

Integrating Stereo with Shape-from-Shading derived Orientation Information

Tom S. F. Haines and Richard C. Wilson

Department of Computer Science
University of York
Heslington, York, UK

thaines,wilson@cs.york.ac.uk

Abstract

Binocular stereo has been extensively studied for extracting the shape of a scene. The challenge is in matching features between two images of a scene; this is the correspondence problem. Shape from shading (SfS) is another method of extracting shape. This models the interaction of light with the scene surface(s) for a single image. These two methods are very different; stereo uses surface features to deliver a depth-map, SfS uses shading, albedo and lighting information to infer the differential of the depth-map.

In this paper we develop a framework for the integration of both depth and orientation information. Dedicated algorithms are used for initial estimates. A Gaussian-Markov random field then represents the depth-map, Gaussian belief propagation is used to approximate the MAP estimate of the depth-map. Integrating information from both stereo correspondences and surface normals allows fine surface details to be estimated.

1 Introduction

Binocular stereo is a long-standing problem in computer vision[17]. It enables the construction of 3D models from two 2D images, by solving the correspondence problem, where matching features are found between two images such that the matched features are at the same location on the object. If camera calibration is then available depth can be reconstructed from such matches. A standard preprocessing step involves rectifying the images, so that they represent the images taken from an idealised horizontal parallel camera pair. Given a rectified image pair features can only match features on the same scan-line. In the dense stereo problem where every pixel is a feature a disparity map is created, where the disparity assigned to each pixel is the offset along the x -axis to its matching pixel in the other image.

Dense stereo algorithms may be divided into two steps. First is the calculation of a matching cost for each disparity at each location, represented by the 3D Disparity Space Image (DSI). In areas with strong cues a DSI gives a clear indication of actual disparity, but in relatively uniform areas it will not distinguish the correct disparity from incorrect disparities. Suggested approaches include normalised cross-correlation[17] and sliding

windows[3]. A common choice is to use individual pixel dissimilarities and rely on modelling assumptions in the second step to recover a reasonable solution.

The second step is the selection of disparities to find a consistent solution. Simply selecting the best matching cost does not work well because of noise and ambiguities, so there is a need for modelling assumptions. Modern approaches use techniques such as dynamic programming[1], graph cuts[4] and belief propagation[21]. The most common modelling assumption is a smoothness term, which is often equivalent to assuming piecewise planar surfaces or fronto-parallel piecewise planar surfaces. Whilst these approaches do well in regions with strong stereo cues uniform areas are generally plane fitted or interpolated, regardless of actual shape.

Shape from Shading (SfS) relies on the shading information available from a single image. It is premised on the intensity of light reflected by a surface being related to the angle between the surface and light source(s). It therefore provides information about surface gradient. The approach was pioneered by Horn[11] and then Ikeuchi and Horn[12]. Whilst complete orientation is desired surface intensity only provides tilt information. Again, as with stereo, modelling assumptions are used to resolve the ambiguities. A smooth surface assumption is again common, though this is smoothing surface orientation rather than disparity. SfS algorithms generally assume Lambertian surface reflectance and constant albedo over the surface in question, and therefore need constant albedo objects.

Stereo algorithms do not perform effectively in areas of uniform texture. Such regions will generally either be interpolated or plane fitted, which is not necessarily a true reflection of the surface shape. In contrast SfS can operate only in areas where albedo can be inferred, so a uniform albedo assumption needs to be used. This makes SfS ideal for filling in areas where stereo has insufficient information[14]. Furthermore, stereo provides information about albedo, which is necessary for SfS. In combining these ideas we have an improved set of modelling assumptions allowing for a surface estimate with greater detail. Leclerc and Bobick[14] have used stereo to provide initialisation and boundary constraints for SfS. Cryer, Tsai and Shah[7] combine SfS and stereo in the frequency domain, using a high pass filter for the SfS and a low pass filter for the stereo. Other work has taken an object centred approach[10, 16]. Here a model is initialised with a stereo algorithm and then optimised to fit both stereo and shading information. The method needs a good initialisation however, and is not effective with only a single stereo pair for initialisation. Jin, Yezzi and Soatto[13] assume the image is divided into areas of texture and constant albedo and apply separate cost functions to each area and solve with level sets. Shao et al[18] use an additional cost for the difference between SfS irradiance in the left image and image irradiance at the corresponding point in the right image. This combines depth and shading into a single cost function.

The paper is organised as follows; section 2 gives the problem formulations for stereo and SfS. Section 3 gives the core details of our algorithm. Section 4 describes albedo estimation whilst sections 5 and 6 detail the SfS and stereo algorithms respectively. Section 7 presents results and, finally, section 8 concludes.

2 Problem Formulation

The goal is to use the methods of stereo and shape from shading in a complementary way. Here we briefly describe the problem formulations of both.

2.1 Stereo

The images captured by the camera pair are initially rectified using the camera calibration before processing. The input to the stereo problem is therefore a rectified image pair, the images notionally referred to as the left, $I_L(x,y)$ and right, $I_R(x,y)$. The output is a disparity map, $D(x,y)$, representing the dense correspondences between images. Rectification has ensured that epipolar lines are horizontal, therefore disparities are offsets on the x -axis, i.e. $I_L(x,y)$ corresponds to $I_R(x+D(x,y),y)$. The process may be divided into two steps, first a $DSI(x,y,d)$ is defined expressing the cost of matching $I_L(x,y)$ and $I_R(x+d,y)$. Modelling assumptions are then used to select an optimal set of matches which produces the solution, $D(x,y)$. Given camera calibration disparity may be converted into a depth map.

2.2 Shape from Shading

SfS uses the image intensity of a single image¹, $L(x,y)$. The goal of SfS is to recover surface orientation for each pixel, $\mathbf{n}(x,y)$. Under a single light source and Lambertian reflectance model the surface normals are related to the image intensity via

$$L(x,y) = A(x,y)(\mathbf{n}(x,y) \cdot \mathbf{s}) \quad (1)$$

where \mathbf{s} is the light-source direction and $A(x,y)$ is the apparent albedo at (x,y) in the image. The goal of SfS is to recover the surface normals given the luminance map, albedo map and the light source direction. As equation 1 only constrains the angle between the light source and surface normal modelling assumptions such as surface smoothness must be introduced to solve the problem. In principle depth can be recovered from the normal map by integrating over the surface. This is neither straightforward nor accurate however.

3 Integrating Depth and Orientation Information

Once we have a field of surface normals to hand we can use it to provide information about scene shape by integration. Traditionally this is done using a global integration method, such as that of Frankot and Chellapa[9]. Depth information is also available however, provided directly by the stereo algorithm. Our goal is then to combine these two sources of information to produce an improved estimate of the surface. We do this within the framework of Gaussian belief propagation. This enables us to define the required surface as the MAP estimate of a Markov random field, and to combine the two sources of information in a probabilistic way.

3.1 Belief Propagation

Belief propagation is a powerful method for finding the posterior distribution of a Markov random field. It has previously been used with discrete distributions to find stereo disparities[8]. In our case we have to recover disparity to sub-pixel accuracy otherwise surface normals will not provide much information, i.e. the change in orientation produced by a unit change in disparity is often an order of magnitude more than the SfS derived orientation information can provide. Using discrete distributions would result in an infeasibly large

¹We use the luminance channel of the Luv colour space for the intensity. Experimentation has shown this to be in reasonable agreement with Lambert's law for non-specular objects with the cameras used. There is an implicit white light assumption being made here.

number of disparity labels. This makes it essential to use continuous density functions representing continuous disparity measurements. One tractable solution is to use Gaussian distributions. The beliefs are then defined by the mean and variance of the Gaussian distribution, allowing orientation information to be effectively used. We adopt this approach in this paper.

Loopy belief propagation works by iteratively passing messages between nodes of the MRF. The message that a node t passes to its neighbour s is [19]

$$m_{ts}^{(n)}(x_s) = \alpha \int_{x_t} \psi_{st}(x_s, x_t) \psi_t(x_t, y_t) \prod_{u \in N/s} m_{ut}^{(n-1)}(x_t) dt \quad (2)$$

Here x_t is the disparity at node t ; $\psi_{st}(x_s, x_t)$ is the compatibility distribution between the disparities at t and s ; $\psi_t(x_t, y_t)$ is the distribution of disparities inferred from the observed evidence y_t ; $m_{ut}^{(n-1)}(x_t)$ is a message from the previous iteration; and the set N/s is the neighbourhood of t excluding s . We can then compute the belief at node t using

$$b_t^{(n)} = \alpha \psi_t(x_t, y_t) \prod_{u \in N} m_{ut}(x_t) \quad (3)$$

We adopt a variant of the Gaussian algebra of Cowell [6]. The Normal distribution is defined as a function of the precision \mathbf{P} and the precision times the mean $\mathbf{P}\boldsymbol{\mu}$, which we will refer to as the p-mean. The precision is equal to the inverse covariance matrix, i.e. $\mathbf{P} = \Sigma^{-1}$. We have

$$\phi[\mathbf{P}\boldsymbol{\mu}, \mathbf{P}] = \alpha \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{P} (\mathbf{x} - \boldsymbol{\mu}) \right] \quad (4)$$

The reason for defining ϕ in this way is that it produces a simple set of rules for manipulating the distributions, which are given in Appendix A.

Under our Gaussian model the stereo algorithm is used to give an initial estimate of the disparities. At a point t in the image the stereo algorithm gives a set of measurements, y_t , which are used to infer a distribution for the disparities, x_t . This is modelled by a Normal distribution, $\psi_t(x_t, y_t) = \phi[\mathbf{P}_t \boldsymbol{\mu}_t, \mathbf{P}_t]$. The mean $\boldsymbol{\mu}$ and precision \mathbf{P} for this distribution are computed from the stereo algorithm as detailed in section 6.

The compatibility distribution between two neighbouring points in the image s and t is also modelled by a normal distribution. If the disparity at t is x_t then we would expect the disparity at s to be $x_t + z_{ts}$ where z_{ts} is the disparity change predicted by integrating the surface normals along the path from t to s . The compatibility distribution $\psi_{st}(x_s, x_t)$ is therefore defined as a Normal distribution with mean $x_t + z_{ts}$ and a fixed precision P_n which reflects the accuracy of the surface normals. We therefore obtain

$$\psi_{st}(x_s, x_t) = \phi \left[P_n \begin{pmatrix} -z_{ts} \\ z_{ts} \end{pmatrix}, P_n \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \right] \quad (5)$$

Since the points are neighbours in the image we can assume that the surface normal direction is constant along the path between them, and use an interpolated surface orientation at the half way point. This is in fact necessary to avoid bias in the result. The two separate processes therefore influence the MRF in different ways; the local measurement process models the depth information and the compatibility between sites is used to incorporate the orientation information.

Since the distributions are Normal, the messages are also Normal distributions. We begin by defining the following quantities:

$$P_0 = P_t + \sum_{u \in N/s} P_{ut} \quad P_0 \mu_0 = P_t \mu_t + \sum_{u \in N/s} P_{ut} \mu_{ut}$$

These are the local precision and p-mean respectively, excluding the message we are currently computing. The new message $m_{ts}^{(n)}(x_s)$ is a Normal distribution which we will define as

$$m_{ts}^{(n)}(x_s) = \phi[\mathbf{P}_{ts} \mu_{ts}, \mathbf{P}_{ts}]$$

Applying Eqn. 2, we obtain the update rules:

$$\begin{aligned} P_{ts} &\leftarrow P_n - P_n(P_n + P_0)^{-1} P_n \\ P_{ts} \mu_{ts} &\leftarrow P_n z_{ts} + P_n(P_n + P_0)^{-1} (P_0 \mu_0 - P_n z_{ts}) \end{aligned} \quad (6)$$

We iteratively apply these rules to find an estimate of the MAP disparity map. The beliefs are given by

$$b_t^{(n)} = \alpha \psi_t(x_t, y_t) \prod_{u \in N} m_{ut}(x_t) = \phi \left[P_t \mu_t + \sum_{u \in N} P_{ut} \mu_{ut}, P_t + \sum_{u \in N} P_{ut} \right]$$

so the mean, and hence the estimated disparity, is

$$\mu(t) = (P_t \mu_t + \sum_{u \in N} P_{ut} \mu_{ut}) (P_t + \sum_{u \in N} P_{ut})^{-1} \quad (7)$$

4 Albedo Estimation

Under the Lambertian reflectance assumption SfS requires an albedo map as input. The *surface texture* consists of albedo and colour; colour is taken to be the (u, v) channels of *Luv* colour space. For an arbitrary texture it is impossible to distinguish texture variation from shading; this is the basis of ‘3D’ effects in user interfaces. It has already been noted that in variable texture regions stereo matching is effective, so additional SFS information is only necessary in relatively uniform regions. Uniform regions allow us to ignore texture variation.

We begin by segmenting the image into uniform regions with mean shift[5] to obtain a set of regions, R . Within each of these regions the colour is uniform and the albedo is assumed to be uniform. The luminance L will however vary across the region because of shading effects. In order to correctly compute the albedo of a region we need to account for shading effects using Equation 1. Given a field of surface normals $\mathbf{n}(x, y)$ we can estimate the albedo at each pixel via the relation $A(x, y) = L / (\mathbf{n}(x, y) \cdot \mathbf{s})$. For individual pixels this is not reliable due to inaccurate normal estimation. As albedo is constant an accurate estimate can be obtained by averaging over each region

$$A_r = \frac{1}{|r|} \sum_{(x, y) \in r} \frac{L(x, y)}{(\mathbf{n}(x, y) \cdot \mathbf{s})} \quad (8)$$

where $r \in R$. This requires a field of surface normals; as it is reasonably robust to noise this may be obtained directly from the stereo algorithm.

	Boot Strap	Smoothed Boot Strap	Our Algorithm
Frame	1.62(13.5%)	1.22(2.2%)	1.08(1.0%)
Head	1.84(13.7%)	1.55(0.2%)	1.90(1.9%)
Head (centre)	1.73(10.0%)	1.47(0.1%)	1.40(0%)

Table 1: Statistics, see text for details.

5 Shape from Shading Algorithm

Equation 1 only constrains the angle between the surface normal and light source, i.e. surface normals must lie on a cone whose angle is defined by the ratio L/A . To remove this ambiguity we introduce two constraints, which are a) that surface normals should vary smoothly across the surface, and b) at occluding boundaries, the surface normals lie in the image plane and point away from the boundary. We adopt the Worthington and Hancock[20] algorithm to solve for the field of surface normals by alternately smoothing and re-projecting onto the cone. Applying this method gives us a fields of surface normals for either image, the framework only uses the left images orientation information however.

6 Stereo Algorithm

The DSI of a single pixel can not be accurately represented with a single Gaussian. A disparity value and its confidence can be however, hence the need for a stereo algorithm to select a *reasonable* disparity for each pixel. A modified version of the algorithm of Meerbergen et al.[15] is used. The modification is such that, in addition to the best disparity, it also outputs all other disparities within a given tolerance of the best, as an indication of confidence. It uses the Birchfield and Tomasi’s[2] sampling invariant dissimilarity measure, the resulting disparities can therefore be considered as ranges, ± 0.5 the given value, rather than as infinitesimal points. A Gaussian can therefore be fitted to each pixels set of disparities for use by the Gaussian belief propagation step. Occluded pixels with no disparities are assigned an evidence of $\psi_i(x_i, y_i) = \phi[0, 0]$.

Both the SfS initialisation and albedo estimation steps require surface normals to be extracted from the initial disparity map, this may be done with camera calibration information. Directly estimating surface normals by differentiating a *discrete* depth-map does not work however. Therefore the belief propagation process is run to obtain an initial smooth surface; for this first run orientation is provided by plane fitting the uniform colour segments. To further reduce noise least squares planes are fitted to a 11×11 window around each normal and the plane perpendiculars used; this is necessary to obtain a robust albedo estimate.

7 Experimental Results

We have evaluated our method using a number of stereo pairs with ground truth data. Standard stereo tests[17] are not fit for our purposes, in part because they do not match the single known light source requirement but also because they give ground truth in terms of either discrete disparities or fitted planes. As we obtain surfaces to a much finer resolution we need ground truth data with disparity maps at sub-pixel resolution.

Using a Cyberware 3030 head scanner and two Canon S70s in the standard stereo position a data set with ground truth data has been captured. It was calibrated both before and after the capture session with a 3D calibration target. The stereo pair backgrounds are

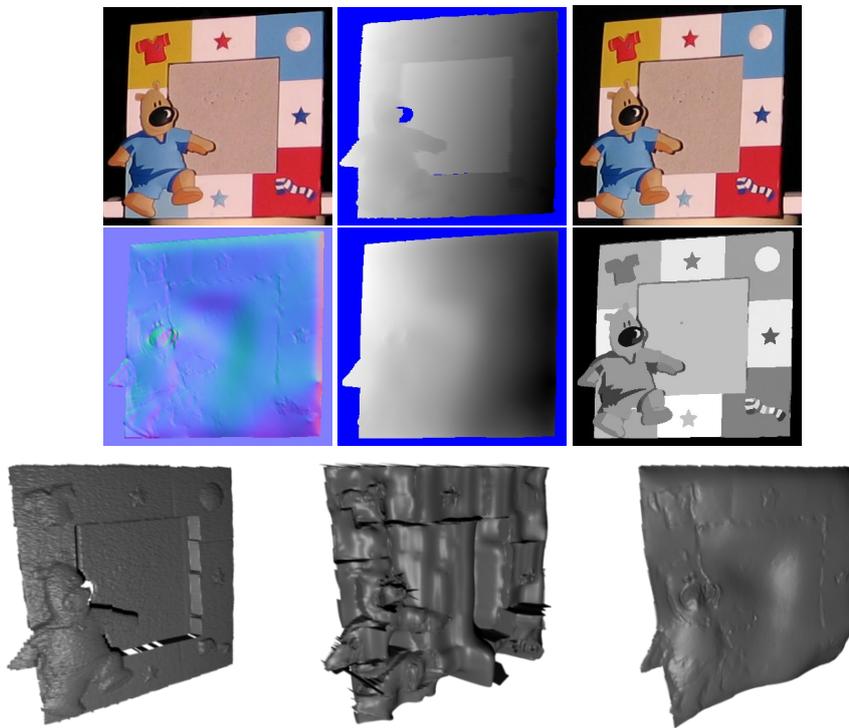


Figure 1: Results for a photo frame, see text for details.

masked out; the ground truth disparity map for the human head is masked out in problem areas, such as eyes and hair.

We illustrate the algorithm with the picture frame given in figure 1 and the head given in figure 2. The figures are arranged as left image, ground truth disparity then right image on the first row, output orientation map, disparity map and albedo map on the second row. The final row contains renders of the 3D models, first the ground truth, then the smoothed boot strap and finally the output. The frame is smoothed considerably by the algorithm. Whilst some overall structure has been lost details not visible in the initialisation are apparent, primarily the decoration on the frame. It additionally shows the effectiveness of the albedo estimation. The algorithm produces a reasonable visual result for the head, unlike the bootstrap algorithm. Again, it captures details not visible in the bootstrap.

Table 1 compares the algorithm statistically, with the boot strap algorithm[15] in the first column then the smoothed version used for albedo estimation followed by the final result. We provide two values in each case. The first is the average disparity difference from ground truth for pixels classified as inliers, the second is in brackets and is the percentage of outliers. We define inliers as disparity values within 8 pixels of the ground truth disparity. For the frame the results are clear cut, but for the head the numbers indicate that our algorithm has made it worse, though the renders indicate otherwise. For the smoothed version the error is equally distributed, but for the output from our algorithm the error is primarily in the ears and edges of the face. This is because the lambertian shading model is insufficient in these regions. The *head (centre)* row shows the statistics

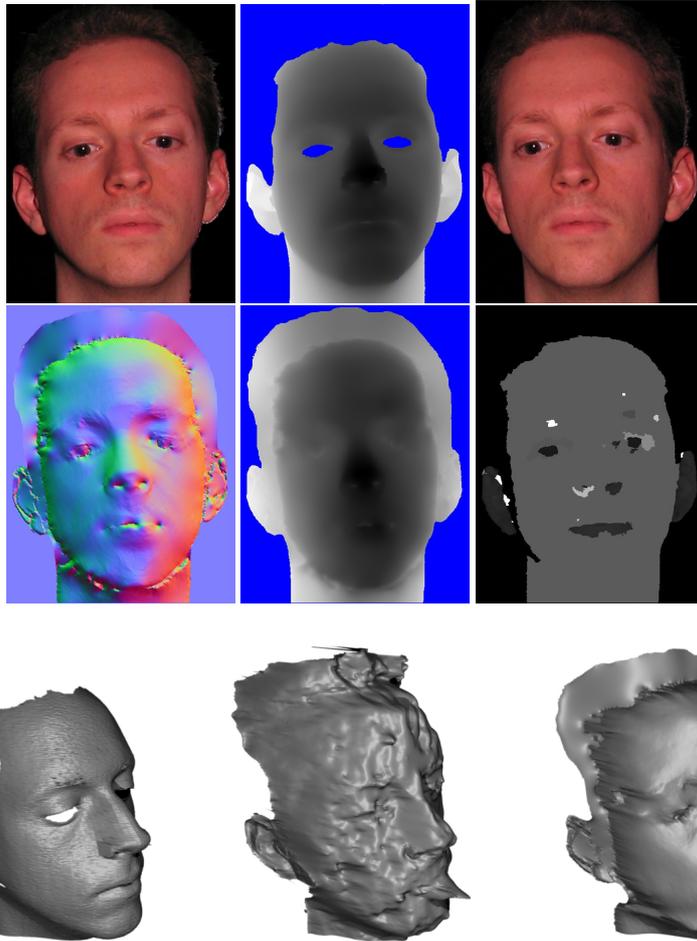


Figure 2: Results for a head, see text for details.

when the ears and edges of the head are masked out.

8 Conclusions

We have presented a method of integrating shape from shading information with stereo information using Gaussian belief propagation. This method efficiently delivers a continuous estimate of disparity and is relatively easy to implement. Our results show an improvement in the fine surface details when shading information is used, leading to more accurate and visually pleasing models.

Much possible future work exists in this area. The greatest weakness of this approach is SfS requiring a single known light source. Two future directions may be found in using another source of orientation information or removing this constraint from SfS, with light source estimation and support for multiple light sources.

A Gaussian Belief Propagation

Multiplying, we get

$$\phi[\mathbf{P}_1\boldsymbol{\mu}_1, \mathbf{P}_1]\phi[\mathbf{P}_2\boldsymbol{\mu}_2, \mathbf{P}_2] = \phi[\mathbf{P}_1\boldsymbol{\mu}_1 + \mathbf{P}_2\boldsymbol{\mu}_2, \mathbf{P}_1 + \mathbf{P}_2] \quad (9)$$

If we add an additional independent variable, we get

$$\text{Ext}(\phi[\mathbf{P}\boldsymbol{\mu}, \mathbf{P}]) = \phi\left[\begin{pmatrix} \mathbf{P}\boldsymbol{\mu} \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{P} & 0 \\ 0 & 0 \end{pmatrix}\right] \quad (10)$$

Finally, if we marginalise over the first variable, we get

$$\text{Marg}_1(\phi[\mathbf{P}\boldsymbol{\mu}, \mathbf{P}]) = \phi[\mathbf{h}_2 - \mathbf{P}_{12}\mathbf{P}_{11}^{-1}\mathbf{h}_1, \mathbf{P}_{22} - \mathbf{P}_{12}\mathbf{P}_{11}^{-1}\mathbf{P}_{12}^T] \quad (11)$$

where $\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{12}^T & \mathbf{P}_{22} \end{pmatrix}$ and $\mathbf{P}\boldsymbol{\mu} = \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{pmatrix}$

Combining the local distribution with previous messages:

$$\begin{aligned} \psi_t(x_t, y_t) \prod_{u \in N/s} m_{ut}^{(n-1)}(x_t) &= \phi[P_{tt}\boldsymbol{\mu}_{tt}, P_{tt}] \prod_{u \in N/s} \phi[P_{ut}\boldsymbol{\mu}_{ut}, P_{ut}] \\ &= \phi[P_{tt}\boldsymbol{\mu}_{tt} + \sum_u P_{ut}\boldsymbol{\mu}_{ut}, P_{tt} + \sum_u P_{ut}] \\ &= \phi[P_0\boldsymbol{\mu}_0, P_0] \end{aligned} \quad (12)$$

Extending the distribution to incorporate x_s , we get

$$\text{Ext}(\phi[P_0\boldsymbol{\mu}_0, P_0]) = \phi\left[\begin{pmatrix} P_0\boldsymbol{\mu}_0 \\ 0 \end{pmatrix}, \begin{pmatrix} P_0 & 0 \\ 0 & 0 \end{pmatrix}\right]$$

Then combining with $\psi_{st}(x_s, x_t)$:

$$\psi_{st}(x_s, x_t)\text{Ext}(\phi[P_0\boldsymbol{\mu}_0, P_0]) = \phi\left[\begin{pmatrix} P_0\boldsymbol{\mu}_0 - P_n z_{ts} \\ P_n z_{ts} \end{pmatrix}, \begin{pmatrix} P_n + P_0 & -P_n \\ -P_n & P_n \end{pmatrix}\right]$$

Finally, we marginalise to find the new message

$$\begin{aligned} m_{ts}^{(n)}(x_s) &= \alpha \text{Marg}_1\left(\phi\left[\begin{pmatrix} P_0\boldsymbol{\mu}_0 - P_n z_{ts} \\ P_n z_{ts} \end{pmatrix}, \begin{pmatrix} P_n + P_0 & -P_n \\ -P_n & P_n \end{pmatrix}\right]\right) \\ &= \phi[P_n z_{ts} + P_n(P_n + P_0)^{-1}(P_0\boldsymbol{\mu}_0 - P_n z_{ts}), P_n - P_n(P_n + P_0)^{-1}P_n] \end{aligned}$$

so the message update rules are

$$\begin{aligned} P_{ts} &\leftarrow P_n - P_n(P_n + P_0)^{-1}P_n \\ P_{ts}\boldsymbol{\mu}_{ts} &\leftarrow P_n z_{ts} + P_n(P_n + P_0)^{-1}(P_0\boldsymbol{\mu}_0 - P_n z_{ts}) \end{aligned} \quad (13)$$

References

- [1] A. A. Amini, T. E. Weymouth, and R. C. Jain. Using dynamic programming for solving variational problems in vision. *Pattern Analysis and Machine Intelligence*, 12(9):955–867, 1990.

- [2] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *Pattern Analysis and Machine Intelligence*, Vol. 20, No. 4:401–406, 1998.
- [3] A. F. Bobick and S. S. Intille. Large occlusion stereo. *International Journal of Computer Vision*, 33(3):181–200, 1999.
- [4] Y.i Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.
- [5] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5:603–619, 2002.
- [6] R. Cowell. *Learning in Graphical Models*, chapter Advanced Inference in Bayesian Networks. MIT Press, 1998.
- [7] J. E. Cryer, P. S. Tsai, and M. Shah. Integration of shape from shading and stereo. *Pattern recognition*, 28(7):1033–1043, 1995.
- [8] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *IEEE CVPR*, 1:261–268, 2004.
- [9] R.T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *Pattern Analysis and Machine Intelligence*, 10:439–451, 1988.
- [10] P. Fua and Y. G. Leclerc. Object-centered surface reconstruction: combining multi-image stereo and shading. *International Journal of Computer Vision*, 16(1):35–56, 1995.
- [11] B. K. P. Horn. *The Psychology of Computer Vision*, chapter Obtaining shape from shading information, pages 115–155. McGraw Hill, 1975.
- [12] K. Ikeuchi and B.K.P. Horn. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17(3):141–184, 1981.
- [13] H. Jin, A. Yezzi, and S. Soatto. Stereoscopic shading: Integrating multiframe shape cues in a variational framework. In *CVPR*, volume 1, pages 169–176, 2000.
- [14] Y. G. Leclerc and A. F. Bobick. The direct computation of height from shading. In *CVPR*, pages 552–558, 1991.
- [15] G. Van Meerbergen, M. Vergauwen, M. Pollefeys, and L. Van Gool. A hierarchical symmetric stereo algorithm using dynamic programming. *International Journal of Computer Vision*, Vol. 47:275–285, 2002.
- [16] D. Samaras, D. Metaxas, P. Fua, and Y. G. Leclerc. Variable albedo surface reconstruction from stereo and shape from shading. In *CVPR*, volume 1, pages 480–487, 2000.
- [17] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1):7–42, 2002.
- [18] M. Shao, R. Chellappa, and T. Simchony. Reconstructing a 3-d depth map from one or more images. *CVGIP: Image Understanding*, 53(2):219–226, 1991.
- [19] Y. Weiss and W. T. Freeman. Correctness of belief propagation in gaussian graphical models of arbitrary topology. *Neural Computation*, 13(10):2173–2200, 2001.
- [20] P.L. Worthington and E.R. Hancock. New constraints on data-closeness and needle map consistency for shape-from-shading. *Pattern Analysis and Machine Intelligence*, 21(12):1250–1267, 1999.
- [21] J.S. Yedidia, W.T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *Information Theory*, 51(7):2282–2312, 2005.