In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'96)*, San Francisco, June 1996, pp. 343–350

Stereo Matching with Non-Linear Diffusion

Daniel Scharstein*
Department of Computer Science
5151 Upson Hall
Cornell University
Ithaca, NY 14853-7501
schar@cs.cornell.edu

Richard Szeliski
Vision Technology Group
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052-6399
szeliski@microsoft.com

Abstract

One of the central problems in stereo matching (and other image registration tasks) is the selection of optimal window sizes for comparing image regions. This paper addresses this problem with some novel algorithms based on iteratively diffusing support at different disparity hypotheses, and locally controlling the amount of diffusion based on the current quality of the disparity estimate. It also develops a novel Bayesian estimation technique which significantly outperforms techniques based on area-based matching (SSD) and regular diffusion. We provide experimental results on both synthetic and real stereo image pairs.

1 Introduction and related work

Most *area-based* approaches to the stereo correspondence problem perform the following three tasks:

1. For each disparity under consideration, compute a per-pixel matching cost;
2. Aggregate support spatially (e.g. by summing over a window, or by diffusion);
3. Across all disparities, find the best match based on the aggregated support.

The focus of this paper is the second step, *aggregation of support*. A central problem is to find the optimal size of the support region. If the region is too small, a wrong match might be found due to ambiguities and noise. If the region is too big, it can no longer be matched as a whole due to foreshortening and occlusion, with the result of lost detail and blurring of object boundaries.

Kanade and Okutomi [10] have proposed *adaptive windows*, square windows that extend by different amounts in each of four directions. The optimal window size is found by a greedy algorithm (gradient descent) based on an estimate of disparity uncertainty in the current window. In this paper we propose a different approach: aggregating support with a non-uniform diffusion process.

A support region can either be two-dimensional at a fixed disparity (favoring fronto-parallel surfaces), or three-dimensional in x-y-d space (supporting slanted surfaces). Two-dimensional evidence aggregation has been done using square windows (traditional), Gaussian convolution [14], and windows with adaptive sizes [10]. Three-dimensional support functions that have been proposed include limited disparity difference [8], limited disparity gradient [12], and Prazdny's coherence principle [13], which can be implemented using two diffusion processes [17].

Some matching costs, such as correlation and non-parametric measures are defined over a certain area of support, and thus combine the cost and aggregation steps into one. Measures that can be accumulated in a separate step, however, have the following advantages:

- Efficiency: the measure can be aggregated with a single convolution (or box-filter) operation [14];
- Parallelizability: the aggregation step can be implemented on highly parallel architectures using local iterative diffusion [17];
- Adaptability: the measure can be aggregated over locally different support regions using either adjustable size windows [10] or a non-uniform diffusion process (this paper).

Other stereo techniques include hybrid and iterative techniques, such as stochastic search [1, 17], as well as hierarchical methods. More than two images are used in multiframe stereo to increase stability of the algorithm [4, 11]. A long version of this paper containing more references is available as Cornell CS TR 96-1575 [15]. For a general survey of stereo vision methods see [5].

2 Disparity space and SSD

Support for a match is defined over a three-dimensional *disparity space* $E(x, y, d)$. Formally, we

*Supported by NSF grant IRI-9057928.

Figure 1: Synthetic stereo pairs *ramp* and *rds* and the underlying disparity pattern.

define the initial disparity space E_0 as

$$E_0(x, y, d) = \rho(I_L(x + d, y) - I_R(x, y)),$$

where I_L, I_R, are the intensity functions of the left and right image respectively, and ρ measures the similarity between the two intensities, e.g., $\rho(l - r) = (l - r)^2$. After aggregating support into a final space $E(x, y, d)$, we can compute a disparity function

$$d(x, y) = \arg\min_{d \in D} E(x, y, d)$$

that represents the matches as offsets to the points in the right image.

The standard sum-of-squared-differences algorithm (SSD) uses square windows for aggregation. As mentioned before, choosing the right window size involves a trade-off between a noisy disparity map and blurring of depth boundaries. We will illustrate this using two synthetic image pairs. Both pairs have the same disparity pattern: a central square floating in front of a background with constant disparity.

Figure 1 shows the two image pairs and the disparity pattern. The *ramp* pair is similar to the image pair in Fig. 5 in [10] and is based on a linear intensity ramp in the direction of the baseline. Gaussian noise has been added to each image independently. The *rds* pair is based on a binary random dot pattern using two gray levels with equal probability.

The two image pairs are quite different. The *ramp* pair has no local texture variation and constant gradients everywhere, except for the boundaries of the central square. The two images can only be matched by comparing absolute intensities, and any algorithm based on band-pass filtered intensities or gradients will fail (as will the human visual system). The *rds* pair, on the other hand, has strong local texture variation, but is highly ambiguous since pixels not in correspondence still have a 50% chance of matching.

Figure 2 shows the performance of the simple SSD algorithm on these two image pairs using two different window sizes, $w = 3$ and $w = 7$. As can be seen, the bigger window size yields a disparity map with less noise, but results in an overall blurring of the features. The effect on the two image pairs is quite different: in the *ramp* pair, the disparities are smoothed

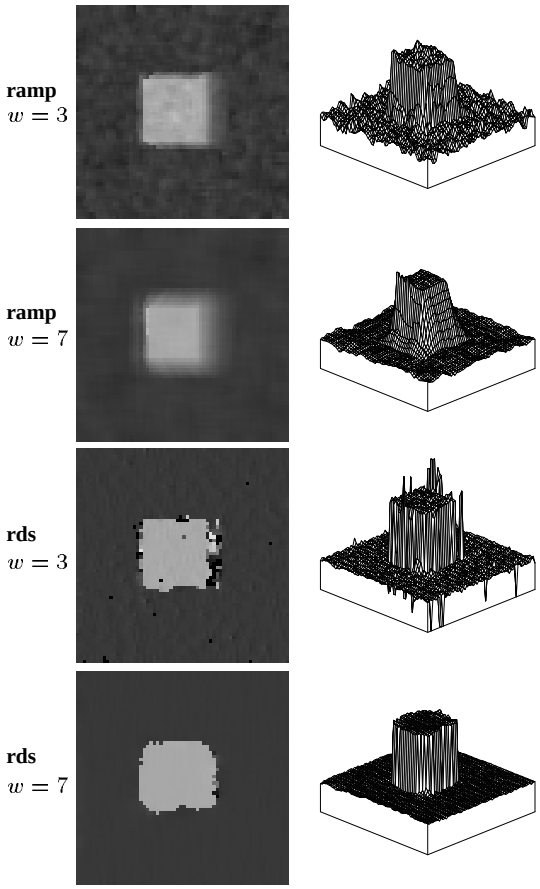

Figure 2: Performance of the SSD algorithm using square windows with sizes $w = 3$ and $w = 7$.

across the boundaries, while in the *rds* pair only the *outlines* of the square are blurred, i.e., the corners are rounded, while the two disparity levels of foreground and background are clearly recovered. The latter effect, smoothing of object boundaries, is more common in real images pairs than the smoothing of disparities. The smoothing of disparities we observed in the *ramp* pair is a direct result of the ramp intensity pattern and the small local variations in intensity.

Since the blurring of outlines is caused by support regions that span object boundaries, a possible solu-

tion to the problem is to use non-uniform and adaptive support regions. Kanade and Okutomi [10] have proposed *adaptive windows*, square windows that extend by different amounts in each of four directions. The optimal window size is found by a greedy algorithm (gradient descent) based on an estimate of disparity uncertainty in the current window. In this paper we propose a different approach: aggregating support with a non-uniform diffusion process.

3 Aggregating support by diffusion

Instead of using a fixed window, support can also be aggregated with a weighted support function such as a Gaussian. A convolution with a Gaussian can be implemented using local iterative diffusion [17] defined by the equation

$$\frac{\partial E}{\partial t} = \nabla^2 E.$$

In a discrete system, this yields the update rule

$$E(i,j,d) \leftarrow (1-4\lambda)E(i,j,d) + \lambda \sum_{(k,l)\in\mathcal{N}_4} E(i+k, j+l, d),$$

where $\mathcal{N}_4 = \{(-1,0),(1,0),(0,-1),(0,1)\}$ is the local neighborhood containing the four direct neighbors, and λ controls the speed of the diffusion. A value of $\lambda < 0.25$ is needed to ensure convergence; we use $\lambda = 0.15$ for the experiments reported in this paper.

Aggregation using a finite number of simple diffusion steps yields fairly similar results to using square windows. Advantages include the rotational symmetry of the support kernel and the fact that points further away have gradually less influence. The problem of boundary blurring still exists, however.

3.1 Membrane model

A problem with simple diffusion is that the size of the support region increases with the number of iterations. While the diffusion would eventually converge to a uniform support covering the whole image, we are interested in an intermediate time step in which the diffusion has only progressed to a certain amount. We can change this behavior by adding a term to the diffusion equation that measures the amount each current value has diverged from its original value, yielding the *membrane equation* [17]

$$\frac{\partial E}{\partial t} = \nabla^2 E + \beta(E_0 - E).$$

The discrete update rule is

$$E(i,j,d) \leftarrow [1-\lambda(\beta+4)]E(i,j,d) + \lambda[\beta E_0(i,j,d) + \sum_{(k,l)\in\mathcal{N}_4} E(i+k, j+l, d)].$$

The β-term ensures that the diffusion converges to a stable solution not too far from the original values.

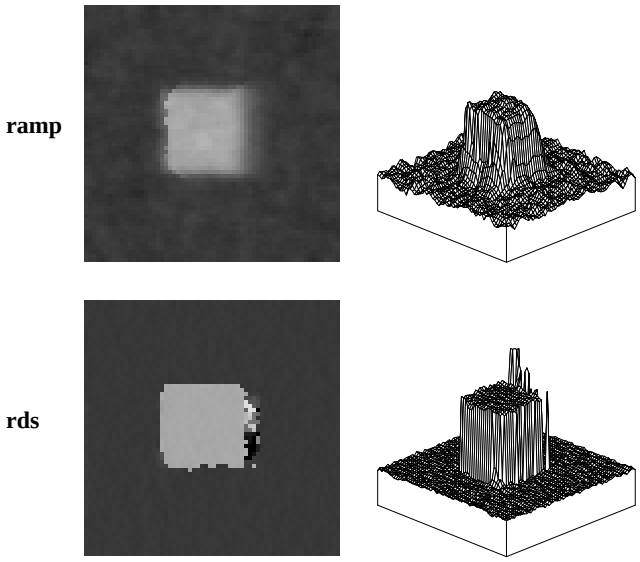

Figure 3: Performance of the membrane model on the *ramp* and *rds* image pairs.

Unless noted otherwise, we use $\beta = 0.5$ in the results reported in this paper. A closed-from solution for the support function can easily be derived using Fourier analysis [15].

Figure 3 shows the results for accumulating support using the membrane model for the *ramp* and *rds* pairs. The number of diffusion iterations is $n = 10$. Using the membrane model alleviates the contour blurring problem to some extent, since the β-term "ties" the center of each support region to its original value. For very noisy images, however, β needs to be chosen quite small to enable enough smoothing for stable matching, making the process more similar to regular diffusion.

3.2 Diffusion with local stopping criteria

A different strategy for preventing both boundary blurring and diffusion to uniformity is to locally stop the diffusion process depending on the distribution of values in each disparity column. To do this, we associate a measure of *certainty* $C(i,j)$ with each location. Intuitively, this measure should reflect how "clear" a minimum there is among the values $E(i,j,d)$ for all d. Given such a measure C, we can aggregate support using *non-uniform diffusion*:

> For each (i,j), compute certainties C and C' before and after a single iteration of diffusion. If $C > C'$, do not diffuse, i.e., restore the old values $E(i,j,d)$ for all d.

The idea is that diffusion takes place only at locations of ambiguous matches. Also, certainties never decrease, thus guarantying convergence.

We have experimented with several different certainty measures. In this paper we will discuss two measures, the *winner margin*, and the *entropy*. The winner margin C_m is the normalized difference between the minimum $E_{\min}$ and the second minimum $E_{\min 2}$ in a disparity column:

$$C_m(i,j) = \frac{E_{\min 2} - E_{\min}}{\sum_d E(i,j,d)}.$$

The second measure C_e is the negative entropy of the probability distribution in the disparity column. We convert to probabilities by taking the inverse exponent and normalizing:

$$C_e(i,j) = -\sum_d p(d) \log p(d),$$

with $p(d) = e^{-E(i,j,d)} / \sum_{d'} e^{-E(i,j,d')}$. We will develop the idea of converting to probabilities further in the next section.

All three methods (local stopping with C_e and C_m and the membrane model) yield fairly similar disparity maps [15]. In Section 5 we numerically analyze their respective performance based on errors in the computed disparities.

4 A Bayesian model of stereo matching

In this section, we develop a Bayesian model for stereo matching that includes both a measurement model corresponding to the matching criterion and a prior Markov Random Field model corresponding to the aggregation function. Our model uses robust (non-Gaussian) statistics to handle gross errors and discontinuities in the surface. We also develop a novel approximation algorithm that results in a non-linear diffusion process, and show how this produces better results than standard diffusion.

As before, stereo reconstruction is specified as the estimation of a discrete disparity field $d_{i,j} = d(x_i, y_j)$ given two input images $I_L(x,y)$ and $I_R(x,y)$. Using a Bayesian framework, we first specify a model of image formation, and then derive estimation algorithms from this model.

The Bayesian model of stereo image formation consists of two parts. The first part, a *prior model* for the disparity surface, uses a traditional Markov Random Field (MRF) to encode preferences for smooth surfaces [7]. It is specified as a Gibbs distribution p_P, the exponential of a potential function E_P:

$$p_P(\mathbf{d}) = \frac{1}{Z_P} \exp\left(-E_P(\mathbf{d})\right),$$

where $\mathbf{d}$ is the vector of all disparities $d_{i,j}$ and Z_P is a normalizing factor. The potential function itself is the sum of clique potentials which only involve neighboring sites in the field. In this paper, we study only first order fields, where

$$E_P(\mathbf{d}) = \sum_{i,j} \rho_P(d_{i+1,j} - d_{i,j}) + \rho_P(d_{i,j+1} - d_{i,j})$$

(see [16] for generalizations to higher order fields).

When $\rho(x)$ is a quadratic, $\rho(x) = x^2$, the field is a Gauss-MRF, and corresponds in a probabilistic sense to a first order regularized (*membrane*) surface model. When $\rho(x)$ is a unit impulse, $\rho(x) = 1 - \delta(x)$, it corresponds to a MRF that favors fronto-parallel surfaces [7]. In between these two extremes are functions derived from *robust statistics* [9], which behave much like surface models with discontinuities [3, 6, 2]. A wide variety of robust penalty functions are possible. In this paper, we use a contaminated Gaussian model,

$$\rho_P(x) = -\log\left((1-\epsilon_P)\exp(-x^2/2\sigma_P^2) + \epsilon_P\right).$$

The second part of our Bayesian model is the *data* or *measurement model* which accounts for differences in intensities between left and right images. This model assumes independent, identically distributed measurement errors,

$$p_M(I_L, I_R|\mathbf{d}) = \prod_{i,j} p_M(I_L(x_i + d_{i,j}, y_j) - I_R(x_i, y_j)).$$

Traditional stereo matching methods use either a squared intensity error metric (Gaussian noise), $\rho_M(x) = \log p_M(x) = x^2$, or an exact binary matching criterion $\rho_M(x) = 1 - \delta(x)$. In this paper, we again use a contaminated Gaussian model,

$$\rho_M(x) = -\log\left((1-\epsilon_M)\exp(-x^2/2\sigma_M^2) + \epsilon_M\right),$$

to model both Gaussian noise and possible outliers due to occlusions or effects such as specularities.

The posterior distribution, $p(\mathbf{d}|I_L, I_R)$ can be derived using Bayes' rule,

$$p(\mathbf{d}|I_L, I_R) \propto p_P(\mathbf{d}) p_M(I_L, I_R|\mathbf{d}).$$

As is often the case, it is more convenient to study the negative log probability distribution

$$\begin{aligned} E(\mathbf{d}) &= -\log p(\mathbf{d}|I_L, I_R) \qquad (1) \\ &= \sum_{i,j} \rho_P(d_{i+1,j} - d_{i,j}) + \rho_P(d_{i,j+1} - d_{i,j}) \\ &\quad + \sum_{i,j} \rho_M(I_L(x_i + d_{i,j}, y_j) - I_R(x_i, y_j)). \end{aligned}$$

While $p(\mathbf{d}|I_L, I_R)$ specifies a complete distribution, usually only a single optimal estimate of $d(x,y)$ is desired. The most commonly studied estimate is the

Maximum A Posteriori (MAP) estimate, which is equivalent to minimizing the energy given in (1).

Two popular techniques for minimizing equations like (1) are the Gibbs Sampler [7, 1] and mean field theory [6]. The Gibbs Sampler and its variants can produce good solutions, but at the cost of long computation times. Mean field techniques are not very good at modeling ambiguous estimates, such as multiple potential matches at each pixel.

4.1 Explicit local distribution model

Instead of using either of these two traditional approaches, we will develop a novel estimation algorithm based on modeling the probability distribution of $d_{i,j}$ at each site. To do this, we associate a scalar value between 0 and 1 with each possible discrete value of d at each pixel (i, j), and require that $\sum_d p(i, j, d) = 1$. Our representation is therefore the same as that used by diffusion-based algorithms, i.e., we explicitly model all possible disparities at each pixel.

To initialize our algorithm, we calculate the probability distribution for each pixel (i, j) based on the intensity errors between matching pixels, i.e.,

$$p_0(i, j, d) \propto \exp\left(-E_0(i, j, d)\right),$$

where

$$E_0(i, j, d) = \rho_M(I_L(x_i + d, y_j) - I_R(x_i, y_j))$$

is the matching cost of pixel (i, j) at disparity d.

Assuming (sub-optimally) independent distributions of adjacent disparity columns, we arrive at the update formula[1]

$$E(i, j, d) \leftarrow E_0(i, j, d) + \sum_{(k,l)\in\mathcal{N}_4} \log[-\sum_{d'} \exp\left(-\rho_P(d'-d) - E(i+k, j+l, d')\right)].$$

For notational and computational convenience, we will introduce a few more additional quantities. The *smoothed probability distribution*

$$p_S(i, j, d) = \sum_{d'} w_P(d' - d)p(i, j, d') \qquad (2)$$

is simply the current probability distribution $p(i, j, d)$ after it has been convolved *vertically* (in disparity) with the smoothing kernel $w_P(d) \propto e^{-\rho_P(d)}$, with $\sum_d w_P(d) = 1$. It has a corresponding *smoothed energy*

$$E_S(i, j, d) = -\log p_S(i, j, d). \qquad (3)$$

[1]Due to space constraints we have to omit the derivation, which can be found in [15].

Finally, the update rule can be written as a pair of equations

$$E(i, j, d) \leftarrow E_0(i, j, d) + \sum_{(k,l)\in\mathcal{N}_4} E_S(i+k, j+l, d), \quad (4)$$

$$p(i, j, d) \leftarrow \frac{e^{-E(i,j,d)}}{\sum_{d'} e^{-E(i,j,d')}}. \quad (5)$$

In practice, since the values of $E(i, j, d)$ are being updated simultaneously at all pixels and disparity, we use a modified version of (4),

$$E(i, j, d) \leftarrow E_0(i, j, d) + \mu[E_S(i, j, d) + \sum_{(k,l)\in\mathcal{N}_4} E_S(i+k, j+l, d)], \quad (6)$$

i.e., we weight the neighboring values less (we use $\mu = 0.5$) and include the current energy estimate.

If we interpret the above equations as a four-step algorithm for iteratively computing the best stereo matches, we see that they are a special instance of a non-linear diffusion process. The smoothing step in (2–3) blurs the current disparity probabilities vertically along a column, thereby enabling different nearby disparities to support each other (depending on the size of σ_P). It also adds a small amount to each probability (ϵ_P), which in effect limits the largest possible value that E_S can take and thus limits the effect of disparity discontinuities.

The update step (6) is identical to a regular diffusion step with β-terms (membrane model). However, the probability re-normalization step ensures that the energies represent meaningful log probabilities. The robust form of the E_0 function also ensures that bad matches have only limited effects, thus allowing for occlusions or other non-modeled errors to occur.

For the above algorithm to work well, the various parameters $\{\sigma_P, \epsilon_P, \sigma_M, \epsilon_M\}$ must be set to appropriate values. σ_M and ϵ_M are based on the expected noise in the image sensor, i.e., σ_M should be proportional to the regular image noise, while ϵ_M should be the probability of gross errors or occlusions (say 1–10%). The choice of σ_P depends on the class of disparity surfaces which may be expected, i.e., a small σ_P favors fronto-parallel surfaces. For the experiments presented in this paper, we set $\sigma_P = 0.1$ and $\epsilon_P = 0.01$.

Figure 4 shows the results of our probabilistic aggregation technique applied to the *ramp* and *rds* images. We use a different σ_M for the two image pairs: $\sigma_M = 2$ for *ramp*; $\sigma_M = 20$ for *rds*, to compensate for the different signal strengths of the two pairs. The other parameters are the same for both image pairs: $\epsilon_M = 0.1, \sigma_P = 0.1, \epsilon_P = 0.01$. The number of diffusion iterations is $n = 10$.

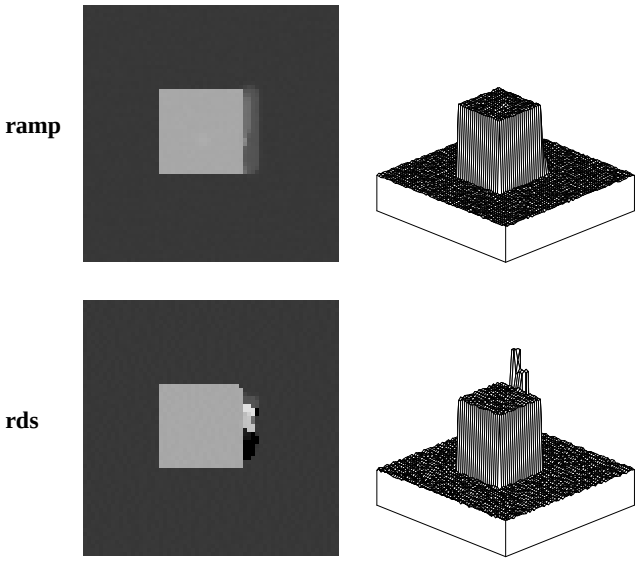

Figure 4: Performance of the probabilistic model on the *ramp* and *rds* image pairs.

5 Experimental results

In this section we numerically evaluate the performance of the different algorithms on synthetic images. We also show results for real image data.

For our experiments we use five synthetic image pairs, including the two pairs from Figure 1. The additional pairs are based on combinations of a new disparity pattern *bars* consisting of two rectangular regions with two different disparities, and a new intensity pattern, *real*, which is taken from a real image depicting ground covered with grass [15].

We compared the following algorithms: SSD, diffusion using the membrane model, diffusion with local stopping, and diffusion using the probabilistic model. For each parameter setting, we ran each algorithm on a test set of 40 images (the 5 image pairs with 8 different levels of additive Gaussian noise). We tried more than 70 different parameter settings, resulting in about 3000 experiments. In each experiment, we compared the computed disparities with the true disparities (ignoring the occluded regions), and collected three different error statistics.

First we analyzed the error statistics for each method separately to gain understanding of the effect of the different parameters. Then we chose the best parameters for each method, and compared the different methods with each other. In this paper we only report a subset of our results based on the root-mean-square (RMS) disparity error. See the long version of this paper for a more detailed analysis [15].

SSD, which we include for comparison, has only one parameter: the size of the support region. The same holds for simple diffusion, where the size of the support region is controlled by the number of iterations. Not surprisingly, higher noise levels require bigger window sizes. The best window size can also depend on the image.

The membrane model behaves similarly to regular diffusion with a fixed number of iterations. For small noise levels, a value of β between 1/3 and 1 usually yields smaller errors than regular diffusion, but not always. Also, as mentioned before, for high noise levels, β needs to be chosen quite small to enable enough smoothing for stable matching.

In analyzing regular diffusion with local stopping criteria, we found that the certainty measure is critical. In our experiments, the winner margin C_m almost always outperformed the measure based on entropy C_e. A problem with our definition of local stopping is that an initial wrong but "certain" match can survive. There is clearly a potential for both better certainty measures and different stopping criteria.

The probabilistic model, which performed by far the best, also has the most parameters. We found, however, that many parameters have only small effects and can be set to default values, including $\epsilon_M = 0.1, \epsilon_P = 0.01$, and $\mu = 0.5$. As expected, a small σ_P worked best for our test images composed from fronto-parallel surfaces. For real images, we found that σ_P needs to be chosen slightly higher. The most important parameter is σ_M, which should reflect the strength of the image signal. We used three different values for the three different textures of our test images. Finally, the number of iterations is less critical, since the method seems to converge relatively fast to a stable solution. Higher numbers of iterations are necessary for images containing regions of uniform intensity, such as the real images discussed below.

For direct comparison of the methods, Figure 5 shows the RMS disparity error versus the noise level for three image pairs. We compare SSD with a window size of 5, the membrane model with $\beta = 0.5$, diffusion with local stopping based on winner margin C_m, and the probabilistic model with $\epsilon_P = 0.01, \sigma_P = 0.1, \epsilon_M = 0.1$, and $\sigma_M = 2, 8, 20$, for *ramp*, *real*, and *rds* textures respectively. The number of iterations is 10 for all methods.

The probabilistic model clearly beats the three other methods. For small noise levels, the occlusion boundaries are recovered almost perfectly. Note that the algorithm recovers the "correct" disparity pattern, even though the notion of true disparities is not well defined for ambiguous images such as random dot stereograms.

We also tested our algorithms on real images. We include results of the probabilistic method on images from the SRI's tree sequence and CMU's town se-

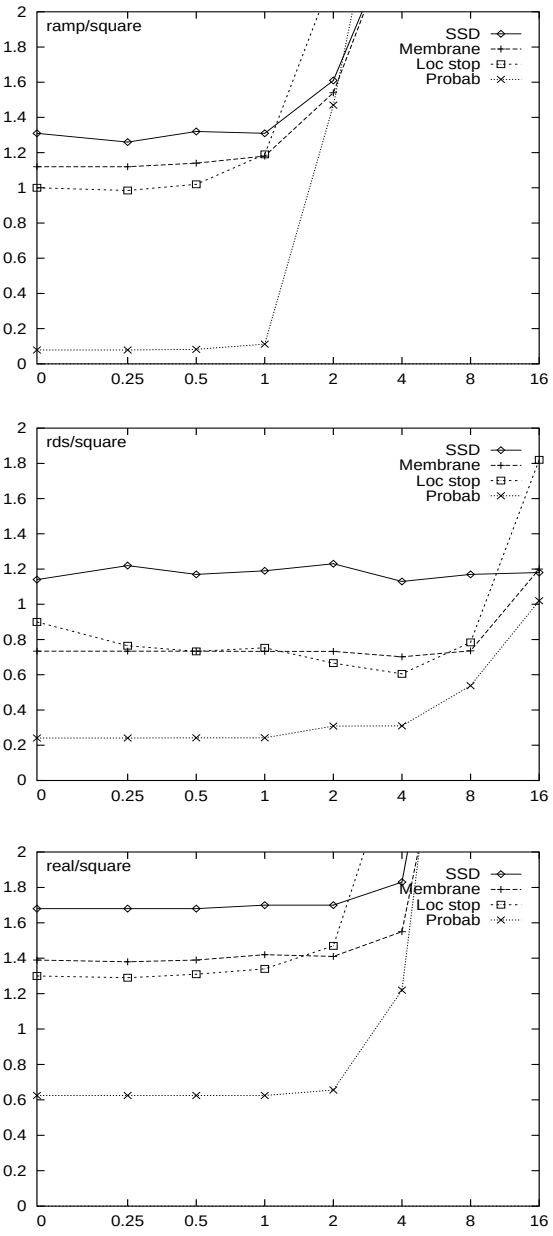

Figure 5: RMS error of the computed disparities versus the standard deviation of image noise for three synthetic image pairs. The error at occluded points is not included.

quence. We used multiple baseline stereo based on five images to initialize the disparity space with the sum of four (appropriately scaled) similarity measures [11]. Figure 6 shows the disparity maps computed by the probabilistic algorithm after 50 iterations, using the following parameters: $\sigma_P = 0.4, \epsilon_P = 0.01, \sigma_M = 5, \epsilon_M = 0.1$. Note that we use a bigger σ_P than before to account for slanted surfaces.

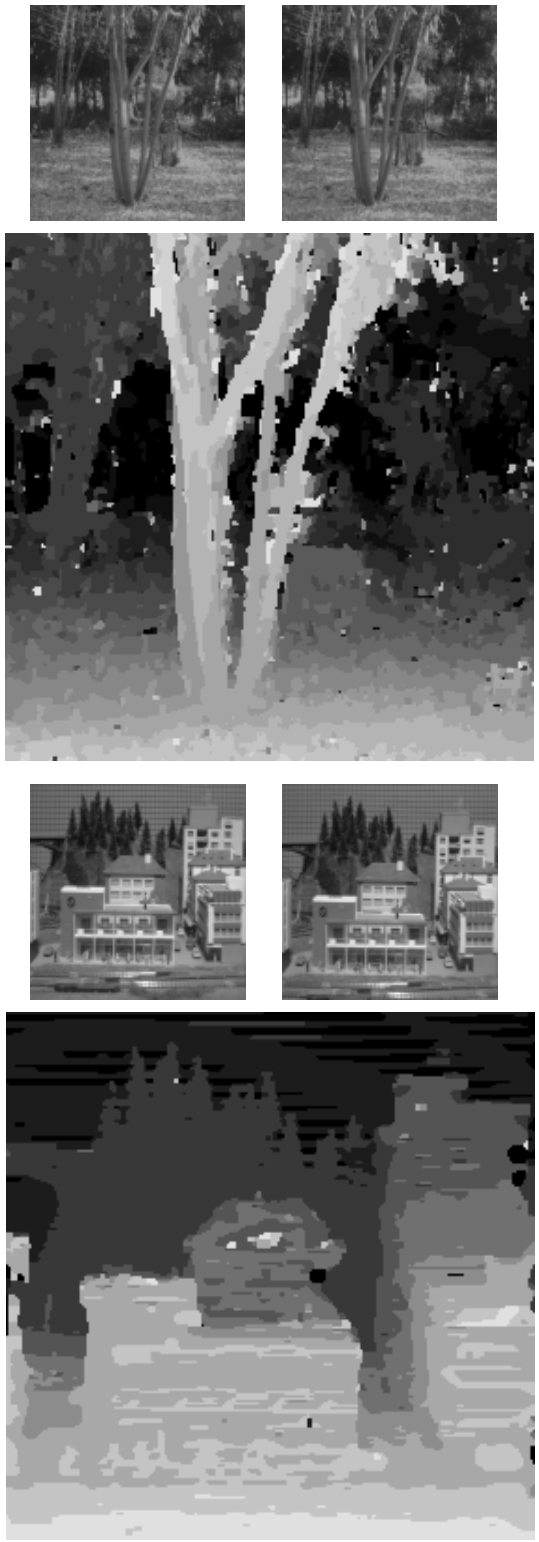

Figure 6: Disparities for tree and town images computed by the probabilistic algorithm.

6 Discussion

As we have shown, linear and non-linear diffusion algorithms are an attractive alternative to the adaptive windows introduced by Kanade and Okutomi [10]. In its simplest form, the membrane algorithm simply requires the iterative summation of neighboring matching costs, with an additional term thrown in to prevent the support region from growing indefinitely. The increased weighting of the central pixel relative to the periphery is sufficient to counteract many of the artifacts introduced by the squared summing window used in SSD. When combined with a local stopping criterion, the resulting non-linear diffusion process has an adaptive support behavior similar to the variable window size algorithm. The inclusion of additional non-linearities in the Bayesian diffusion algorithm improves the performance even more.

In addition to their simplicity and computational efficiency, our non-linear diffusion algorithms can also handle stereograms with more ambiguity than the adaptive window SSD algorithm. Kanade and Okutomi's algorithm is based on locally adjusting the subpixel disparity estimate simultaneously with growing the window size. This presupposes that the algorithm is somehow initialized in the vicinity of the true disparity. This is achieved in their synthetic image sequences by using small disparities, and in their real sequences by using a multi-frame version of the basic SSD algorithm [11]. Image pairs with rapidly varying textures and many potential matches such as the random-dot stereograms used in our experiments could not be handled by their current algorithm.

In future work, we plan to study better local stopping criteria based on improved certainty measures. We would also like to investigate multi-resolution versions of our diffusion algorithms to help fill in regions which have few features to match.

7 Conclusions

In this paper, we have demonstrated that diffusion-based aggregation of support is a useful alternative to both traditional area-based correlation and to more recent adaptive window size-based techniques. Our algorithms are simple to implement and computationally efficient, and result in better quality estimates, especially near discontinuities in the disparity surface. The addition of local termination conditions to the basic diffusion process results in a behavior similar to that of adaptively sized windows. Furthermore, our novel non-linear diffusion algorithm derived from a Bayesian model of stereo matching results in markedly improved performance. We believe that further study of the basic support and aggregation methods in stereo matching is central to developing algorithms with improved performance over a wide range of imagery.

References

[1] S. T. Barnard. Stochastic stereo matching over scale. *Intl. Journal Comp. Vision*, 3(1):17–32, 1989.

[2] M. J. Black and A. Rangarajan. The outlier process: Unifying line processes and robust statistics. In *IEEE Conf. on Comp. Vision and Pattern Recog.*, pages 15–22, June 1994.

[3] A. Blake and A. Zisserman. *Visual Reconstruction.* MIT Press, Cambridge, MA, 1987.

[4] R. C. Bolles, H. H. Baker, and D. H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *Intl. Journal Comp. Vision*, 1:7–55, 1987.

[5] U. R. Dhond and J. K. Aggarwal. Structure from stereo—a review. *IEEE Trans. on Systems, Man, and Cybern.*, 19(6):1489–1510, Nov./Dec. 1989.

[6] D. Geiger and F. Girosi. Mean field theory for surface reconstruction. *IEEE Trans. Pattern Anal. and Machine Intel.*, 13(5):401–412, May 1991.

[7] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. and Machine Intel.*, 6(6):721–741, Nov. 1984.

[8] W. E. L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Trans. Pattern Anal. and Machine Intel.*, 7(1):17–34, Jan. 1985.

[9] P. J. Huber. *Robust Statistics.* John Wiley & Sons, New York, NY, 1981.

[10] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Trans. Pattern Anal. and Machine Intel.*, 16(9):920–932, Sept. 1994.

[11] M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE Trans. Pattern Anal. and Machine Intel.*, 15(4):353–363, 1993.

[12] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby. PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14:449–470, 1985.

[13] K. Prazdny. Detection of binocular disparities. *Biological Cybernetics*, 52(2):93–99, 1985.

[14] D. Scharstein. Matching images by comparing their gradient fields. In *Intl. Conf. on Pattern Recogn.*, volume 1, pages 572–575, Oct. 1994.

[15] D. Scharstein and R. Szeliski. Stereo matching with non-linear diffusion. Computer Science TR 96-1575, Cornell University, Mar. 1996.

[16] R. Szeliski. *Bayesian Modeling of Uncertainty in Low-Level Vision.* Kluwer Academic Publishers, Boston, MA, 1989.

[17] R. Szeliski and G. Hinton. Solving random-dot stereograms using the heat equation. In *IEEE Conf. on Comp. Vision and Pattern Recog.*, pages 284–288, June 1985.