-quadratics: Geometry from the bottom up. In *Proceedings of the IEEE Workshop on Interpretation of 3D Scenes*, pages 170–176, Austin, Texas, 1989.
- [3] M.A. Fischler and R.C. Bolles. A RANSAC-Based Approach to Model Fitting and Its Application to Finding Cylinders in Range Data. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 81)*, pages 637–643, 1981.
- [4] R.A. Fisher. Dispersion on a sphere. *Proceedings of the Royal Society London*, A127:295–305, 1953.
- [5] P. Fua. Reconstructing complex surfaces from multiple stereo views. In *Proceedings of the Fifth International Conference on Computer Vision (ICCV '96)*, pages 1078 – 1085, Cambridge, Massachusetts, USA, June 1996.
- [6] Y. Liu, R. Emery, D. Chakrabarti, W. Burgard, and S. Thrun. Using EM to learn 3D models with mobile robots. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2001.
- [7] T. P. Minka. Expectation-Maximization as lower bound maximization. Unpublished online document – <http://www.stat.cmu.edu/minka/papers/em.html>, 1998.
- [8] D. Murray. *Patchlets: a method of interpreting correlation stereo 3D data*. PhD thesis, University of British Columbia, Vancouver, Canada, 2004.
- [9] D. Murray and J. J. Little. Using real-time stereo vision for mobile robot navigation. *Autonomous Robots*, 8(2):161–171, 2000.
- [10] H. Pfister, M. Zwicker, J. van Baar, and M. Gross. Surfels: Surface elements as rendering primitives. In Kurt Akeley, editor, *Siggraph 2000, Computer Graphics Proceedings*, pages 335–342. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.
- [11] Szymon Rusinkiewicz and Mark Levoy. Qsplat: A multiresolution ppoint rendering system for large meshes. In *Computer Graphics, SIGGRAPH 2000 Proceedings*, 2000.
- [12] P. Sander. *On Reliably Inferring Differential Structure from Three-Dimensional Images*. PhD thesis, McGill University, Montreal, Canada, 1988.
- [13] R. Sara and R. Bajcsy. Fish-scales: Representing fuzzy manifolds. In *Proceedings of the 6th IEEE International Conference on Computer Vision (ICCV '98)*, pages 811–817, 1998.
- [14] M. Shimizu and M. Okutomi. Precise sub-pixel estimation on area-based matching. In *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV 2001)*, volume I, pages 90–97, Vancouver, BC, Canada, 2001.
- [15] R. Szeliski and D. Tonnesen. Surface modeling with oriented particle systems. *Computer Graphics*, 26(2):185–194, July 1992.