Probabilistic Signal Models to Regularise Dynamic Programming Stereo

Georgy Gimel'farb^{1*} and Uri Lipowezky²

¹ CITR, Department of Computer Science, Tamaki Campus, University of Auckland
Private Bag 92019, Auckland 1, New Zealand
`g.gimelfarb@auckland.ac.nz`

² Tiltan System Engineering Ltd.
35 Hayarkon Street, Beney - Beraq 51204, Israel
`Uri@tiltan-se.co.il`

Abstract. Ill-posedness of the binocular stereo problem stems from partial occlusions and homogeneous textures of a 3D surface. We consider the symmetric dynamic programming stereo regularised with respect to partial occlusions. The regularisation is based on Markovian models of epipolar profiles and stereo signals that allow for measuring similarity of stereo images with due account of binocular and monocular visibility of the surface points. Experiments show that the probabilistic regularisation yields mostly accurate elevation maps but fails in excessively occluded or shaded areas.

1 Introduction

Computational binocular stereo is an ill-posed problem because the same stereo pair can be produced by very different optical surfaces. The ill-posedness stems from partial occlusions yielding no stereo correspondence and from uniform or repetitive textures resulting in multiple equivalent correspondences. To obtain a unique solution closely approaching visual or photogrammetric reconstruction, the stereo problem has to be regularised.

We consider dynamic programming stereo (DPS) that reconstructs a 3D surface as a collection of independent continuous epipolar profiles. The reconstruction is based on the best correspondence between stereo images, each profile maximising the total similarity of signals (grey values or colours) in the corresponding pixels or of local image features derived from the signals [1,2,3,4,6,7,8,9].

The intensity-based symmetric DPS (SDPS) can be regularised with respect to partial occlusions by modelling the profiles with explicit Markov chains of the 3D points and signals [5]. We compare three models of the epipolar profiles with respect to the overall accuracy of stereo reconstruction. Experiments are conducted with a large-size digitised aerial stereo pair of an urban scene in Fig. 1 having various partially occluded areas and the known ground control.

* This work was supported by the University of Auckland Research Committee grant 3600327/9343.

Fig. 1. Left (2329×1442) and right (1887×1442) stereo images “Town”

2 Markovian Models of an Epipolar Profile

Let x_{left} and x_{right} be discrete x -coordinates of pixels along the corresponding epipolar scan-lines in the left and right image of a stereo pair, respectively. Let x denote a discrete x -coordinate of points of an epipolar profile. Assuming the symmetric epipolar geometry [4], $x = (x_{\text{left}} + x_{\text{right}})/2$. Let $p = x_{\text{left}} - x_{\text{right}}$ be the x -disparity (parallax) of the corresponding pixels. The N -point discrete

Fig. 2. Fragment of the graph of profile variants and the allowable transitions between the nodes

continuous profile $\mathbf{P} = ((x_i, p_i : i = 0, \dots, N - 1)$ is modelled by a Markov chain of transitions between the successive nodes $v_i(s_i) = (x_i, p_i, s_i)$ of a graph of profile variants (GPV) shown in Fig. 2. Here, $s \in \{B, ML, MR\}$ denotes a visibility state indicating, respectively, the binocularly visible point (BVP) and the monocularly visible point (MVP) observed only in the left or right image. Only the BVP $v(B) = (x, p, B)$ involves the pair of corresponding pixels with the x -coordinates $x_{\text{left}} = x + p/2$ and $x_{\text{right}} = x - p/2$. The MVPs $v(ML) = (x, p, ML)$ and $v(MR) = (x, p, MR)$ are depicted by the single pixels with the x -coordinate x_{left} or x_{right} , respectively.

Transition probabilities for a stationary Markov chain describing the shape of the random profiles $\mathbf{P}$ depend only on the successive visibility states [4]:

$$\Pr(v_{i+1}(s_{i+1})|v_i(s_i)) \equiv \pi(s_{i+1}|s_i) \quad (1)$$

On the assumption that the transitions to the MVPs with $s = ML$ and $s = MR$ are equiprobable, the random profile is specified by the two transition probabilities [4]: $\pi_{B|B}$ and $\pi_{M|M}$ where M stands for MR or ML . Each profile $\mathbf{P}$ specifies particular stereo correspondence between the left and right images of a stereopair, and similarity between the corresponding pixels is measured with a specific Markovian model of mutual signal adaptation. The adaptation estimates and excludes photometric distortions of the stereo images relative to the surface texture [3,4]. After the images are mutually adapted, the probability of transition to each BVP $v_i(B) = (x_i, p_i, B)$ depends on the point-wise residual absolute difference Δ_i between the corresponding signals. As follows from the GPV in Fig. 2, the transition probabilities to the BVPs $\Pr(v_i(B), \Delta_i|v_{i-1}(s_{i-1}))$ define uniquely the transition probabilities to the MVPs.

The regularised SDPS relates the point-wise signal similarity for each node $v_i(s_i)$ of the GPV to the log-likelihood ratio

$$l(v_i(s_i), \Delta_i|v_{i-1}(s_{i-1})) = \ln \Pr(v_i(s_i), \Delta_i|v_{i-1}(s_{i-1})) - \ln \Pr(s_i|s_{i-1})_{\text{rand}}$$

of the transition probabilities for the profile specified by a given stereopair and the purely random profile.

Below we experimentally compare three probabilistic models of the signal-depending and random profiles assuming, for simplicity, that $\pi_{B|B} + \pi_{M|M} = 1$. The simplest model of the transition probability $\Pr(v_i(B), \Delta_i | v_{i-1}(s_{i-1})) = F_B(\Delta_i)$ such that $\sum_{\Delta=0}^{\Delta_{\max}} F_B(\Delta) = 1$ is introduced in [5]:

$$F_B(\Delta) \propto \min \{1 - \tau, \max \{\tau, \exp(-\gamma \cdot \Delta)\}\} \quad (2)$$

where γ is a scaling factor and the threshold $\tau > 0$ excludes zero probabilities (in our experiments $\tau = 10^{-10}$).

The adaptation of corresponding signals amplifies the relative number of zero deviations. This can be accounted for by using an additional parameter $\alpha = F_B(0)$ in the transition probability:

$$F_B(\Delta) = \begin{cases} \alpha & \text{if } \Delta = 0 \\ (1 - \alpha) \frac{\max\{\tau, \exp(-\gamma \cdot \Delta)\}}{\sum_{\delta=1}^{\Delta_{\max}} \max\{\tau, \exp(-\gamma \cdot \delta)\}} & \text{otherwise} \end{cases} \quad (3)$$

where $\Delta_{\max}$ is the maximum absolute deviation ($\Delta_{\max} = 255$).

The profile models to be compared yield the following likelihood ratios:

- for the transition models in [5]:

$$\begin{aligned} l(v_i(B), \Delta_i | v_{i-1}(s)) &= \log F_B(\Delta_i) - \log \pi_{B|B} \\ l(v_i(M), \Delta_i | v_{i-1}(s)) &= \log (1 - F_B(\Delta_i)) - \log \pi_{M|M} \end{aligned} \quad (4)$$

- for the conditional Markov model of the profile points depending on the adapted signals:

$$\begin{aligned} l(v_i(B), \Delta_i | v_{i-1}(s)) &= \log \left(\frac{\pi_{B|B}^{\circ} \cdot F_B(\Delta_i)}{\pi_{B|B}^{\circ} \cdot F_B(\Delta_i) + \pi_{M|M}^{\circ} \cdot P_M} \right) - \log \pi_{B|B} \\ l(v_i(M), \Delta_i | v_{i-1}(s)) &= \log \left(\frac{\pi_{M|M}^{\circ} \cdot P_M}{\pi_{B|B}^{\circ} \cdot F_B(\Delta_i) + \pi_{M|M}^{\circ} \cdot P_M} \right) - \log \pi_{M|M} \end{aligned} \quad (5)$$

where $\pi_{B|B}^{\circ}$ and $\pi_{M|M}^{\circ}$ denote the transition probabilities specifying the actual shape of the profile specified by the stereo images ($\pi_{B|B}^{\circ} + \pi_{M|M}^{\circ} = 1$), and P_M is the probability of signal deviations for the MVPs ($P_M = 1/\Delta_{\max}$ if the equiprobable deviations are assumed), and

- for the joint Markov models of the profile points and adapted signals:

$$\begin{aligned} l(v_i(B), \Delta_i | v_{i-1}(s)) &= \log (\pi_{B|B}^{\circ} \cdot F_B(\Delta_i)) - \log (\pi_{B|B} \cdot P_M) \\ l(v_i(M), \Delta_i | v_{i-1}(s)) &= \log \pi_{M|M}^{\circ} - \log \pi_{M|M} \end{aligned} \quad (6)$$

3 Experimental Results and Conclusions

The original photos of the urban scene in Fig. 1 containing different types of open and partially occluded areas are obtained from the altitude 960 m using

Table 1. Accuracy of the regularised SDPS reconstruction in terms of the cumulative percentage of the control points with the absolute error less than or equal to ε . The notation used: $\bar{\varepsilon}$, σ , and $\varepsilon_{\max}$ are the mean absolute error, standard deviation, and maximum error, respectively, C_B is the cross-correlation of the corresponding signals for the reconstructed surface (in the parentheses – after this latter is smoothed by post-processing); ν_B is the relative number of the BVPs in the surface

Model (4) – (2) with $\pi_{B B} = 0.10$ and $\gamma = 0.1$:														
$\bar{\varepsilon} = 3.66$; $\sigma = 6.84$; $\varepsilon_{\max} = 44$; $C_B = 0.944$ (0.856); $\nu_B = 81.8\%$														
ε :	0	1	2	3	4	5	10	15	20	25	30	35	40	44
%	18.0	60.1	71.0	78.6	82.2	84.9	90.9	93.0	95.8	97.1	98.2	99.0	99.5	100.0
Model (5) – (3) with $\pi_{B B} = 0.10$, $\pi_{B B}^{\circ} = 0.90$, $\gamma = 1$, and $\alpha = 0.90$:														
$\bar{\varepsilon} = 3.69$; $\sigma = 6.59$; $\varepsilon_{\max} = 48$; $C_B = 0.945$ (0.849); $\nu_B = 81.6\%$														
ε :	0	1	2	3	4	5	10	15	20	25	30	35	40	48
%	18.5	58.7	68.4	75.7	80.7	82.8	90.6	94.5	95.8	97.4	98.4	99.0	99.5	100.0
Model (5) – (3) with $\pi_{B B} = 0.25$, $\pi_{B B}^{\circ} = 0.75$, $\gamma = 1$, and $\alpha = 0.90$:														
$\bar{\varepsilon} = 3.72$; $\sigma = 6.73$; $\varepsilon_{\max} = 53$; $C_B = 0.948$ (0.833); $\nu_B = 78.4\%$														
ε :	0	1	2	3	4	5	10	15	20	25	30	35	40	53
%	19.3	57.2	66.6	77.0	80.4	83.6	89.0	94.5	96.9	98.2	99.0	99.2	99.2	100.0
Model (6) – (3) with $\pi_{B B} = 0.10$, $\pi_{B B}^{\circ} = 0.90$, $\gamma = 1$, and $\alpha = 0.90$:														
$\bar{\varepsilon} = 2.96$; $\sigma = 5.34$; $\varepsilon_{\max} = 37$; $C_B = 0.950$ (0.861); $\nu_B = 79.6\%$														
ε :	0	1	2	3	4	5	10	15	20	25	30	37		
%	19.6	63.2	74.2	80.7	84.3	87.7	93.2	94.8	96.6	98.4	99.5	100.0		
Model (6) – (3) with $\pi_{B B} = 0.25$, $\pi_{B B}^{\circ} = 0.75$, $\gamma = 1$, and $\alpha = 0.90$:														
$\bar{\varepsilon} = 2.74$; $\sigma = 4.82$; $\varepsilon_{\max} = 36$; $C_B = 0.953$ (0.862); $\nu_B = 77.6\%$														
ε :	0	1	2	3	4	5	10	15	20	25	30	36		
%	18.8	62.1	74.9	82.2	85.6	88.0	94.3	96.3	97.7	99.0	99.5	100.0		

the photogrammetric camera Wild RC30 with the focal length 153.26 mm. For the SDPS reconstruction the digitised images are transformed to the epipolar geometry and scaled down to the resolution of 330 mm per pixel. The left and right images are of the size 2329×1441 and 1887×1441 pixels, respectively. The x -disparity range for these images is [50, 200].

This scene has 383 uniformly distributed ground control points (GCP) found by visual (photogrammetric) processing. Most of them are at the roofs of the buildings because these latter present a real challenge to the reconstruction. These GCPs allow to analyse the accuracy of the regularised SDPS in most complicated conditions and indicate typical reconstruction errors. Table 1 presents results of the reconstruction using the above regularising models with different parameters. The range of successive signal adaptation is $\pm 20\%$ of the grey level difference estimated for the surface texture. To simplify the comparisons, numerical values of the likelihood functions $l_B(\Delta) \equiv l(v_i(B), \Delta|v_{i-1}(s))$ and

Table 2. Likelihood values for the residual signal deviations

Model (4) – (2) with $\pi_{B B} = 0.10$ and $\gamma = 0.1$:														
Δ :	0	1	2	3	4	5	6	7	8	9	10	11	12	13
$l_B(\Delta)$:	2.30	2.07	1.84	1.61	1.38	1.15	0.92	0.69	0.46	0.23	0.00	-0.23	-0.46	-0.69
$l_M(\Delta)$:	-22.9	-1.48	-0.89	-0.59	-0.40	-0.27	-0.18	-0.12	-0.07	-0.03	0.00	0.02	0.03	0.03
Δ :	14	15	16	17	18	19	20	30	40	50	100	150	200	255
$l_B(\Delta)$:	-0.92	-1.15	-1.38	-1.61	-1.84	-2.07	-2.30	-4.61	-6.91	-9.21	-20.7	-20.7	-20.7	-20.7
$l_M(\Delta)$:	0.06	0.07	0.08	0.09	0.09	0.09	0.10	0.10	0.11	0.11	0.11	0.11	0.11	0.11
Model (5) – (3) with $\pi_{B B} = 0.10$, $\pi_{B B}^\circ = 0.90$, $\gamma = 1$, and $\alpha = 0.90$														
Δ :	0	1	2	3	4	5	6	7	8	9	10	11	12	13
$l_B(\Delta)$:	2.30	2.30	2.28	2.25	2.17	1.98	1.60	0.98	0.16	-0.76	-1.73	-2.72	-3.72	-4.72
$l_M(\Delta)$:	-7.53	-4.88	-3.89	-2.93	-2.00	-1.19	-0.58	-0.20	-0.02	0.06	0.09	0.10	0.10	0.10
Δ :	14	15	16	17	18	19	20	30	40	50	100	150	200	255
$l_B(\Delta)$:	-5.72	-6.72	-7.72	-8.72	-9.72	-10.7	-11.7	-13.7	-13.7	-13.7	-13.7	-13.7	-13.7	-13.7
$l_M(\Delta)$:	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
Model (5) – (3) with $\pi_{B B} = 0.25$, $\pi_{B B}^\circ = 0.75$, $\gamma = 1$, and $\alpha = 0.90$														
Δ :	0	1	2	3	4	5	6	7	8	9	10	11	12	13
$l_B(\Delta)$:	1.38	1.37	1.33	1.24	1.04	0.63	-0.01	-0.84	-1.77	-2.75	-3.74	-4.73	-5.73	-6.73
$l_M(\Delta)$:	-6.25	-3.62	-2.65	-1.74	-0.94	-0.35	0.00	0.17	0.24	0.27	0.28	0.29	0.29	0.29
Δ :	14	15	16	17	18	19	20	30	40	50	100	150	200	255
$l_B(\Delta)$:	-7.73	-8.73	-9.73	-10.7	-11.7	-12.7	-13.7	-15.7	-15.7	-15.7	-15.7	-15.7	-15.7	-15.7
$l_M(\Delta)$:	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29
Model (6) – (3) with $\pi_{B B} = 0.10$, $\pi_{B B}^\circ = 0.90$, $\gamma = 1$, and $\alpha = 0.90$														
Δ :	0	1	2	3	4	5	6	7	8	9	10	11	12	13
$l_B(\Delta)$:	7.64	4.48	3.48	2.48	1.48	-0.02	-1.02	-2.02	-3.02	-4.02	-5.02	-6.02	-7.02	-8.02
$l_M(\Delta)$:	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20
Δ :	14	15	16	17	18	19	20	30	40	50	100	150	200	255
$l_B(\Delta)$:	-9.02	-10.0	-11.0	-12.0	-13.0	-14.0	-15.0	-16.0	-16.0	-16.0	-16.0	-16.0	-16.0	-16.0
$l_M(\Delta)$:	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20	-2.20
Model (6) – (3) with $\pi_{B B} = 0.25$, $\pi_{B B}^\circ = 0.75$, $\gamma = 1$, and $\alpha = 0.90$														
Δ :	0	1	2	3	4	5	6	7	8	9	10	11	12	13
$l_B(\Delta)$:	6.54	3.88	2.88	1.88	0.88	-0.12	-1.12	-2.12	-3.12	-4.12	-5.12	-6.12	-7.12	-8.12
$l_M(\Delta)$:	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10
Δ :	14	15	16	17	18	19	20	30	40	50	100	150	200	255
$l_B(\Delta)$:	-9.12	-10.1	-11.1	-12.1	-13.1	-14.1	-15.1	-17.1	-17.1	-17.1	-17.1	-17.1	-17.1	-17.1
$l_M(\Delta)$:	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10	-1.10

$l_M(\Delta) \equiv l(v_i(M), \Delta|v_{i-1}(s))$ for the model parameters in use are given in Table 2. For the models (5) and (6), $l_B(\Delta) > l_M(\Delta)$ if $\Delta \geq 6 \dots 8$.

Figure 3 presents the greycoded digital x -parallax map (DPM), or the range image of the reconstructed x -disparities, and the orthoimage formed by fusing the left and right images in Fig. 1 in accord with the DPM. The fusion takes account of the visibility of each point. These results are obtained for the best in Table 1 regularising model (6) using the transition probability (3) with $\gamma = 1$

Fig. 3. Greycoded range image of the reconstructed DPM and the orthoimage 2108×1442 of the “Town” (vertical white pointers show positions of the 383 GCPs)

and $\alpha = 0.90$ and for the regularising parameters $\pi_{B|B} = 0.25$ and $\pi_{B|B}^{\circ} = 0.75$. In this case the reconstructed DPM has the mean absolute error $\bar{\varepsilon} = 2.74$ with the standard deviation $\sigma = 4.82$. The absolute error for 74.9% of the GCPs is

Fig. 4. Reconstruction around the GCP (368, 1277): the fragments 351×351 of the left (a) and right (b) images (white and grey pointers indicate the GCPs and reconstruction results, respectively; for a small error, the white pointer is superposed on the grey one), the greycoded DPM (c), and the orthoimage (d) with the GCPs (the larger the bottom rectangle of the pointer, the larger the error)

equal to or less than 2. The maximum error is $\varepsilon_{\max} = 36$ but only 5.7% of the GCPs have the error greater than or equal to 10.

To exclude local y -discontinuities due to homogeneous texture, the reconstructed DPMs are post-processed. The post-processing consists of the in-column median filtering within the moving window 3×15 with the subsequent in-line median filtering of the results of the in-column filtering.

A few typical reconstruction errors are shown in Fig. 4. Here, partially occluded areas around the tall buildings are quite large comparing to the adjacent open areas with a high contrast of the texture. In such cases, especially, if the deep shadows create large uniform areas at and around the walls, the SDPS reconstruction fails. But some of these areas are challenging even for visual stereo

perception although the corresponding points can be easily found by visual comparisons of the images. But the reconstruction is accurate when the surface texture is not uniform and the occluded parts are not similar do not prevail. Large reconstruction errors ($\varepsilon \geq 6$, that is, larger than 4% of the x -disparity range) are encountered in less than 10% of all the GCPs.

The errors are localised to within small areas of the overall scene so that the cross-correlation of the corresponding pixels in the stereo images is relatively high, namely, 0.953 and 0.862 without and with the post-processing, respectively. But the cross-correlation of the corresponding pixels for the GCPs is only 0.654 so that the accurate visual matching does not mean the highest correlation-based similarity of the stereo images.

Our experiments show that the regularised SDPS yields accurate overall reconstruction accuracy with the absolute errors less than 3 pixels (2% of the total x -disparity range) for more than 80–82% of the surface points depending on a proper choice of the regularising signal model. But the regularisation fails in the cases where too large occluded or deeply shaded areas are involved.

Because large local errors do not effect notably the overall similarity of the corresponding points in the stereo images, it is the regularisation rather than stereo matching that plays a crucial role in solving the stereo problem.

References

1. Bensrhair, A., Miché, P., Debrie, R.: Fast and automatic stereo vision matching algorithm based on dynamic programming method. *Pattern Recognition Letters* **17** (1996) 457–466 [814](#)
2. Cox, I. J., Hingorani, S. L., Rao, S. B.: A maximum likelihood stereo algorithm. *Computer Vision and Image Understanding* **63** (1996) 542–567 [814](#)
3. Gimel'farb, G. L.: Intensity-based computer binocular stereo vision: signal models and algorithms. *Int. Journal of Imaging Systems and Technology* **3** (1991) 189–200 [814](#), [816](#)
4. Gimel'farb, G.: Stereo terrain reconstruction by dynamic programming. In: Jaehne, B., Haussecker, H., Geisser, P. (Eds.): *Handbook of Computer Vision and Applications*. Academic Press, San Diego, vol. 2 (1999) 505–530 [814](#), [815](#), [816](#)
5. Gimel'farb, G.: Binocular stereo by maximizing the likelihood ratio relative to a random terrain. In: Klette, R., Peleg, S., Sommer, G. (Eds.): *Robot Vision* (Lecture Notes in Computer Science **1998**). Springer, Berlin (2001) 201–208 [814](#), [817](#)
6. Li, Z.-N.: Stereo correspondence based on line matching in Hough space using dynamic programming. *IEEE Trans. on Systems, Man, and Cybernetics* **24** (1994) 144–152 [814](#)
7. Rojas, A., Calvo, A., Muñoz, J.: A dense disparity map of stereo images. *Pattern Recognition Letters* **18** (1997) 385–393 [814](#)
8. Vergauwen, M., Pollefeys, M., Van Gool, L.: A stereo vision system for support of planetary surface exploration. In: Schiele, B., Sagerer, G. (Eds.): *Computer Vision Systems Second Int. Workshop (ICVS 2001)*, Vancouver, Canada, July 7–8, 2001 (Lecture Notes in Computer Science **2095**). Springer, Berlin (2001) 298–312 [814](#)
9. Yip, R. K. K., Ho, W. P.: A multi-level dynamic programming method for stereo line matching. *Pattern Recognition Letters* **19** (1998) 839–855 [814](#)