### A Correlation-Based Model Prior for Stereo

Yanghai Tsin Real-Time Vision and Modeling Siemens Corporate Research yanghai.tsin@siemens.com

# Takeo Kanade School of Computer Science Carnegie Mellon University tk+@cs.cmu.edu

### **Abstract**

All non-trivial stereo problems need model priors to deal with ambiguities and noise perturbations. To meet requirements of increasingly demanding tasks such as modeling for rendering, a proper model prior should impose preference on the true scene structure, while avoiding artificial bias such as fronto-parallel. We introduce a geometric model prior based on a novel technique we call kernel correlation. Maximizing kernel correlation is shown to be equal to distance minimization in the M-estimator sense. As a model prior, kernel correlation is demonstrated to have good properties that can result in renderable, very smooth and accurate depth map. The results are evaluated both qualitatively by view synthesis and quantitatively by error analysis.

#### 1. Introduction

Stereo vision has been one of the central topics in computer vision. Despite the contemporary development in range sensing, such as by laser range finders or by structured light techniques, stereo vision continues to be important as a passive and rich sensor that provides both photometric and geometric information.

Discrete stereo vision, usually formulated as labeling each pixel in an image with a small set of disparities, has matured due to the recent progress in energy minimization techniques such as graph cut [5] and belief propagation [21]. Two comprehensive reviews of the state-of-the-art are available [18, 7]. In [18] Scharstein and Szeliski primarily address the performance issue, while in [7] Brown *et. al.* put emphasis on correspondence methods, speed, and occlusion detection methods.

However, discrete depth outputs cannot always satisfy all 3D modeling requirements. For example, for rendering purpose, it is very difficult to synthesize from discrete depth map visually acceptable images from viewing angles other than a small neighborhood around the original view point. We give such an example in Figure 1, first row. In this case we are rendering using the hand labeled "ground-truth" disparity and we get nothing more than two parallel planes.

A common conception for such imperfection is that

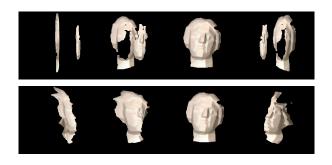


Figure 1: Rendered using: first row, the hand-labeled "ground-truth" disparity (from the Tsukuba stereo dataset [15]); second row, disparity of the new method.

their isn't enough visual support for a better reconstruction. There is little texture, and the total disparity difference in the whole region is just one pixel. We demonstrate in this paper that such coarse disparity estimation is not the limit that a computational method can achieve. By using a model prior proposed in this paper, a continuous valued, more accurate 3D model is computable from just two views (second row of Figure 1).

In the next section, we introduce our new model prior. This is followed by introduction of the new stereo vision framework. In Section 4 we compare our model prior with existing methods. We give experimental results and performance evaluation in Section 5. Finally the paper is closed by conclusions and discussions.

### 2. Kernel Correlation

### 2.1. Kernel Correlation Definition

Kernel correlation (KC) between two points  $x_i$  and  $x_j$  is defined as the correlation of kernel functions (K(x,y)) centered at the two points,

$$KC(x_i, x_j) = \int K(x, x_i) \cdot K(x, x_j) dx.$$
 (1)

We limit ourselves to kernel functions that are usually used in the Parzen window density estimation [16], instead of the broader "kernel" definition in machine learning community. Furthermore, for ease of illustration we focus our discuss on the isotropic Gaussian kernel:

$$K_G(x, x_i) = (\pi \sigma^2)^{-D/2} \exp(-\|x - x_i\|^2 / \sigma^2).$$
 (2)

Here ||x-y|| is the Euclidean distance between two vectors x and y, and D is the dimension of the vectors. KC of Gaussian kernels has a very simple form,

$$KC_G(x_i, x_j) = (2\pi\sigma^2)^{-0.5D} \exp\left\{-\|x_i - x_j\|^2 / 2\sigma^2\right\}.$$
(3)

Further, we define KC between a point  $x_i$  and a point set  $\mathcal{X}$  as the *leave-one-out KC*,

$$KC(x_i, \mathcal{X}) = \sum_{x_i \in \mathcal{X}, x_i \neq x_i} KC(x_i, x_j),$$
 (4)

and define KC of the whole point set as

$$KC(\mathcal{X}) = \sum_{x_i \in \mathcal{X}} KC(x_i, \mathcal{X}) = \sum_{x_i \in \mathcal{X}} \sum_{x_j \neq x_i} KC(x_i, x_j).$$
 (5

The form  $e_{ij} = \exp\{-\|x_i - x_j\|^2/2\sigma^2\}$  is known as the *affinity* or *proximity* between points and it has biological significance [28]. Symmetric affinity matrix in the form of

$$\begin{bmatrix} 1 & e_{12} & \dots & e_{1j} & \dots & e_{1N} \\ e_{21} & 1 & \dots & e_{2j} & \dots & e_{2N} \\ \dots & \dots & \dots & \dots & \dots \\ e_{N1} & e_{N2} & \dots & e_{Nj} & \dots & 1 \end{bmatrix}$$
(6)

has been used by Scott and Longuet-Higgins [19] to determine correspondences between two point sets. When the kernel is Gaussian, the three definitions of KC can have the following correspondence to the affinity matrix. KC between points  $x_i$  and  $x_j$  is proportional to the  $e_{ij}$  and  $e_{ji}$  terms in the affinity matrix. The leave-one-out KC  $(KC(x_i,\mathcal{X}))$  is the sum of the  $i^{th}$  row or  $i^{th}$  column excluding the diagonal element, and KC of the point set  $KC(\mathcal{X})$  is the sum of all the off-diagonal elements.

### 2.2. Properties of Kernel Correlation

#### 2.2.1 MaxKC for Robust Distance Minimization

According to (3) and (4), for the Gaussian case, we have

$$KC_G(x_i, \mathcal{X}) \propto \sum_{x_j \in \mathcal{X}, x_j \neq x_i} \exp\{-\|x_j - x_i\|^2 / 2\sigma^2\}.$$
 (7)

This is a function of distances between pairs of points. Note that the influences of  $x_j$ 's on  $x_i$  are different. Distant points have negligible contributions, and only neighboring points have influences. In fact, maximizing KC is equivalent to minimizing distance between  $x_i$  to all  $x_j$ 's in the M-estimator [10] sense. The benefit of minimizing (7) is

that the update of  $x_i$  is robust against distant points; for example, 3D points across the depth discontinuity. To see this we study the local update rule of KC. A necessary condition for a local maximum at step n+1 is,

$$\frac{\partial KC_G(x_i, \mathcal{X})}{\partial x_i} \propto \sum_{x_i \in \mathcal{X}, x_j \neq x_i} w_{ij}^{(n)}(x_i - x_j) = 0,$$

where  $w_{ij}^{(n)}=\exp{\{-\|x_i^{(n)}-x_j\|^2/2\sigma^2\}}$ . This leads to the weighted average updating rule

$$x_i^{(n+1)} = \sum_{j \neq i} w_{ij}^{(n)} x_j / \sum_{j \neq i} w_{ij}^{(n)}.$$
 (8)

Now it is easy to see that  $\lim_{\|x_j - x_i\| \to \infty} w_{ij}^{(n)} x_j = 0$ , or distant points have no influence on local updates. As a result Gaussian KC maximization is robust to points from a different structure.

#### 2.2.2 KC Maximization by Local Updates

Another appealing property of KC is that  $KC(\mathcal{X})$  can be maximized by maximizing the leave-one-out KC  $(KC(x_i,\mathcal{X}))$  iteratively over all  $x_i$ . Maximizing  $KC(x_i,\mathcal{X})$  simultaneously increases the sum of the  $i^{th}$  row and  $i^{th}$  column (excluding diagonal) of the affinity matrix, while leaving all other elements intact. As a result,  $KC(\mathcal{X})$ , which is proportional to the total sum of the off-diagonal elements in the affinity matrix, is increased. This property makes it possible to maximize KC by local updates.

Since  $KC(\mathcal{X})$  is bounded for a finite point set, iterative update of leave-one-out KC is guaranteed to converge to a fixed point.

#### 2.2.3 KC as an Object Space Model Prior

Kernels are defined directly in the object space in our stereo formulation. To see the smoothing effect by the KC prior, we give a 2D example. As in Figure 2, imagine we have a 2D scene composed of two disjointed sinusoid depth curves (green solid curves). Imagine that the red circles  $(x_i, z_i)$ are the noisy observation of the true scene. We compare three smoothing methods. The first is the weighted average method over the immediate 2-neighbors. The smoothed result is plotted as cyan diamonds. The output is still noisy and there is oversmoothing at the discontinuity. The second is kernel weighted average ([9]) using a Gaussian kernel with scale  $\sigma = 4$ . The output is plotted as blue crosses. The noise is sufficiently suppressed in this case. However, the oversmoothing across boundary is even worse. Finally, we use KC with kernel scale  $\sigma = 4$ . The output at a point  $x_i$  is defined as the maximum leave-one-out KC position:  $argmax_z KC((x_i, z), \mathcal{X})$ , where  $\mathcal{X}$  is the observed noisy point set in 2D. The KC smoothing result is shown as black dots in the figure. Notice that noise is sufficiently suppressed, and at the same time depth discontinuity is preserved.

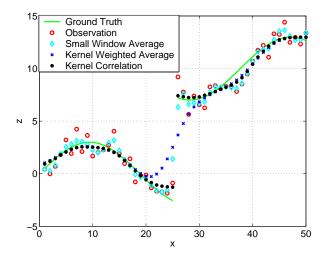


Figure 2: Comparing smoothing methods. For the kernel weighted average (K.W.A.) and kernel correlation methods,  $\sigma = 4$ .

The depth discontinuity preservation property of KC is due to its nature that pairs of points with large distance don't interact. In the above example, points across the structure may have similar reference view distances that cause oversmoothing in the first two methods, but their large distances in the object space prohibit them from interacting in the KC method.

We conclude that 1) sufficiently large window averaging is needed in order to handle noise; 2) object space smoothing with a robust mechanism is needed in order to preserve discontinuity. We will show in Section 4 that, unfortunately, existing model priors don't possess both of these properties.

### 2.3. Approximating Kernel Correlation

Direct computation of  $KC(\mathcal{X})$  requires enumerating all pairs of points, an  $O(N^2)$  computation. Fortunately, for the kernels we adopt the influence of a point is negligible beyond a certain distance. Thus we can approximate KC using only local interactions. For the case of Gaussian kernel, we have two choices. First, we can approximate KC by using (7) and (5). We find the neighbors of  $x_i$  within a radius and approximate the leave-one-out KC within the window. The second way is to use discretized version of (1), the original definition of correlation. The second approach is especially useful when the model prior is applied to multiple view reconstruction, where points from different views interact. In this case nearest neighbor finding in a large dynamic dataset can be a difficult problem itself.

### 3. Kernel Correlation in Stereo

### 3.1. The New Stereo Energy Function

Our new energy function follows the general energy function framework: weighted contribution from evidence (color matching error) and model prior. But we define the model prior as the kernel correlation of the reconstructed point set,

$$E_{KC}(\boldsymbol{d}) = \sum_{i} C(x_i, d_i) - \lambda \cdot KC(\mathcal{X}(\boldsymbol{d})).$$
 (9)

Here  $d = \{d_i\}$  is the set of depths to be computed.  $\mathcal{X}(d)$  is the point set obtained by projecting the pixels into 3D according to the depth map d.  $\lambda$  is a weighting term. The evidence term  $C(x_i, d_i)$  is determined by the color in the reference view,  $d_i$  the depth at pixel  $x_i$ , and colors of the corresponding pixels in the other visible views. As we have seen in Section 2.2.3, our prior has the ability to suppress noise by using a large window, and because it is defined in 3D space it can preserve discontinuity at the same time.

#### 3.2. Solving the Energy Function

Minimizing the energy function (9) is not trivial. It is a continuous value optimization problem and discrete optimization methods like max-flow graph cut do not apply. If we are content with a discrete solution, we show in Appendix C of [27] that the energy function belongs to the energy function group  $\mathcal{F}^2$  [13]. But the energy function (9) violates the *regularity condition* when the kernel scale in the depth direction is too large, namely, greater than  $0.91\Delta d$ , where  $\Delta d$  is the disparity resolution.

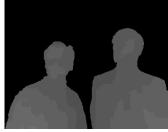
In this paper we report stereo results obtained by minimizing (9) with iterative gradient descent. At each step we sequentially update the depth of each pixel by minimizing the sum of the color matching error and the negative leave-one-out KC. From Section 2.2.2 we know that the energy function (9) decreases at each step. Thus the convergence of the gradient descent approach to a fix point is guaranteed.

In our experiments, we initialize our algorithm using a simple stereo method: intensity window correlation. Figure 3 demonstrates the efficacy of our algorithm. The initial results provided by a  $11 \times 11$  window correlation were very noisy. Note that the depth map of the final result is accurate enough that it captures the disparity difference between the tie and shirt of the right person. (Please see the electronic version of the paper if the printed version is of poorer quality.) For comparison we show the reconstruction by a graph cut method in the fourth column, where discretization effect is quite obvious.









Reference image

Initialization by intensity window correlation

Final result

Result by graph cut + Potts

Figure 3: Stereo results by minimizing the new KC energy function.

# 4. Comparing Existing Priors

This section serves to compare the KC model prior to existing methods. For an in-depth review of various assumptions made in stereo, the reader is referred to [11].

### 4.1. Implicit Fronto-Parallel Model

We first compare KC stereo with intensity window correlation (IWC) [11]. In an IWC algorithm, matching errors are spatially averaged in a local window to produce evidence supporting different depth hypotheses. The IWC implicitly assumes local constancy of disparity, or the frontoparallel model prior. To achieve a smoother reconstruction in IWC, larger windows are required. However, pixels in a large window may violate the fronto-parallel assumption, resulting in matching errors. Even worse, large window may straddle depth discontinuity and result in reconstruction failure. The coupling of intensity matching and model prior in an IWC algorithm thus makes it difficult to ensure both accurate color matching and smooth 3D model. In contrast, the "correlation" of KC is defined on the geometry alone and color matching is defined on individual pixels and their correspondences. We can enforce very smooth 3D shapes by adjusting both the weight  $\lambda$  and increasing the size of the correlation window, yet still ensure depth discontinuity due to the embedded M-estimator mechanism. Also due to the decoupling, color matching error can be kept accurate.

KC stereo shares the same assumption with the very successful two-step algorithm [20] in image-based rendering: Geometric variation is slower than color changes. If the assumption is violated, both rendering and reconstruction are difficult for any known algorithm and the discussion is beyond the scope of this paper. Under such assumption, it makes sense to enforce smooth geometric models while keeping intensity sampling sharp. In this sense KC stereo is the converse of the two step rendering algorithm: reconstruction versus rendering.

### 4.2. Locally Planar Embedding Model

There exist stereo algorithms that compare warped intensity patches according to local geometry, such as the oriented particle method [8]. However, it still assumes locally planar embedding. It remains difficult to decide the window size and orientation, and it is not known how to fill the gaps between neighboring oriented particles.

### 4.3. Segmentation Based Methods

We have seen several successful examples of stereo matching proceeded by a segmentation step [3, 25, 14]. The underlying assumption is that the segmented regions can each be fit by a parametric model. The segmentation can be color [25] or graph-cut based [3, 14]. If the image is oversegmented, a merging step is applied after the depth for each region is estimated. The advantage of the KC prior is that it is not restricted to any single parametric form. As a result, it does not need to know in advance the parametric models, or the number of regions. This directly results in the automatic discovery of the crease (highlighted by circles) in Figure 10, which is missed by [14].

#### 4.4. Parametric Model

Next we compare stereo algorithms that adopt parametric model priors. When scene geometry is simple enough and can be specified by user interaction, stereo is transferred to a model fitting problem. Reconstruction of very good polyhedral architectural models are reported in [26]. For scenes with complex geometry it is impossible to adopt such an approach. More general priors such as those based on splines [22] have also been proposed. Their success depends on choices such as functional forms and support of each function. The adaptive spline method [23] partially addresses the support issue, with the introduced difficulty of "cracks" at the boundary of splines. KC model prior is non-parametric and very general. KC prior with large kernel scale share the benefits of a parametric model: Contri-

butions of pixels in a large region are considered.

#### 4.5. Disparity Gradient Model

The next class of model priors are defined by constraining the magnitudes of the first or second order derivatives of the reconstructed structure [12, 4, 1]. A derivative-based prior term generally has two problems: over-smoothing and the Gibbs effect (ringing or over-shooting at the discontinuity region). To handle depth discontinuity, the model prior needs to couple with a line process.

#### 4.6. Potts Model

Current state-of-the-art stereo algorithms usually adopt the Potts model prior [5]. The combination of Potts model and graph cut algorithm produces the best performing stereo algorithms as reported in [18]. By using the Potts model, we hope that when we introduce finer discretization of disparity, we would get more and more accurate 3D reconstruction. We show in Figure 4 (first two columns) that this is not true. We increased the number of disparity levels for about one hundred times and we still get approximately the same discrete fronto-parallel model. If we reduce the relative weight (corresponding to our  $\lambda$ ), noise will dominate very quickly. KC model prior on the other hand supports very smooth and accurate reconstruction (Figure 4, last column).

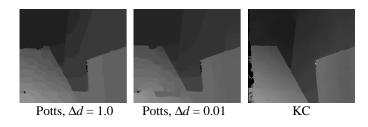


Figure 4: Potts model does not result in more accurate depth map with finer disparity resolution (first two columns), while by using the KC model prior an algorithm outputs much smoother and yet more accurate results, including along discontinuities.

The first goal of introducing a model prior is to propagate confident estimation into ambiguous regions. Figure 5 gives an illustration why KC prior works better than the Potts prior at fine resolution. Let us assume scene points A and C have unique texture and can be accurately determined by color matching, while color matching for point B is ambiguous or with low confidence. Suppose the disparity discretization is fine enough and the three points are at three different disparity levels,  $L_1, L_2, L_3$ . The right part of Figure 5 gives the energy corresponding to the Potts energy and the KC energy as a function of disparity d. The Potts model

prior gives a constant energy when B is labeled either as  $L_1$  or  $L_3$  because B has a different disparity from either A or C. When B is labeled as  $L_2$ , the true disparity, however, the Potts model gives double penalty because B has a different disparity from both A and C. The Potts model prior is contradicting with the true scene structure in this case. Potts prior introduces unwanted fronto-parallel bias. In contrast, KC prior with large enough kernel scale is consistent with the true scene structure. KC model prior thus helps to propagate confident estimation into ambiguous region.

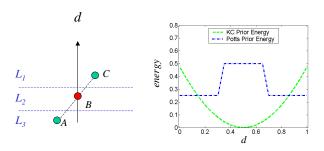


Figure 5: Propagation of confident estimation. Points A and C have unique texture. Confident estimation at A and C are propagated to B by the KC model prior, but not by the Potts prior.

The second goal of introducing a model prior is to suppress noise. Disparity estimation from stereo images alone are extremely noisy. This makes the noise-suppression role of a model prior crucial. There are two ways of suppressing the noise. The first way is increasing the weight. Doing so, however, magnifies both good preferences (such as smoothness) and bad biases (such as fronto-parallel). With increasing weight, a Potts stereo algorithm will output less and less disparity levels, until finally a single fronto-parallel reconstruction. The second way is to allow a pixel to interact with a large window of neighbors. With increased window size, the average behavior within a large set of points can be extracted. We show in our experiments that very clean reconstructions can be achieved by this second way. Unfortunately, existing geometric priors are usually defined on immediate neighbors only, possibly due to the difficulty to analyze complex Markov random fields with links beyond immediate neighbors. We believe interaction with only immediate neighbors is one of the reasons why those methods produce much noisier models than ours. KC, on the other hand, allows for increasing the window size, yet preventing unwanted effects from too far points because of its robustness property shown in 2.2.

## 5. Experiments

#### 5.1. Qualitative Results

Our first example was given in Figure 1, a head model contained in the Tsukuba dataset [15]. From two views,we recovered a more realistic 3D shape. Here we see a renderable reconstruction from a frontal pair.

Our reconstructed foreground model for the five frame Dayton sequence was compared with two other methods in Figure 3. We choose Gaussian kernels with scale 1.5 in both (u,v) directions (image plane) and the disparity direction, and  $\lambda=200.0$ . The kernel is defined in 3D space. To approximate the leave-one-out KC, pixels within a  $7\times 7$  window are projected to 3D and their influence on the center point is considered. Each update of gradient descent takes about 10 seconds on a 2.2 GHz PC and the algorithm converges after 20 steps. The reference view (view 3) is warped to view 1 and view 5 in Figure 6. The synthesized images are almost indistinguishable from the original ones.

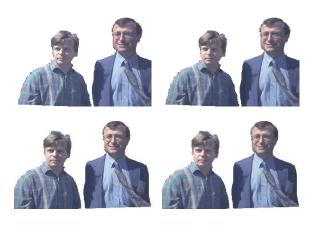


Figure 6: Left: original image. Right: synthesized image.

Next, we apply our new stereo formulation in the Euclidean space. We work with five frames of the Lodge sequence [6], whose reference view is shown in Figure 7. The depth of the scene is between  $[5\times 10^6, 9\times 10^6]$ . We choose depth increment to be  $5\times 10^4$ , which corresponds to approximately 80 discrete depth levels in the whole range. The 3D kernel is chosen to be isotropic with kernel scale  $\sigma=7.5\times 10^4$ , and we choose  $\lambda=10$ . Figure 8 shows several synthesized images using the reconstructed model. The smooth shape of the building and the occlusion in the scene are accurately captured by our single reference view algorithm.

#### 5.2. Quantitative Evaluation

#### 5.2.1 Bad Pixel Statistics

We present quantitative evaluation of the performance of our new stereo algorithm with real image sequences. We





Figure 7: Lodge sequence: the reference view and recovered depth.













Figure 8: Synthesized views for the Lodge sequence.

use four standard test images with ground-truth data used in [18]. A disparity map is evaluated by the percentage of "bad pixels", where a bad pixel is a pixel whose estimated disparity has an estimated error greater than 1 (not including 1).

To avoid large color matching errors due to aliasing, we adopt the color matching method introduced by Birchfield and Tomasi [2]. To eliminate the "foreground-fattening" effect at depth discontinuity areas, we incorporate the static cues [5] in our energy function. All results reported are generated by using the same set of parameters. We set the kernel scale to be  $\sigma_{uv}=4$  in the directions parallel to the image plane,  $\sigma_d=0.5$  in the disparity dimension (an anisotropic kernel), and  $\lambda=10$ . Pixels within a  $13\times13$  window contributes to the KC value. We observe very clean disparity maps (Figure 9) in all four cases.

To quantify our results, we count the percentage of the bad pixels in three regions: all valid regions, textureless regions and depth discontinuity regions. Percentages of bad pixels in these three regions are listed in Table 2. The numbers in parentheses are the ranks of our algorithm within the top 20 best performing algorithms. Given the gradient descent optimization strategy we are using, we consider this

performance satisfactory.

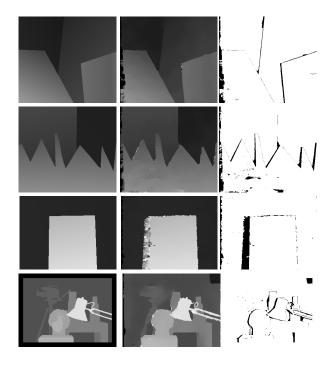


Figure 9: Estimated disparity maps. Left to right: ground-truth; our result; bad-pixels;

Table 1: *Performance of the kernel-correlation based stereo algorithm.* 

	Tsukuba	Map	Sawtooth	Venus
All	2.21 (8)	0.52 (11)	1.16 (6)	0.86(2)
Discontinuity	7.66 (3)	5.98 (11)	3.99 (4)	5.07 (3)
Textureless	1.99 (9)		0.58 (12)	0.86 (3)

Table 2: Bad Pixel Statistics. The numbers in the parentheses are the ranks of our algorithm compared to the top 20 best stereo algorithms as of May 22nd, 2003, time of the experiment.

#### 5.2.2 Good Pixel Statistics

To show the advantage of our algorithm over the graph cut method, we take a cross-section of the estimated disparity of the Venus pair. Together we show the result of the swapping graph cut algorithm [5] in the same plot, Figure 10. The discretization effect is clearly visible for the graph cut method, while our result is a much better approximation to the ground-truth. Also, pay attention to the rightmost part of the plot (columns > 400). The ground-truth disparity clearly shows a increasing trend of disparity, caused by a fold in the scene object (highlighted by the circles). Our

reconstruction follows the change closely. But graph cut produces a constant disparity in the whole region.

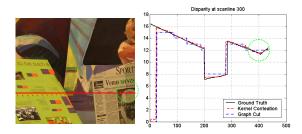


Figure 10: Disparity of scanline 300 of the Venus pair.

To show that this improved accuracy is not an isolated phenomenon, we study the statistics of the "good pixels", or the pixels whose disparity estimation error is less than or equal to 1. We show histograms and standard deviations of the estimation errors in Figure 11. We do not compare the Tsukuba dataset because the dataset does not have sub-pixel ground-truth disparity map. The first row shows the results using our new algorithm, while the second row shows the results generated by the swapping graph cut algorithm [5], implemented by Scharstein and Szeliski [18]. In all cases the errors of our method have smaller standard deviations than those generated by graph cut, especially in the map pair and the Venus pair, where our results have a standard deviation of less than half of the standard deviation of the graph cut algorithm.

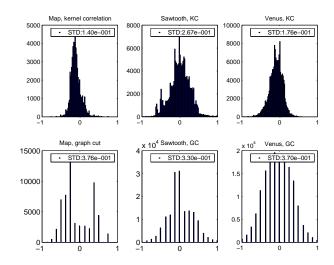


Figure 11: Histogram of the disparity estimation errors of "good" pixels. First row: KC stereo algorithm. Second row: graph cut.

# 6. Summary

We introduced the kernel correlation based model prior in this paper. It is equal to an M-estimator with a large interaction window. KC model prior shows better capability in resolving ambiguity and suppressing noise. Very accurate and smooth reconstructed models are reported.

There are some open issues in the approach. The first one is the kernel scale selection. The scales are selected empirically in our experiments and 3D points at different disparities use the same kernel function. This is a problem for distant points whose disparity differences tend to be small. In our experiments we observe oversmoothing in the background, such as the camcorder in the Tsukuba set (last row, center image of 9). The second problem is the speed problem. Our current code (not optimized) needs 10 to 30 seconds for each iteration. However, we see very large potentials for speeding up the algorithm because the algorithm is highly parallel. Third, we need to study better energy minimization algorithms for the new stereo formulation. Finally, Szeliski and Scharstein [24] demonstrated that stereo matching can be improved by correctly interpolating the color matching error (the first term in our energy function (9)). An integrated consideration of both terms is of our interest.

### References

- [1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *IJCV*, 2(3):283–310.
- [2] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE TPAMI*, 20(4):401–406, April 1998.
- [3] S. Birchfield and C. Tomasi. Multiway cut for stereo and motion with slanted surfaces. In *ICCV'99*, pages 489–495, September 1999.
- [4] T. E. Boult. What is regular in regularization? In *ICCV'87*, pages 457–462.
- [5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE TPAMI*, 23(11):1222–1239, November 2001.
- [6] A. Broadhurst, T. W. Drummond, and R. Cipolla. A probabilistic framework for space carving. In *ICCV 2001*, volume I, pages 388–393.
- [7] M.Z. Brown, D. Burschka, and G.D. Hager. Advances in computational stereo. *IEEE TPAMI*, 25(8):993–1008, August 2003.
- [8] P. Fua. Reconstructing complex surfaces from multiple stereo views. In CVPR'95, pages 1078–1085.
- [9] T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning, Data Mining, Inference and Prediction. Springer, 2001.

- [10] P. J. Huber. *Robust Statistics*. John Wiley & Sons, New York, New York, 1981.
- [11] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE TPAMI*, 16(9):920–932, September 1994.
- [12] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321–331, January 1988.
- [13] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? In ECCV'02, pages Part III, 65–81.
- [14] M. H. Lin and C. Tomasi. Surfaces with occlusions from layered stereo. In CVPR'03, pages 710–717, June 2003.
- [15] Y. Nakamura, T. Matsuura, K. Satoh, and Y. Ohta. Occlusion detectable stereo - occlusion patterns in camera matrix. In CVPR'96, pages 371–378.
- [16] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065– 1076, 1962.
- [17] A. Rényi. On measures of entropy and information. In Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, volume 1, pages 547–561. University of California Press, 1961.
- [18] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1):7–42, May 2002.
- [19] G.L. Scott and H.C. Longuet-Higgins. An algorithm for associating the features of two images. *Proceedings: Biological Sciences*, 244(1309):21–26, April 1991.
- [20] J. Shade, S. Gortler, L.-W. He, and R. Szeliski. Layered depth images. In SIGGRAPH'98, pages 231–242.
- [21] J. Sun, H.-Y. Shum, and N.-N. Zheng. Stereo matching using belief propagation. In *ECCV'02*, pages Part II, 510–524.
- [22] R. Szeliski and J. Coughlan. Hierarchical spline-based image registration. In CVPR'94, pages 194–201.
- [23] R. Szeliski and H.-Y. Shum. Motion estimation with quadtree splines. *IEEE TPAMI*, 18(12):1199–1210, December 1996.
- [24] R. Szeliski and D. Scharstein. Symmetric sub-pixel stereo matching. ECCV'02, pages Part II, 525–540.
- [25] H. Tao, H.S. Sawhney, and R. Kumar. A global matching framework for stereo computation. In *ICCV 2001*, volume I, pages 532–539, July 2001.
- [26] C. J. Taylor, P. E. Debevec, and J. Malik. Reconstructing polyhedral models of architectural scenes from photographs. In *ECCV'96*, volume 2, pages 659–668.
- [27] Y. Tsin. Kernel correlation as an affinity measure in pointsampled vision problems. *Technical Report*, CMU-RI-03-36, 2003.
- [28] S. Ullman. *The Interpretation of Visual Motion*. MIT Press, Cambridge, Massachusetts, 1979.