The Panum Proxy algorithm for dense stereo matching over a volume of interest

A. Agarwal and A. Blake
Microsoft Research Ltd.
7 J J Thomson Ave, Cambridge, CB3 0FB, UK
http://research.microsoft.com/vision/cambridge

Abstract

Stereo matching algorithms conventionally match over a range of disparities sufficient to encompass all visible 3D scene points. Human vision however does not do this. It works over a narrow band of disparities — Panum's fusional band — whose typical range may be as little as 1/20 of the full range of disparities for visible points. Points inside the band are fused visually and the remainder of points are seen as "diplopic" — that is with double vision. The Panum band restriction is important also in machine vision, both with active (pan/tilt) cameras, and with high resolution cameras and digital pan/tilt.

A probabilistic approach is presented for dense stereo matching under the Panum band restriction. First it is shown that existing dense stereo algorithms are inadequate in this problem setting. Secondly it is shown that the main problem is segmentation, separating the (left) image into the areas that fall respectively inside and outside the band. Thirdly, an approximation is derived that makes up for missing out-of-band information with a "proxy" based on image autocorrelation. Lastly it is shown that the Panum Proxy algorithm achieves accuracy close to what can be obtained when the full disparity band is available.

1. Introduction

In attentional stereo vision, the viewer steers a *volume of interest* around the scene. This is a problem that has received a good deal of attention in the realms of oculomotor control [5, 8] and sparse stereo *e.g.* [14]. In the area of dense stereo however *e.g.* [12, 6, 2, 3, 10, 15] the issue of restricting attention to a volume, with a limited range of depth or equivalently disparity, has not been addressed. It is of considerable importance from the point of view of efficiency, particularly with high resolution or head-mounted cameras, in restricting computation to a volume of interest which may be only a small fraction of the visible volume. In principle also, it is most unsatisfying that conventional stereo algorithms need to explore an irrelevant background, simply in order to establish significiant properties of the foreground — a form of the celebrated "frame" problem of Artificial Intelligence.

1.1. The Panum band

The geometry of the situation is illustrated in figure 1. For a particular field of view of each camera, potential matches between left and right images form a diamond-shaped region in each epipolar plane. In human vision [13] the space of possible matches is restricted further to the "Panum band" (see figure). This is typically around 5 mrad wide, and cuts down the number of possible foveal matches by around an order of magnitude. High quality stereo cameras with narrow fields of view can also benefit from a Panum band restriction in a similar way.

The motivation for studying Panum band stereo is then threefold.

1. It is conceptually appealing to develop a stereo algorithm which focuses on a volume of interest, in the manner known to prevail in human vision. Why should a stereo algorithm expend needless attention to the entire background of a scene?

2. Computational cost for stereo matching grows linearly (or faster) with volume of interest. This is true both for both main components of stereo matching: cost computation and global optimization (whether by graph-cut (GC), dynamic programming (DP) or belief propagation (BP)). Restricting the size of the matching volume is therefore critical for efficiency. For stereo geometry similar to human vision, the saving in computational cost is at least an order of magnitude, due to the reduced range of *depth* (disparity). Usually there is a further factor of saving, due to the concomitant restriction in image *area* over which matching occurs. The best stereo algorithms (GC, DP or BP [15]) do not currently come close to real time. This is not going to be solved any time soon by Moore's law because camera resolution is increasing faster than processing power.

3. Computational cost, we have argued, necessitates the restriction of stereo matching to a Panum volume. However, existing dense stereo algorithms are not capable of satisfac-

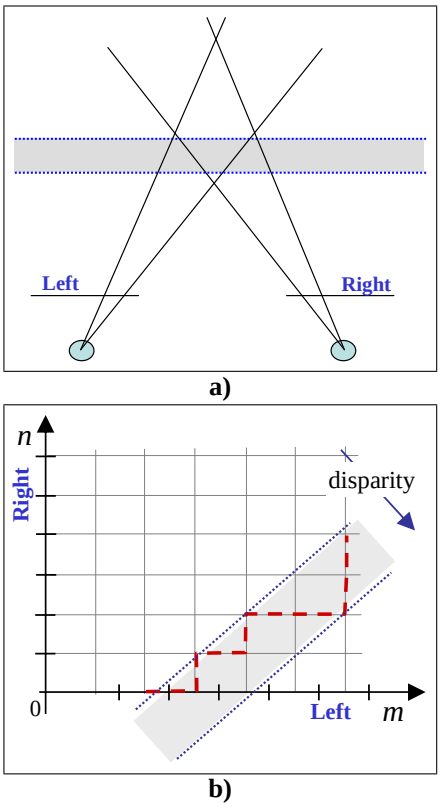

Figure 1. **The space of possible matches restricted to a Panum band.** a) View from above of rays in a single epipolar plane, forming a diamond-shaped match space; the Panum band forms a ribbon across the diamond and thus cuts down the set of possible matches. b) The match-space represents the situation in (a) in a standardised diagram, in which the diamond-shaped match-space becomes a square, whose sides are respectively a left and right epipolar line, and restricted to the Panum band as shown. A possible matching path is shown dashed.

tory operation over a Panum band, as this paper will show. A new algorithm is needed.

1.2. The Panum Proxy algorithm

The principle of the Panum Proxy stereo algorithm is therefore as follows.

1. Compute match scores or likelihoods for disparities within the Panum band.

2. Aggregate those scores to compute a total likelihood, at each point, that there is a within-band (foreground) match.

3. The same cannot be done for the background likelihood, as that would require match scores outside the band. However, it is shown that an autocorrelation-like measure can be used to estimate the background likelihood.

4. Use the true foreground likelihood and the estimated background likelihood, in a graph cut algorithm, to achieve a segmentation.

5. Once segmentation is complete, perform conventional stereo matching *e.g.* [3], but *restricted* to the image regions that have been labelled as in-band.

Note that the restriction to the Panum band in the segmentation step **4** is indeed essential, because the complexity of segmentation is dominated by the cost of computing stereo match scores, and this is linear in match volume.

The resulting stereo disparity map can be used, for example, to synthesise a new view, as in figure 2, in which case the view is fused within the Panum band, but diplopic outside it, just as in human stereo vision.

Figure 2. **Fusion and diplopia with the Panum Proxy algorithm.** Results of the Panum Proxy algorithm are illustrated here for a frame from one of the six Microsoft stereo datasets. The matched stereogram shows fusion within the Panum band but diplopia — double vision — elsewhere.

2. Probabilistic framework for stereo matching

First we outline the notation for probabilistic stereo matching. Pixels in the rectified left and right images are $\mathbf{L} = \{L_m\}$ and $\mathbf{R} = \{R_n\}$ respectively, and jointly we denote the two images $\mathbf{z} = (\mathbf{L}, \mathbf{R})$. Left and right pixels are associated by any particular matching path (fig. 1). Frequently in stereo matching the so-called "ordering constraint" is imposed, and this means that each move in figure 1b) is allowed only in the positive quadrant [1, 12]. Stereo "disparity" is $\mathbf{d} = \{d_m,\ m = 0, \ldots, N\}$ and disparity is simply related to image coordinates as $d_m = m - n$.

In algorithms that deal explicitly with occlusion [10, 7] an array $\mathbf{x}$ of state variables $\mathbf{x} = \{x_m\}$, takes values $x_m \in \{\mathrm{M}, \mathrm{O}\}$ according to whether the pixel is matched or occluded.

This sets up the notation for a path in epipolar match-space which is a sequence $((d_1, x_1), (d_2, x_2), \ldots)$ of disparities and states. A Gibbs energy $E(\mathbf{z}, \mathbf{d}, \mathbf{x}; \Theta, \Phi)$ can be defined for the posterior over all epipolar paths taken together and notated $(\mathbf{d}, \mathbf{x})$, given the image data $\mathbf{z}$. Parameters Φ and Θ relate respectively to prior and likelihood terms in the posterior. Then the Gibbs energy can be glob-

ally minimised to obtain a segmentation $\mathbf{x}$ and disparities $\mathbf{d}$.

2.1. Prior distribution over matching paths

A Bayesian model for the posterior distribution $p(\mathbf{x}, \mathbf{d} \mid \mathbf{z})$ is set up as a product of prior and likelihood:

$$p(\mathbf{x}, \mathbf{d} \mid \mathbf{z}) \propto p(\mathbf{x}, \mathbf{d}) p(\mathbf{z} \mid \mathbf{x}, \mathbf{d}). \quad (1)$$

The prior distribution $p(\mathbf{x}, \mathbf{d}) \propto \exp -\lambda E_0(\mathbf{x}, \mathbf{d})$ is frequently decomposed, in the interests of tractability, as a Markov model. An MRF (Markov Random Field) prior for $(\mathbf{x}, \mathbf{d})$ is specified as a product of clique potentials $V_{m,m'}$ over all pixel pairs $(m, m') \in \mathcal{N}$ deemed to be neighbouring in the left image. The potentials are chosen to favour matches over occlusions, to impose limits on disparity change along an epipolar line, and to favour figural continuity between matching paths in adjacent epipolar line-pairs.

2.2. Stereo matching likelihood

The stereo likelihood is:

$$p(\mathbf{z} \mid \mathbf{x}, \mathbf{d}) \propto \prod_m \exp -U_m^{\mathrm{M}}(x_m, d_m) \quad (2)$$

where the pixelwise negative log-likelihood *ratio*, for match vs. non-match, is

$$U_m^{\mathrm{M}}(x_m, d_m) = \begin{cases} M(L_m^{\mathrm{P}},\ R_n^{\mathrm{P}}) & \text{if } x_m = \mathrm{M} \\ M_0 & \text{if } x_m = \mathrm{O}, \end{cases} \quad (3)$$

where $M(\ldots)$ is a suitable measure of goodness of match between two patches, often based on normalised sum-squared difference (SSD) or correlation scores [15].

3. Restricting conventional stereo matching to a Panum band

We looked at two dense stereo matching algorithms which are considered competitive [15], one referred to as **BVZ** [4] that uses graph-cut optimization; the other **KZ** also using graph-cut but also with explicit allowance for occlusion [10]. The question is whether these algorithms can be applied to the Panum problem simply by reducing the disparity range available for matching. Following conventions for stereo testing, we took the four image pairs *Tsukuba*, *sawtooth*, *venus* and *map* on the Middlebury database[1], together with supplied ground truth, and calculated error measures. Over foreground, an error is counted wherever computed disparity is in error by more than 1 pixel. For background regions, the true disparity is of course out of range, so an incorrect disparity is considered to be as follows:

[1] http://cat.middlebury.edu/stereo

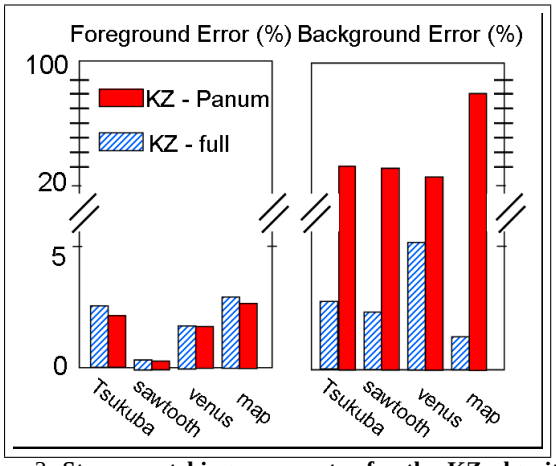

Figure 3. **Stereo matching error rates for the KZ algorithm constrained to the Panum band.** The error rate data show that background error is greatly magnified when the Panum band constraint is imposed, while foreground error barely changes.

BVZ: not at the endstop of the Panum band;

KZ: wherever the state is not occluded: $x_m \neq \mathrm{O}$, and the disparity is not at the endstop of the Panum band.

In each case we used the operating parameters recommended for the algorithms: for **BVZ**, disparity gradient penalty [3] $\lambda = 20$, and for **KZ**, $\lambda = 10$ with occlusion penalty [10] $K = 50$.

Results for the **KZ** algorithm are shown in figure 3. Results for the (simpler) **BVZ** algorithm are similar, but omitted here. In both cases, disparity error over foreground regions is not much affected by the Panum band restriction (in fact improved slightly because of the added constraint). Over background regions, error for both algorithms rises substantially. The conventional stereo algorithms simply fail over the background, generating many random disparities.

The conclusion from this experiment is that the conventional algorithms, when restricted to the Panum band, work perfectly well over foreground regions. All that is required to make the algorithms usable, is reliable identification of those pixels whose disparities fall within the band. In other words, successful Panum-band stereo could be achieved if only segmentation into foreground (within band) and background could be achieved reliably. Therefore the remainder of the paper considers the problem of foreground/background segmentation under the Panum band constraint.

3.1. Can graph-cut stereo be adapted for segmentation?

One possibility, for a more subtle adaptation of the existing **KZ** algorithm, is that its ability to label occlusions could

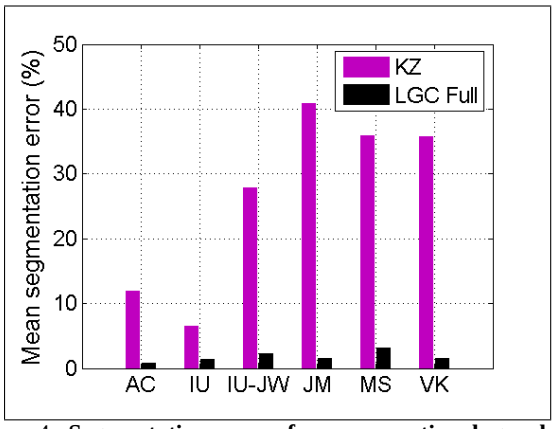

Figure 4. **Segmentation error from conventional graph-cut stereo.** The **KZ** algorithm is tuned here to use its occlusion labels to indicate background, but error rates are very high compared with what is attainable using LGC segmentation with the full range of disparities.

be extended to label background points. This is reasonable because, given the restricted Panum band, both occlusions and background points represent failures to obtain a stereo match. In order to give the **KZ** algorithm every chance of success, parameter value K was explored to minimise labelling error rate and this yielded parameters $\lambda = 10$, $K = 10$, quite different from the optimal operating point for regular use of **KZ** for stereo matching. Results are given in figure 4, showing segmentation error for each of the six test videos in the Microsoft stereo-segmentation database[2]. Labelling error-rates (equal error-rate) for the 6 datasets vary between 7% and 41%, and are in all cases many times worse than are obtainable from full, unconstrained stereo segmentation, in the form of LGC (Layered Graph Cut) [9]. Since the aim, with Panum-band stereo, is to approach the quality of full, unconstrained stereo, the performance of **KZ** in this mode is far from acceptable.

3.2. Segmentation of the in-band image region

Given the results and discussion so far, the aim of the remainder of the paper is to develop a segmentation algorithm, to label all "foreground" points with an accuracy approaching full LGC, but without any computation out of the Panum band. Segmentation could be done in one of two ways. Either it could proceed simultaneously with computation of disparity; or in a separate pass, preceding the computation of disparity. Simultaneous segmentation and disparity determination perhaps has the attraction of greater elegance. On the other hand, separate segmentation could be achieved by marginalising the stereo likelihood over disparities $\mathbf{d}$, and then performing energy minimisation with respect to labels $\mathbf{x}$ only. A separate labelling pass should surely be more efficient, since full consideration of disparity need then only occur within the foreground region.

[2] research.microsoft.com/vision/cambridge/i2i

4. Stereo segmentation

First we summarise the full LGC (Layered Graph Cut) algorithm [9] for segmentation by marginalisation of stereo likelihoods. Then in the next section the full LGC energy function is approximated to stay within the Panum band restriction.

For LGC, the matched state M is further subdivided into foreground match F and background match B. LGC determines segmentation $\mathbf{x}$ as the minimum of an energy function $E(\mathbf{z}, \mathbf{x}; \Theta)$, in which stereo disparity $\mathbf{d}$ does not appear explicitly. Instead, the stereo match likelihood (2) in section 2.2 is marginalised over disparity, aggregating support from each putative match, to give a likelihood $p(\mathbf{L} \mid \mathbf{x}, \mathbf{R})$ for each of the three label-types occurring in $\mathbf{x}$: foreground, background and occlusion $(\mathrm{F}, \mathrm{B}, \mathrm{O})$. Segmentation is therefore a ternary problem, and it can be solved (approximately) by iterative application of a binary graph-cut algorithm, augmented for a multi-label problem by so-called α-expansion [4]. The energy function for LGC is composed of two terms:

$$E(\mathbf{z}, \mathbf{x}; \Theta, \Phi) = V(\mathbf{z}, \mathbf{x}; \Theta) + U^{\mathrm{S}}(\mathbf{z}, \mathbf{x}, \Phi) \qquad (4)$$

representing energies for spatial coherence/contrast and stereo likelihood.

4.1. Encouraging coherence

The coherence energy $V(\mathbf{z}, \mathbf{x}; \Theta)$ is a sum, over cliques, of pairwise energies with potential coefficients $F_{m,m'}$ now defined as follows. Cliques consist of horizontal, vertical and diagonal neighbours on the square grid of pixels. For vertical and diagonal cliques it acts as a switch active across a transition in or out of the foreground state: $F_{m,m'}[x, x'] = \gamma$ if exactly one of the variables x, x' equals F, and $F_{m,m'}[x, x'] = 0$ otherwise. Horizontal cliques, along epipolar lines, inherit the same cost structure, except that certain transitions are disallowed on geometric grounds. These constraints are imposed via infinite cost penalties:

$$F_{m,m'}[x = \mathrm{F}, x' = \mathrm{O}] = \infty; \quad F_{m,m'}[x = \mathrm{O}, x' = \mathrm{B}] = \infty.$$

where [9] $\gamma = \log(2\sqrt{W_{\mathrm{M}} W_{\mathrm{O}}})$ and parameters W_{M} and W_{O} are the mean widths (in pixels) of matched and occluded regions respectively.

4.2. Encouraging boundaries where contrast is high

A tendency for segmentation boundaries in images to align with contours of high contrast is achieved by defining prior penalties $F_{k,k'}$ which are suppressed where image

contrast is high [3, 4, 11], multiplying them by a discount factor $C^*_{m,m'}(L_m, L_{m'})$ which suppresses the penalty by a factor $\epsilon/(1+\epsilon)$ wherever the contrast across $(L_m, L_{m'})$ is high — see [9] for details. Previously, maximal discounting has been obtained [3] by setting $\epsilon = 0$. Here, as in stereo segmentation [10], $\epsilon = 1$ tends to give the best results, though sensitivity to the precise value of ϵ is relatively mild.

4.3. Foreground likelihood

The remaining term in (4) is $U^{\mathrm{S}}(\mathbf{z}, \mathbf{x})$ which captures the influence of stereo matching likelihood on the probability of a particular segmentation. It is defined to be

$$U^{\mathrm{S}}(\mathbf{z}, \mathbf{x}) = \sum_m U^{\mathrm{S}}_m(x_m) \tag{5}$$

$$\text{where } U^{\mathrm{S}}_m(x_m) = -\log p(L_m \mid x_m = \mathrm{F}, \mathbf{R}). \tag{6}$$

Now, marginalising out disparity, foreground likelihood is

$$p(L_m|x_m = \mathrm{F}, \mathbf{R}) = \sum_d p(L_m|d_m = d, \mathbf{R}) p(d_m = d|x_m = \mathrm{F}) \tag{7}$$

where, from (2),

$$p(L_m \mid d_m = d, \mathbf{R}) \propto f(L, d, \mathbf{R}) = \exp -U^{\mathrm{M}}_m(x_m, d_m), \tag{8}$$

using the log-likelihood ratio defined in (3). As a shorthand, we write:

$$p(L \mid \mathrm{F}) = \sum_d p(L \mid d, \mathbf{R}) p(d \mid \mathrm{F}) \tag{9}$$

and, as before, in terms of likelihood ratios, this becomes:

$$\mathcal{L}(L \mid \mathrm{F}) \equiv \frac{p(L \mid \mathrm{F})}{p(L \mid \mathrm{O})} = \sum_d f(L, d, \mathbf{R}) p(d \mid \mathrm{F}) \tag{10}$$

where $f(L, d, \mathbf{R})$ is the match/non-match likelihood ratio as above.

4.4. Background likelihood

Since the distribution $p(d_m = d \mid x_m = \mathrm{F})$ is defined to be zero outside the Panum fusional area, it is perfectly possible, under the Panum assumptions, to compute $\mathcal{L}(L \mid \mathrm{F})$ in (10). However, the same cannot be said for

$$\mathcal{L}(L \mid \mathrm{B}) \equiv \frac{p(L \mid \mathrm{B})}{p(L \mid \mathrm{O})} = \sum_d \mathcal{L}(L \mid d) p(d \mid \mathrm{B}) \tag{11}$$

since the corresponding summation is entirely *outside* the Panum band $\mathcal{D}_{\mathrm{F}}$ of disparities, in that $p(d \mid \mathrm{B})$ is non-zero only outside the Panum band. Each pixel L_m would therefore have to be compared with pixels in the right image $\mathbf{R}$ that are unreachable because they are outside the band.

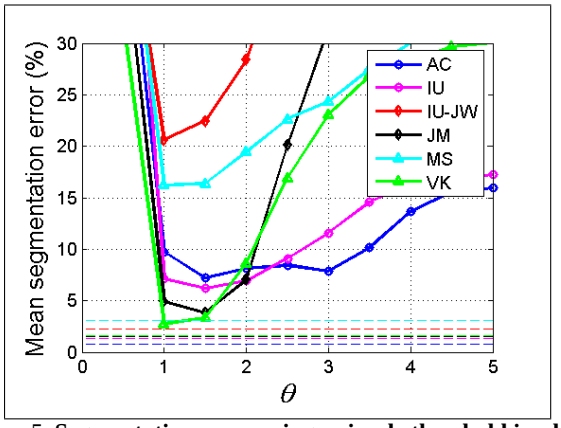

Figure 5. **Segmentation error using a simple threshold in place of background likelihood.** Error curves are shown as a function of threshold θ for six subjects from the Microsoft database. (Error-rates are total foreground and background error, averaged over each sequence.) Horizontal dashed lines show corresponding error rates for full (non-Panum) LGC segmentation, as a benchmark. The substantial shortfall suggests that it should be possible to improve considerably on simple thresholding.

4.5. A simple threshold as proxy for the background likelihood?

Before going to some trouble to approximate the background likelihood, it is worth looking at the simplest possible approach, and treating the problem as novelty detection. In that view, we have a model $\mathcal{L}(L \mid \mathrm{F})$ for the positive class, and no model of the background class. Then the likelihood ratio classifier $\mathcal{L}(L \mid \mathrm{F}) > \mathcal{L}(L \mid \mathrm{B})$ is simplified to a threshold rule, replacing the background likelihood by a constant $\mathcal{L}(L \mid \mathrm{B}) = \theta$. Segmentation under this model, for variable threshold θ, is exhibited in figure 5. It appears that a constant threshold $\theta = 1$ yields close to the best error for each of the 6 datasets, so there would be no need for an adaptive algorithm. However, the best error rate achieved is between 2 and 8 times higher than the error achieved (dashed lines) by full LGC. Again, therefore, there is strong motivation to look for a model and an algorithm that performs better under the Panum-band restriction.

5. The Panum Proxy algorithm

In the previous section, it was shown that the Panum-band constraint means that information required for computing background likelihood $\mathcal{L}(L|\mathrm{B})$ is missing, and that replacing $\mathcal{L}(L|\mathrm{B})$ with a simple threshold constant gives poor results. Therefore in this section an approximation for $\mathcal{L}(L|\mathrm{B})$ is developed.

5.1. Deriving the approximate likelihood

We assume that $p(d \mid \mathrm{F})$ is uniform over the Panum band so that $p(d \mid \mathrm{F}) = 1/|\mathcal{D}_{\mathrm{F}}|$ and similarly, for the background,

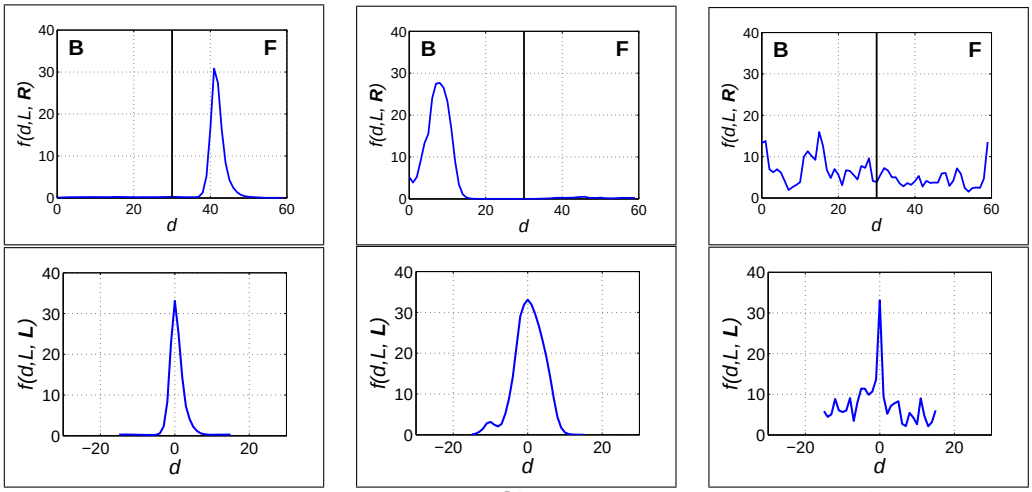

Figure 6. **Using the left image as a proxy for the right, to approximate likelihoods.** Upper: likelihood function $f(L, d, \mathbf{R})$ for three sample image points; Panum band is $30 \leq d \leq 60$. Lower: "autocorrelation-like" proxy $f(L, d, \mathbf{L})$ for the same three sample points. In the first two examples a,b), the proxy satisfactorily mimics the shape of the true likelihood. In c) the sample happens to fall on a textureless area, with consequent stereo ambiguity, and the proxy fails, in that $f(L, d, \mathbf{L})$ has a dominant peak whereas there is none in $f(L, d, \mathbf{R})$. A test will be developed for this case.

$p(d \mid \mathrm{B}) = 1/|\mathcal{D}_\mathrm{B}|$. Then, defining $\mathcal{D} = \mathcal{D}_\mathrm{B} \cup \mathcal{D}_\mathrm{F}$, we can write

$$S(L) \equiv \sum_{d \in \mathcal{D}} f(L, d, \mathbf{R}) \quad (12)$$

$$= |\mathcal{D}_\mathrm{F}| \mathcal{L}(L \mid \mathrm{F}) + |\mathcal{D}_\mathrm{B}| \mathcal{L}(L \mid \mathrm{B}), \quad (13)$$

from (10) and (11). If $S(L)$ were known, then it would be possible, having computed $\mathcal{L}(L \mid \mathrm{F})$ as in (10), to compute $\mathcal{L}(L \mid \mathrm{B})$ from the constraint (12). Of course S in (12) cannot be computed exactly, because the summation extends outside the Panum band. However, and this is the key idea of the Panum Proxy, we can approximate it by using the left image $\mathbf{L}$ as a *proxy* for the right image $\mathbf{R}$ in the match likelihood ratio:

$$\tilde{S}(L) = \sum_{d=-d_\mathrm{S}}^{d_\mathrm{S}} f(L, d, \mathbf{L}) \quad (14)$$

— see figure 6 for diagrams illustrating how this works. The approximation rests on the assumption that each match is a good one, since it is matching the left image with itself. Note that the value of $f(L, d, \mathbf{L})$ at $d = 0$ is an upper bound on the value of $f(L, d, \mathbf{R})$ at the true match value of d, since the match of the left image directly onto itself is of course perfect; d_S has to be chosen just big enough to capture the peak of the match-likelihood, but it is reasonably assumed that $d_\mathrm{S} \ll |\mathcal{D}_\mathrm{F}|$ so that the additional work in computing (14) is smaller than the amount of matching work done already in the Panum band. Note that a factor of 2 can be saved in computing (14) by exploiting the symmetry of autocorrelation, that is that $f(L_m, d, \mathbf{L}) = f(L_{m+d}, -d, \mathbf{L})$. Finally, having estimated $S(L)$, we can estimate the background likelihood-ratio from the approximate constraint

$$\tilde{S}(L) = |\mathcal{D}_\mathrm{F}| \mathcal{L}(L \mid \mathrm{F}) + |\mathcal{D}_\mathrm{B}| \mathcal{L}(L \mid \mathrm{B}). \quad (15)$$

$$\text{giving } \mathcal{L}(L \mid \mathrm{B}) = \left(\tilde{S}(L) - |\mathcal{D}_\mathrm{F}| \mathcal{L}(L \mid \mathrm{F})\right) / |\mathcal{D}_\mathrm{B}|. \quad (16)$$

5.2. Complementary likelihood

Now given the weakness of evidence, resulting from the Panum band restriction, for distinguishing background match from occlusion, we do not attempt to distinguish the hypotheses B and O. Therefore we lump them together as the complementary hypothesis $\overline{\mathrm{F}} = \mathrm{B} \cup \mathrm{O}$, so that

$$p(L \mid \overline{\mathrm{F}}) p(\overline{\mathrm{F}}) = p(L \mid \mathrm{B}) p(\mathrm{B}) + p(L \mid \mathrm{O}) p(\mathrm{O}), \quad (17)$$

and again dividing by $p(L \mid \mathrm{O})$:

$$\mathcal{L}(L \mid \overline{\mathrm{F}}) p(\overline{\mathrm{F}}) = \mathcal{L}(L \mid \mathrm{B}) p(\mathrm{B}) + p(\mathrm{O}). \quad (18)$$

and this is expressed as

$$\mathcal{L}(L \mid \overline{\mathrm{F}}) = (1 - \nu) \mathcal{L}(L \mid \mathrm{B}) p(\mathrm{B}) + \nu, \quad (19)$$

where $\nu = p(\mathrm{O})/(1 - p(\mathrm{F}))$, for which a typical value would be $\nu = 0.1$, reflecting the empirical fact that normally a small proportion of background points are occluded.

5.3. Kurtosis test

Earlier in figure 6, we saw that although $f(L, d, \mathbf{L})$ is often a good predictor of the shape of $f(L, d, \mathbf{R})$, as in figure 6a,b), it can fail where there is no clear peak in $f(L, d, \mathbf{L})$, as in figure 6c). The kurtosis $k = k(L, \mathbf{L})$, of $f(L, d, \mathbf{L})$ as a function of d, is computed as a diagnostic. Figure 7 shows that high kurtosis is associated with low error in the proxy. Therefore the likelihood estimate is predicted to be reliable if $k > k_0$.

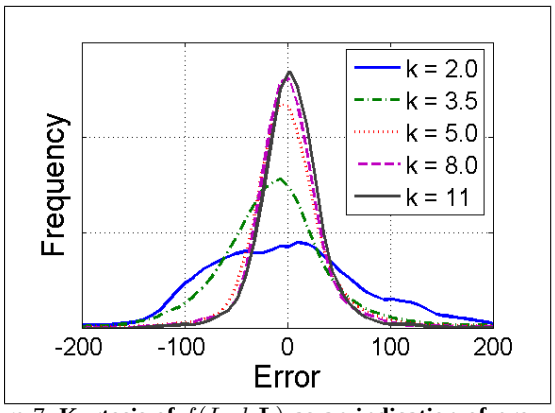

Figure 7. **Kurtosis of $f(L, d, \mathbf{L})$ as an indication of proxy accuracy.** High kurtosis k is associated with reduced magnitude of error $\tilde{S} - S$, suggesting a validity check based on kurtosis.

In fact low kurtosis occurs in practice over relatively textureless image areas, just the situation that gives rise to ambiguous disparity, as in figure 6c). A threshold value of $k_0 = 2.5$ has proved effective, catching 86% of points on the tails of the error distribution (defined to be those outside 1 standard deviation). Then the definition of $\tilde{S}$ from (14) is replaced by

$$\tilde{S}(L) = r(k) \sum_{d=-d_S}^{d_S} f(L, d, \mathbf{L}) \; + \; (1 - r(k)) \; |\mathcal{D}| \; \mathcal{L}(L|F), \tag{20}$$

where $r(k)$ is soft threshold function, taking the value $r(k) = 1$ when $k \gg 0$ and $r(k) = 0$ when $k \ll 0$. In this way, the estimated complementary likelihood $\mathcal{L}(L|\overline{\mathrm{F}})$ (19) is unchanged in the reliable case $r(k) = 1$. In the unreliable case $r(k) = 0$, $\tilde{S}(L) = |\mathcal{D}| \; \mathcal{L}(L|F)$, and $\mathcal{L}(L \mid \mathrm{B}) = \mathcal{L}(L \mid \mathrm{F})$ from (16), and then (19) defaults towards the no-information condition $\mathcal{L}(L|\overline{\mathrm{F}}) = \mathcal{L}(L|\mathrm{F})$ as $r(k) \to 0$.

5.4. Positivity check

The other condition that must be dealt with is the possible negativity of the estimated $\mathcal{L}(L|\mathrm{B})$ (16). In the case of negativity, we simply replace (16) with

$$\mathcal{L}(L|\mathrm{B}) = \mathcal{L}(L|\mathrm{F})/\eta \tag{21}$$

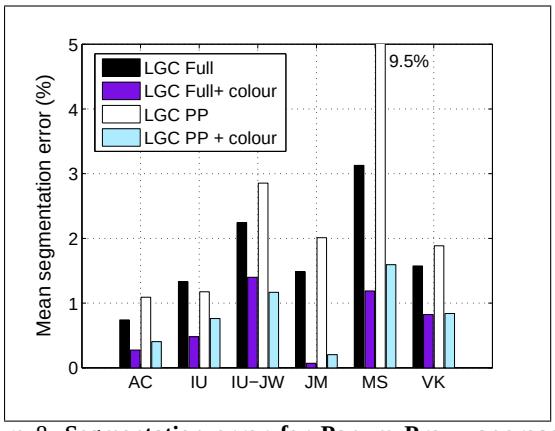

Figure 8. **Segmentation error for Panum Proxy approaches that of full stereo.** The Panum Proxy (**LGC PP**) algorithm achieves error rates that approach very nearly the level achieved by LGC segmentation with the full range of disparities (**LGC full**), at considerably reduced computational cost. For one test set (**MS**) the error rate for (**LGC PP**) is relatively high, though this is restored by adding in colour information.

and use this to evaluate the complementary hypothesis (19). The value of η is set using the statistics of $\mathcal{L}(L|\mathrm{F})/\mathcal{L}(L|\mathrm{B})$ in the negativity condition, collected from a variety of images, and this gives a working value of $\eta = 3$.

6. Results

First we show mean error rates, averaged over the entire stereo video sequence for each of the six subjects from the Microsoft database. These are extensive tests, representing measurements taken from several hundred stereo pairs. Figure 8 shows that error for the Panum Proxy algorithm approaches quite closely that for full Layered Graph Cut (LGC), with unrestricted stereo disparity. Note that error rates are mostly an order of magnitude better than for the conventional graph cut stereo algorithm **KZ** (figure 4). Compared with the naive thresholding scheme (figure 5), in which background likelihood is replaced by a constant, error rates for the Panum Proxy algorithm are lower by factors ranging from 1.5 to 7 across the six subjects. Error rates fall even further when colour information is used in the segmentation following the paradigm, used in LGC [9]. These results can be examined in more detail along their timelines, and we show this here just for the **VK** dataset, in figures 9 and 10.

Discussion We have shown that stereo within a Panumband can be solved effectively using a conventional stereo algorithm, together with a pre-segmentation step that selects those pixels that are within the band. This allows stereo to operate within a volume of interest, fusing over that volume, and with diplopic vision elsewhere, as in figure 2. Results

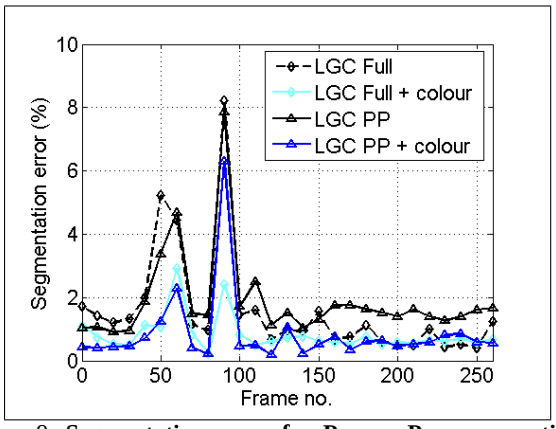

Figure 9. **Segmentation error for Panum Proxy, over time, for one subject (VK).** The Panum Proxy (**LGC PP**) algorithm achieves error rates close to those achieved by LGC segmentation with the full range of disparities (**LGC full**), especially when colour information is fused in with stereo. See also figure 10.

Best case: frame 120 Worst case: frame 90

Dataset **VK**

Figure 10. **Some examples of segmentation.** Segmentations are shown for the frames with lowest and highest error, for the dataset of fig 9. Results are for **LGC PP** plus colour.

of the Panum Proxy algorithm are close in quality to what is obtainable under unconstrained conditions, using the full range of available disparity. It remains for future work to test the algorithm under more stringent circumstances, with greater ranges of disparity in the scenes.

Acknowledgements We acknowledge helpful discussions with A. Criminisi, A. Fitzgibbon and V. Kolmogorov.

References

[1] H.H. Baker and T.O. Binford. Depth from edge and intensity based stereo. In *Proc. Int. Joint Conf. Artificial Intelligence*, pages 631–636, 1981. 2

[2] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. In *Proc. European Conf. Computer Vision*, number 800 in Lecture notes in computer science, pages 45–58. Springer-Verlag, 1996. 1

[3] Y.Y. Boykov and M-P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In *Proc. Int. Conf. on Computer Vision*, pages 105–112, 2001. 1, 2, 3, 5

[4] Y.Y. Boykov, O. Veksler, and R.D. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(11), 2001. 3, 4, 5

[5] C.M. Brown, D. Coombs, and J. Soong. Real-time smooth pursuit tracking. In A. Blake and A.L. Yuille, editors, *Active Vision*, pages 123–136. MIT, 1992. 1

[6] 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. 1

[7] A. Criminisi, J. Shotton, A. Blake, and P.H.S. Torr. Gaze manipulation for one to one teleconferencing. In *Proc. Int. Conf. on Computer Vision*, pages 191–198, 2003. 2

[8] T. Uhlin K. Pahlavan and J-O. Eklundh. Dynamic fixation. In *Proc. Int. Conf. on Computer Vision*, pages 412–419, 1993. 1

[9] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross, and C. Rother. Bi-layer segmentation of binocular stereo video. In *Proc. Conf. Computer Vision and Pattern Recognition*, 2005. 4, 5, 7

[10] V. Kolmogorov and R. Zabih. Computing visual correspondences with occlusions using graph cuts. In *Proc. Int. Conf. on Computer Vision*, 2001. 1, 2, 3, 5

[11] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *Proc. European Conf. Computer Vision*, pages 82–96, 2002. 5

[12] 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. 1, 2

[13] T. Poggio and W. Reichardt. Visual control of orientation behaviour in the fly. *Quart. Rev. Biophys.*, 9(3):377–438, 1984. 1

[14] I.D. Reid and D.W. Murray. Tracking foveated corner clusters using affine structure. In *Proc. Int. Conf. on Computer Vision*, pages 76–83, 1993. 1

[15] 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. 1, 3