Estimating uncertainty in dense stereo disparity maps

A. Blake, P.H.S. Torr, I.J. Cox and A. Criminisi

March 2003

Technical Report
MSR-TR-2003-93

Dense stereo is a well studied problem in computer vision. Generally dense stereo algorithms provide only a single estimate of disparity, ignoring uncertainty in the disparity map. Here however, we present a new, linear-time, exact method for recovering entire distributions for disparity at all pixels. This is accomplished by using a recent extension, due to Durbin et al., of the well-known forward-backward algorithm to work with two unsynchronised input streams, rather than just one as in the conventional case. The two input streams, in the stereo context are simply two corresponding epipolar lines, one from each stereo image. Specifically we consider the problem of view interpolation. The availability of a distribution over disparity is particularly appealing here. In that case, disparities themselves are not the required end product, but merely an intermediate representation. It is therefore unnecessary to estimate a unique disparity map. Instead, the image intensity at each cyclopean pixel can be estimated as a mean of predicted intensities for all possible disparities. These principles are illustrated for a teleconferencing application: enabling eye contact by positioning a virtual camera at the centre of a display screen used in a two-way conference. We show that the new approach can significantly improve the quality of the interpolated cyclopean image, compared with using unique estimated disparities.

Microsoft Research Ltd.
7 J J Thomson Ave
Cambridge, UK CB3 0FB
http://www.research.microsoft.com

1 Introduction

It is well known that stereo image matching is a hard problem, and much progress has been made over the past 25 years or so. The variety of algorithms for dense stereo matching is too extensive to list here but has recently been comprehensively documented and evaluated [14]. None of the algorithms reported and investigated there deal with uncertainty in disparity. However, representation of uncertainty is of critical importance for robust estimation and data fusion.

According to the evaluation of stereo algorithms [14], two of the most powerful algorithms use respectively graph cuts [12, 7] and loopy propagation [15] but these approaches are currently too computationally intensive for real time applications. Conversely, Epipolar line Dynamic Programming [8, 4] is efficient and offers the best hope for real time applications. For this reason, we focus on epipolar line algorithms, and seek a generalisation that represents the uncertainty in disparities.

We illustrate the new approach with the synthesis of the *cyclopean* view [6], the view from a virtual camera placed symmetrically, midway between the (calibrated) cameras of a physical stereo pair. Previously proposed solutions to this problem can be broadly categorized as model-based or stereo-based. One model-based approach is to use a detailed head model and reproject it into the cyclopean view; whilst this can be successful [16, 17], it is limited to imaging heads, and would not, for example, deal with a hand in front of the face. A more general approach therefore is to use low level stereo matching.

Given left and right images $\mathbf{L}(\mathbf{x})$ and $\mathbf{R}(\mathbf{x})$, over the image domain $\mathbf{x} \in D \subset \mathbf{R}^2$, the cyclopean image $\mathbf{I}(\mathbf{x})$ cannot be generated by any simple pointwise interpolation of $\mathbf{L}$ and $\mathbf{R}$. In the absence of occlusion, one can think of the right image and left images as symmetrically warped forms of the cyclopean image, so that

$$\mathbf{I}(\mathbf{x}) = \mathbf{R}\left((\mathbf{x}) - \frac{1}{2}\mathbf{d}(\mathbf{x})\right) = \mathbf{L}\left(\mathbf{x} + \frac{1}{2}\mathbf{d}(\mathbf{x})\right), \tag{1}$$

in which case the cyclopean image could be generated from the left or right image using an estimate of the disparity field $\mathbf{d}$. This requires the disparity field $\mathbf{d}$ to be known, for example by dense pointwise matching (under epipolar constraints) of corresponding points in $\mathbf{L}$ and $\mathbf{R}$. However, three-dimensional occlusion of surface points means that $\mathbf{d}(\mathbf{x})$ cannot necessarily be defined for all pixels $\mathbf{x} \in D$, which is a problem for cyclopean reconstruction. An elegant account of the stereo matching problem, in a cyclopean frame and with occlusion, is given by Belhumeur [1]. It is of course impossible to reconstruct, entirely accurately, the cyclopean view of a patch that is occluded in one or other physical camera; all that can be done is to apply a default assumption about the shape of such a patch.

It has also been noted by others [13] that a good reconstruction of the cyclopean view can often be achieved, even when the underlying disparities are ambiguous. This is acknowledged in the approach taken here: its aim is to improve the robustness of reconstruction by treating the disparity field $\mathbf{d}$ as a first-class random variable, rather than merely as a fixed (unknown) parameter.

If $\mathbf{d}$ were taken to be a parameter, then estimation of the cyclopean image would proceed in two stages. First the estimation of the warp field, for example by MAP estimation [1]:

$$\hat{\mathbf{d}} = \arg\min_{\mathbf{d}} p(\mathbf{d} \mid \mathbf{z}) \ \text{ where } \ \mathbf{z} = (\mathbf{L}, \mathbf{R}), \tag{2}$$

followed by deterministic reconstruction of cyclopean intensities:

$$\hat{\mathbf{I}} = \mathbf{I}(\hat{\mathbf{d}}, \mathbf{z}). \tag{3}$$

Instead, treating $\mathbf{d}$ as a random variable, there is no need for premature commitment to a specific value of $\mathbf{d}$. Then the aim of inference is to estimate a posterior distribution $p(\mathbf{d} \mid \mathbf{z})$. The cyclopean image can be estimated robustly as $\hat{\mathbf{I}}$, the posterior mean over the disparity distribution:

$$\hat{\mathbf{I}} = \mathcal{E}[\mathbf{I}|\mathbf{z}] = \mathcal{E}_{\mathbf{I}}\left[\mathcal{E}_{\mathbf{d}|\mathbf{z}}[\mathbf{I}|\mathbf{d}, \mathbf{z}]\right]. \tag{4}$$

Bayesian estimate of cyclopean intensity

Given the observation $\mathbf{z} = (\mathbf{L}, \mathbf{R})$, and knowing the observer model $p(\mathbf{z} \mid \mathbf{I}, \mathbf{d})$, it is required to estimate the cyclopean image $\hat{\mathbf{I}}$, via the posterior $p(\mathbf{d} \mid \mathbf{z})$ for disparity. The full solution has the following steps.

1. Marginalise over $\mathbf{I}$ to get $p(\mathbf{z} \mid \mathbf{d})$; for Gaussian noise processes this can be done analytically [1]
2. Bayes: $p(\mathbf{d} \mid \mathbf{z}) \propto p(\mathbf{z} \mid \mathbf{d})p(\mathbf{d})$
3. Bayes: $p(\mathbf{I} \mid \mathbf{d}, \mathbf{z}) \propto p(\mathbf{z} \mid \mathbf{I}, \mathbf{d})p(\mathbf{I})$
4. Marginalise over $\mathbf{d}$: $p(\mathbf{I} \mid \mathbf{z}) \propto \int_{\mathbf{d}} p(\mathbf{I} \mid \mathbf{d}, \mathbf{z})p(\mathbf{d} \mid \mathbf{z})$
5. Compute expectations, as above (4), to obtain $\hat{\mathbf{I}}$.

Of these, the substantial step is the second one. This is where, if only the mode of $p(\mathbf{d} \mid \mathbf{z})$ were required, the Viterbi algorithm along epipolar lines could be used [1]. To obtain the full distribution, a variation of message passing [9] or Forward-Backward inference [11] is used, and details follow.

2 Notation and framework

The framework used from here onward, is based on matching pairs of epipolar lines. Intensity functions $\mathbf{L}$ and $\mathbf{R}$ will now refer to corresponding epipolar lines from left and right images respectively. The objective is to infer the intensity function $\mathbf{I}$ for the corresponding line in the cyclopean image.

Images and disparities The intensity function in the left epipolar line is

$$\mathbf{L} = \{L_m,\ m = 0, \ldots, N\}$$

and in the right it is

$$\mathbf{R} = \{R_n,\ n = 0, \ldots, N\}.$$

The cyclopean epipolar line is

$$\mathbf{I} = \{I_k,\ k = 0, \ldots, 2N\},$$

doubly oversampled, for reasons that will become apparent. Stereo disparity is a vector $\mathbf{d} = \{d_k,\ k = 0, \ldots, 2N\}$ with components expressed in cyclopean coordinates. Stereo disparity $\mathbf{d}$ in turn induces a vector $\mathbf{h} = \{h_k,\ k = 0, \ldots, 2N\}$ of image warps from cyclopean coordinates into the coordinates of the left and right images:

$$h_k = (m(k), n(k)) = \left(\frac{1}{2}(k - d_k),\ \frac{1}{2}(k + d_k)\right) \tag{5}$$

and conversely, stereo disparity can be expressed as the coordinate difference $d_k = n - m$ — see figure 1. The cyclopean coordinate k corresponding to left and right coordinates m, n is simply $k = m + n$. The warp function $\mathbf{h}$ associates image intensities at locations $k, m(k), n(k)$ in the cyclopean, left and right images. The precise nature of this association, which is probabilistic, is described later.

Even and odd warps When $k + d_k$ is even, $(m(k), n(k))$ are integers, so that h_k warps directly onto the left and right coordinate grids. In this case the warp is termed to be even. For odd warps however, there is no such direct mapping, so odd warps in the cyclopean image are deemed to be transitory states, and this is made clearer below.

Disparity process The space of matches between images is constrained and conditioned by a probability distribution $P(d_k|d_{k-1})$ specifying a random process for disparities at successive locations on the cyclopean epipolar line. It is assumed that, a priori, d_k depends directly on the immediately preceding disparity d_{k-1}.

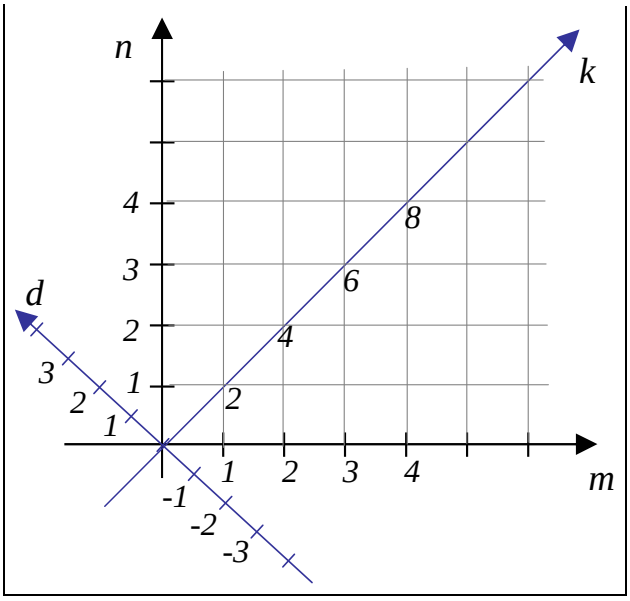


Figure 1: **Disparity and the cyclopean image** The diagram shows left and right epipolar lines with pixel coordinates m, n, cyclopean coordinates k and disparity vectors d as displacements orthogoal to the cyclopean coordinate axis.

Observations The data $\mathbf{z}$ from which cyclopean image inference is made consists of the left and right scan lines, ie

$$\mathbf{z} = (\mathbf{L}, \mathbf{R}). \tag{6}$$

Each image typically contains color pixels, i.e. $L_m \in \mathcal{R}^3$. (In a more general setting they could be other features $L_m \in \mathcal{R}^f$ derived from filtering a group of pixels, for example to obtain improved invariance to illumination variations or nonuniform camera sensitivity.) Left/right pixels are assumed to have been generated by adding noise to the value of the corresponding cyclopean pixel, so

$$L_m = I_k + \mu_m \text{ and } R_n = I_k + \nu_n, \tag{7}$$

where μ_m and ν_n are noise random variables, assumed i.i.d. and Gaussian across all m, n (and isotropic if pixels are color values).

Observation density The joint likelihood for observations is denoted $p(\mathbf{z}|\mathbf{I}, \mathbf{d})$ and can be decomposed into a product over pixels and pixel pairs using the independence of μ_m and ν_n. Details are given later.

Observation history The history of observations "preceding" the match h_k is denoted

$$\mathbf{z}_{h_k} \equiv \left(L_1, \ldots, L_{m(k)}, R_1, \ldots, R_{n(k)}\right) \tag{8}$$

and the complementary set of (future) observations is denoted

$$\mathbf{z}^{h_k} \equiv \left(L_{m(k)+1}, \ldots, L_N, R_{n(k)+1}, \ldots, R_N\right). \tag{9}$$

Image prior The marginal prior distribution for cyclopean image values is denoted $p_0(I|k)$ and useful special cases include the stationary case $p_0(I|k) = p_0(I)$ and the uniform case $p_0(I) = 1/V$, where V is the volume of color-space. Prior p_0 could be modelled as a Gaussian or mixture of Gaussians learned from patches in appropriate neighbourhoods of the left and right images. In any case, all pixels in the cyclopean image are taken to have mutually independent prior distributions.

3 Process and observation models

Process model The process model specifies the distribution of disparity gradients[1] $\delta_k = d_k - d_{k-1}$ [2, 10]. Varying degrees of subtlety can be incorporated; here we take the simplest reasonable model, restricting $|\delta_k| \leq 1$ (ordering constraint) and penalising the occlusion case $|\delta_k| = 1$. This gives a process distribution

$$P_d(d_k|d_{k-1}) = \begin{cases} 1 & \text{if } h_{k-1} \text{ is odd and } \delta_k = 0 \\ q & \text{if } h_{k-1} \text{ is even and } |\delta_k| = 1 \\ 1-2q & \text{if } h_{k-1} \text{ is even and } \delta_k = 0 \end{cases} \tag{10}$$

with some appropriate occlusion probability $q < 1/3$. The oddness condition forces odd warps to be treated correctly, as a transitory condition resolved by an immediate step to an even warp.

Observation model The observation density $p(\mathbf{z}|\mathbf{d})$ is defined as a marginalisation of the full conditional density with respect to intensity:

$$p(\mathbf{z}|\mathbf{d}) \equiv \int_{\mathbf{I}} p(\mathbf{z}|\mathbf{I}, \mathbf{d}) p_0(I) \tag{11}$$

and decomposes (it can be shown) as a product:

$$p(\mathbf{z}|\mathbf{d}) = \prod_k f(\mathbf{z}; d_k, d_{k-1}) \tag{12}$$

[1] In [10] disparity gradient is defined to be $2\delta_k$, giving a limiting gradient of $|\delta_k| = 2$ in the occluding case.

where the observation function f is defined

$$f(\mathbf{z}; d_k, d_{k-1}) = \begin{cases} p\left([\mathbf{z}_{h_k} \backslash \mathbf{z}_{h_{k-1}}] \mid d_k, d_{k-1}\right) & \text{if } h_k \text{ is even} \\ 1 & \text{otherwise} \end{cases} \tag{13}$$

(where $\backslash$ denotes set difference). In the case of odd warps, $[\mathbf{z}_{h_k} \backslash \mathbf{z}_{h_{k-1}}] = \emptyset$, and hence the observation function must be defined to be unity. It can be shown that, for the observation model above, f has the form[2]

$$f(\mathbf{z}; d_k, d_{k-1}) = \begin{cases} \left(\frac{\lambda}{\pi}\right)^{s/2} \exp -\lambda \|L_{m(k)} - R_{n(k)}\|^2 & h_k \text{ even}, \quad \delta_k = 0 \\ p_0(L_{m(k)}) & h_k \text{ even}, \quad \delta_k = -1 \\ p_0(R_{n(k)}) & h_k \text{ even}, \quad \delta_k = 1 \\ 1 & h_k \text{ odd}. \end{cases}$$

where $s = 1$ for monochrome observations and $s = 3$ for color.

4 Forward-backward algorithm

The conventional forward-backward algorithm [11] has been generalised to matching problems, initially for matching genome sequences, by Durbin et al. [5], and is used here to deal with the stereo problem. Conventionally, the history of observations corresponding to the k^{th} step would depend only on k, having the form $z_1, \ldots, z_k$ for some appropriately defined observations k. In the stereo problem this has to be generalised to a history $\mathbf{z}_{h_k}$ (8) which depends not only on the "time" k, but also on the *value* d_k of the inferred quantity at time k. A forward-backward algorithm for this generalised setting follows (with a derivation in the appendix).

Forward algorithm Define forward probabilities

$$\alpha_k(d_k) = P(d_k | \mathbf{z}_{h_k}). \tag{14}$$

They are generated iteratively (see appendix for derivation) as

$$\alpha_k(d_k) \propto \sum_{d_{k-1}} \alpha_{k-1}(d_{k-1}) p(d_k | d_{k-1}) f(\mathbf{z}; d_k, d_{k-1}) \tag{15}$$

[2] In fact the equation given here is an approximation, in the interests of simplicity, for the case that the prior for I is weak.

and note that sparsity (10) of $p(d_k|d_{k-1})$ means that the sum is evaluated over at most 3 values of d_{k-1}.

Backward algorithm Backward probabilities are defined to be

$$\beta_k(d_k) = p(\mathbf{z}^{h_k} \mid d_k) \tag{16}$$

and propagated as

$$\beta_k(d_k) \propto \sum_{d_{k-1}} \beta_{k+1}(d_{k+1}) p(d_{k+1} \mid d_k) f(\mathbf{z}; d_{k+1}, d_k). \tag{17}$$

Posterior distributions Finally, forward and backward variables could be combined conventionally [11] to compute the posterior marginal distributions of disparity

$$\gamma_k(d_k) \equiv p(d_k \mid \mathbf{z}) = \frac{\tilde{\gamma}_k(d_k)}{\sum_{d_k} \tilde{\gamma}_k(d_k)} \tag{18}$$

where

$$\tilde{\gamma}_k(d_k) = \alpha_k(d_k)\ \beta_k(d_k). \tag{19}$$

However, this is not enough to estimate cyclopean intensity, because of the need to allow for occlusion. We need not only the marginal density, but the *joint* density of the disparity d_k with the disparity gradient δ_k. This joint density $P(d_k, d_{k-1}|\mathbf{z})$ is also available [11] from forward and backward probabilities, yielding:

$$\xi_k(d_k, \delta_k) \equiv p(d_k, \delta_k \mid \mathbf{z}) = \frac{\tilde{\xi}_k(d_k, \delta_k)}{\sum_{d_k, \delta_k} \tilde{\xi}_k(d_k, \delta_k)} \tag{20}$$

where

$$\begin{aligned} &\tilde{\xi}_k(d_k, \delta_k) \\ &\quad = \alpha_{k-1}(d_k - \delta_k)\ P_d(d_k \mid d_k - \delta_k)\ \beta_k(d_k)\ f(\mathbf{z}; d_k, d_{k-1}). \end{aligned} \tag{21}$$

5 Estimation of cyclopean image intensity

Having obtained the probability distribution for disparities, $p(d_k, \delta_k \mid \mathbf{z})$, this can now be used to estimate cyclopean image intensities. At cyclopean pixel k we estimate the intensity I_k as an expectation, first over the observation noise variables, then over the posterior distribution for

disparity. The expectation over noise variables is denoted $\tilde{I}_k(d_k, \delta_k)$ and in the simplest case of uniform intensity prior $p_0(I)$ is given by

$$\tilde{I}_k(d_k, \delta_k) \equiv \mathcal{E}[I_k|h_k] = \begin{cases} \frac{1}{2}\left(L_{m(k)} + R_{n(k)}\right) & \text{if } \delta_k = 0 \\ L_{m(k)} & \text{if } \delta_k = -1 \\ R_{n(k)} & \text{if } \delta_k = +1, \end{cases} \tag{22}$$

where the three conditions cover the cases of: no occlusion, right occlusion, left occlusion. There are variations in these formulae for the case of Gaussian priors. The formula is defined strictly only when the warp is even (so m, n are integers). This gives two options: either cyclopean intensities must be estimated from the even warps only, or left/right intensities must be interpolated for fractional indices, for example:

$$L_{m+1/2} = \frac{L_m + L_{m+1}}{2}. \tag{23}$$

Finally, cyclopean intensity is estimated as the posterior expectation (4), so that the estimated cyclopean intensity at site k is:

$$\hat{I}_k \equiv \mathcal{E}_{p(\mathbf{d})}[\tilde{I}_k(d_k, \delta_k)] = \sum_{d_k, \delta_k} \tilde{I}_k(d_k, \delta_k)\xi_k(d_k, \delta_k). \tag{24}$$

As this sum includes odd warps, interpolated left/right intensities (23) must be used where needed.

6 Application and results

Dynamic programming (DP) stereo [4], which has previously been demonstrated for cyclopean view interpolation [3] in video, nowadays runs at around 10 frames/sec on a 2 GHz Pentium computer. To obtain that speed, observations consist of single, monochrome pixels, and the consequent quality of reconstruction is not consistently satisfactory for teleconferencing, as figure 2 shows. Two kinds of error are clearly visible: (i) artefacts produced by mismatches, and (ii) defects introduced where the background is occluded in one or other view. The Forward-Backward algorithm is capable of alleviating errors of the first kind, as we will demonstrate.

In this section comparisons are made between cyclopean intensity images computed from

DP: the Dynamic Programming algorithm[3] of [3, 4], and

[3]Our realisation of the DP algorithm implements the same model as described here for FB, including the strategy of (22) for rendering occluded areas.

FB: the forward-backward algorithm proposed here ,

using exactly the same model and parameters q, λ in each case. Note that whereas the behaviour of FB depends on both q and λ, the behaviour of DP depends only on the single parameter[4]

$$c = \frac{1}{\lambda}\left(\log\left((1-2q)/q\right) - \frac{1}{2}\log\left(\frac{\lambda}{\pi}\right)\right) \tag{25}$$

This reflects the fact that all posterior distributions for $\mathbf{d}$ with a given value of $c(q,\lambda)$ share the same posterior mode even though other properties of those posterior distributions differ. (One view of the additional parameter is that it controls the degree to which the probability density for disparity spreads about its mode.)

The performance of both algorithms is considerably improved by replacing the pixelwise square-difference measure $||L_{m(k)} - R_{n(k)}||^2$ by an average square-difference

$$\frac{1}{N^2}||L_{m(k)} - R_{n(k)}||^2_{N\times N}$$

over a square window, and for this study we have settled on a window of size $N = 5$. There are of course further useful elaborations that can be made to the measure, for example using color and various normalisations, and possibly a larger window still, at increased computational cost. However, the purpose of this study is to compare algorithms, so it is the comparison of FB with DP that is of paramount interest, more than the absolute performance of either.

Note that the computational costs of DP and FB are closely comparable. The forward pass of FB has a similar structure to the forward pass of DP (Viterbi). The backward pass of DP consists of pointer following and so has negligible cost. The backward pass (17) of FB would appear to have a cost similar to the forward pass, but in fact the likelihood $f(\ldots)$ terms in (17) can be cached and re-used from the forward pass, so the backward pass has negligible cost for FB also.

Comparisons of FB with DP are given in figures 3 and 4. For practical convenience, in place of λ we specify a noise parameter σ in gray-level units, so that:

$$\lambda = \frac{1}{2(\sigma/255)^2} \tag{26}$$

It is clear from inspection that artefacts of reconstruction in the DP algorithm are partly suppressed by the FB algorithm. The mechanism by which this is done is illustrated in figure 6 showing how ambiguity of disparity is represented as a probability distribution. Note the bifurcation of paths in

[4] In [4] c is referred to in the pseudo-code as the constant **Occlusion**.

Left Virtual Cyclopean (DP, $c = 0.005$) Right

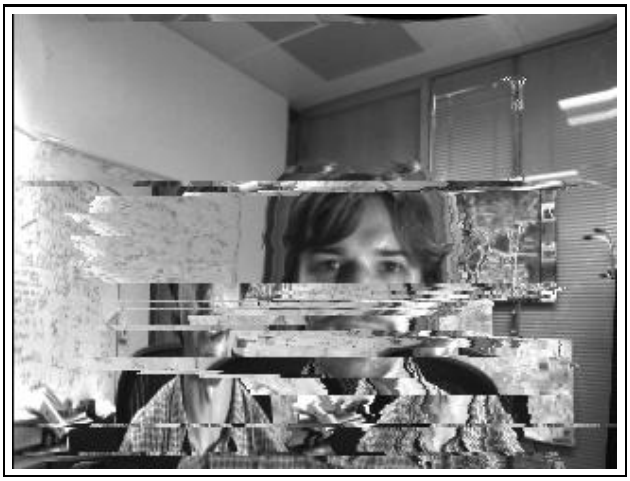
Figure 2: **Fast cyclopean view synthesis by dynamic programming** Left, right and typical estimated cyclopean image using dynamic programming with single pixel, monochrome observations. Note that gaze is correct in the cyclopean view. The algorithm runs at near real-time rate, but can produce artefacts in the cyclopean image — in this case texture is broken up at the top of the head, and in the background to the right of the head. (See later for explanation of occlusion constant c.)

$q = 0.02, \sigma = 5$ $q = 0.1, \sigma = 6$ $q = 0.3, \sigma = 7$

Figure 3: **Forward-Backward cyclopean interpolation** for the data of figure 2. The three sets of (q, σ) parameters used here all correspond to occlusion parameter $c = 0.005$ as used for DP in figure 2. Note that i) performance of FB appears superior to that of DP in that artefacts are reduced, and ii) performance of FB appears relatively insensitive to changing parameters.

the disparity distribution, typical in the presence of ghost matches. Such bifurcation could further be used to signal failures in stereo matching.

Finally, figure 5 shows a comparison with ground truth, using the "Tsukuba" calibrated motion sequence of [13]. The first and last images of the sequence are used to provide the baseline, and the centre image gives cyclopean ground truth. This is even so when the window size is reduced to 1×1 pixels as here. Note that FB peforms consistently better than DP on values of the c parameter



DP: $c = 0.005$

DP: $c = 0.01$

FB: $c = 0.005$ ($q = 0.1, \sigma = 6$)

FB: c=0.01 ($q = 0.1, \sigma = 8$)

Figure 4: **Forward-Backward cyclopean interpolation** The first and second row show comparable results for cyclopean images from FB and DP, for two different settings of parameter c. FB appears to perform better than DP, to some extent for $c = 0.01$ and to a substantial extent for $c = 0.005$.

tested. However, this data, along with all 4 ground truth sequences in [13], rank as very easy tests compared with the our teleconferencing data, so both FB and DP obtain good reconstructions. This is because the range of disparities is about 10 times greater in the teleconference data. At the time of writing, there is no publically available ground truth data for this much more difficult problem setting

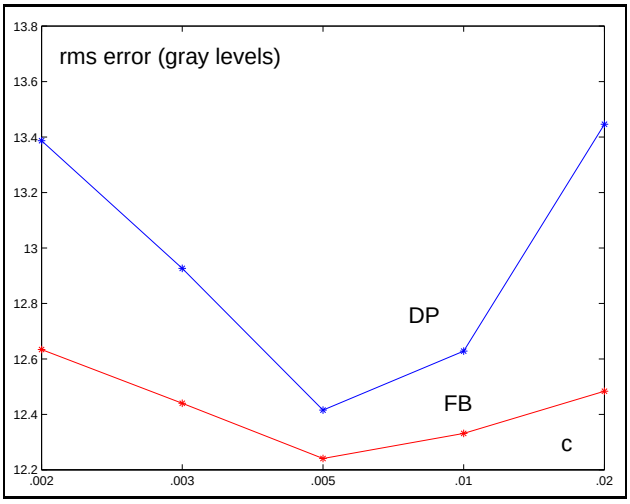


Figure 5: **Testing against cyclopean ground truth.** Testing ground truth with the "Tsukuba" data (see text for details) shows a consistent improvement for FB over DP, and reduced sensitivity to the value of parameter c.

7 Conclusion

This paper describes an algorithm for estimating uncertainty in dense disparity from corresponding epipolar lines in stereo images. The algorithm allows ambiguity of disparity to be represented explicitly, as a probability distribution. This has been achieved by using an extension due to [5] of the forward backward algorithm to deal with two unsynchronised input streams, rather than just the usual single stream. The use of the forward backward algorithm allows for uncertainty to be computed efficiently, with effort comparable to conventional matching by dynamic programming.

One application of the new algorithm is to the problem of view interpolation, replacing dynamic programming with forward-backward inference. Some practical evidence has been presented that this can reduce the severity of artefacts in the reconstructed cyclopean image. The new framework is particularly striking where ambiguity in matching is severe. In such locations, the uncertain representation can even pick out ambiguous and bifurcating paths in the disparity space. Other possible applications include i) detecting failures in stereo matching as heavy-tailed or multimodal distributions and ii) data fusion in robot navigation, combining uncertainty in observations with uncertainty in disparity.

Acknowledgements The authors would like to thank P. Dawid and C. Meek for useful comments and discussion.

References

[1] P.N. Belhumeur. A Bayesian approach to binocular stereopsis. *Int. J. Computer Vision*, 19(3):237–260, 1996.

[2] P.J. Burt and B. Julesz. A disparity gradient limit for binocular fusion. *Science*, 208:615–617, 1980.

[3] I. Cox, M. Ott, and J.P. Lewis. Videoconference system using a virtual camera image. *US Patent*, 5,359,362, 1993.

[4] I.J. Cox, S.L. Hingorani, and S.B. Rao. A maximum likelihood stereo algorithm. *Computer vision and image understanding*, 63(3):542–567, 1996.

[5] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological sequence analysis*. Cambridge University Press, 1998.

[6] B. Julesz. *Foundations of Cyclopean Perception*. University of Chicago Press, 1971.

[7] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *Proc. Europ. Conf. Computer Vision*, Copenhagen, Denmark, May 2002.

[8] Y. Ohta and T. Kanade. Stereo by intra- and inter-scan line search using dynamic programming. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7(2):139–154, 1985.

[9] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, Palo Alto, 1988.

[10] S.B. Pollard, J. Mayhew, and J. Frisby. PMF:a stereo correspondence algorithm using a disparity gradient. *Perception*, 14:449–470, 1985.

[11] L. Rabiner and B.H. Juang. A tutorial on hidden Markov models. *Proc. IEEE*, 77(2):257–286, 1989.

[12] S. Roy and I.J. Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *Proc. Int. Conf. Computer Vision*, pages 492–499, 1998.

[13] D. Scharstein. Stereo vision for view synthesis. In *Proc. Conf. Comp. Vision Pattern Rec.*, 1996.

[14] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Computer Vision*, 47(1–3):7–42, 2002.

[15] J. Sun, H. Y. Shum, and N. N. Zheng. Stereo matching using belief propagation. In *Proc. Europ. Conf. Computer Vision*, Copenhagen, Denmark, May 2002.

[16] T. Vetter. Synthesis of novel views from a single face image. *Int. J. Computer Vision*, 28(2):103–116, 1998.

[17] R. Yang and Z. Zhang. Eye gaze correction with stereovision for video tele-conferencing. In *Proc. Europ. Conf. Computer Vision*, volume 2, pages 479–494, Copenhagen, Denmark, May 2002.

A Derivation of forward algorithm

A derivation of the iterative formula (15) for forward probabilities follows here. Note that $\alpha_k(d_k)$ in (14) has the property that

$$\alpha_k(d_k) \propto p(d_k, \mathbf{z}_{h_k}), \tag{27}$$

the constant of proportionality $p(\mathbf{z}_{h_k})$ being independent of d_k. Then, the right hand side of (15) can be expressed, using (14) and (13), as

$$\begin{aligned} &\mathrm{RHS} \propto \\ &\sum_{d_{k-1}} p(d_{k-1}, \mathbf{z}_{h_{k-1}}) P_d(d_k|d_{k-1}) p\left([\mathbf{z}_{h_k} \backslash \mathbf{z}_{h_{k-1}}] \mid d_k, d_{k-1}\right). \end{aligned} \tag{28}$$

It follows from the assumed independence of the disparity process that

$$P_d(d_k|d_{k-1}) = P_d(d_k|d_{k-1}, \mathbf{z}_{h_{k-1}}) \tag{29}$$

so that

$$p(d_{k-1}, \mathbf{z}_{h_{k-1}}) P_d(d_k|d_{k-1}) = p(d_k, d_{k-1}, \mathbf{z}_{h_{k-1}}). \tag{30}$$

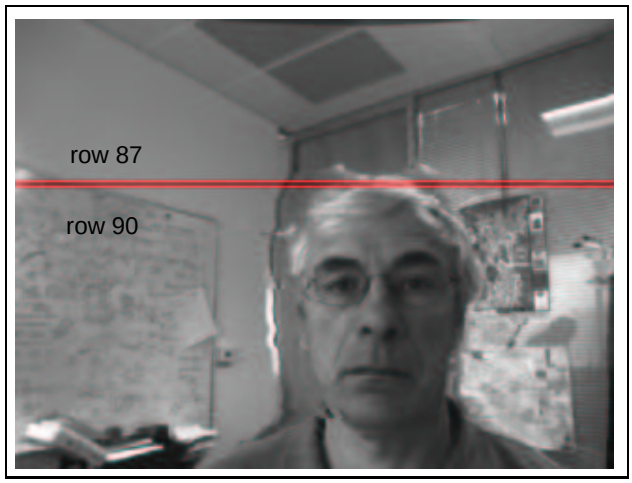
Using the mutual independence of observations,

$$p\left([\mathbf{z}_{h_k} \backslash \mathbf{z}_{h_{k-1}}] \mid d_k, d_{k-1}\right) = p\left([\mathbf{z}_{h_k} \backslash \mathbf{z}_{h_{k-1}}] \mid d_k, d_{k-1}, \mathbf{z}_{h_{k-1}}\right), \tag{31}$$

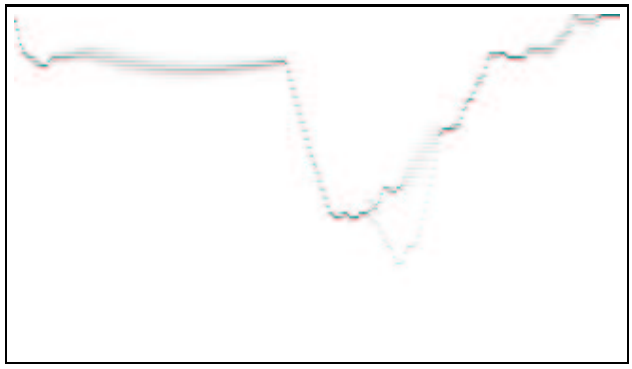
so now

$$\begin{aligned} &\mathrm{RHS} \\ &= \sum_{d_{k-1}} p(d_k, d_{k-1}, \mathbf{z}_{h_{k-1}})\, p\left([\mathbf{z}_{h_k} \backslash \mathbf{z}_{h_{k-1}}] \mid d_k, d_{k-1}, \mathbf{z}_{h_{k-1}}\right) \\ &= \sum_{d_{k-1}} p\left([\mathbf{z}_{h_k} \backslash \mathbf{z}_{h_{k-1}}], \mathbf{z}_{h_{k-1}}, d_k, d_{k-1}\right) \\ &= \sum_{d_{k-1}} p\left(\mathbf{z}_{h_k}, d_k, d_{k-1}\right) \\ &= p\left(\mathbf{z}_{h_k}, d_k\right), \end{aligned}$$

which is proportional to $\alpha_k(d_k)$ in (14), as required.



row 87



row 90

Figure 6: **Forward-Backward disparity distributions** for the data of figures 2 and 3. Disparity distributions are given for the two scan lines shown. Note that there is appreciable ambiguity of disparity, visible as vertical spreading in the diagrams, and this is what distinguished FB from DP.