# Stereo Matching Using Iterated Graph Cuts and Mean Shift Filtering

Ju Yong Chang, Kyoung Mu Lee, and Sang Uk Lee

School of Electrical Eng., ASRI,
Seoul National University, 151-600, Seoul, Korea
jangbon@diehard.snu.ac.kr, kyoungmu@snu.ac.kr, sanguk@sting.snu.ac.kr

**Abstract.** In this paper, we propose a new stereo matching algorithm using an iterated graph cuts and mean shift filtering technique. Our algorithm consists of following two steps. In the first step, given an estimated sparse RDM (Reliable Disparity Map), we obtain an updated dense disparity map through a new constrained energy minimization framework that can cope with occlusion. The graph cuts technique is employed for the solution of the proposed stereo model. In the second step, we re-estimate the RDM from the disparity map obtained in the first step. In order to obtain accurate reliable disparities, the crosschecking technique followed by the mean shift filtering in the color-disparity space is introduced. The proposed algorithm expands the RDM repeatedly through the above two steps until it converges. Experimental results on the standard data set demonstrate that the proposed algorithm achieves comparable performance to the state-of-the-arts, and gives good results especially in the areas such as the disparity discontinuous boundaries and occluded regions, where the conventional methods usually suffer.

## 1 Introduction

Stereo matching is one of the classical problems in computer vision and has many potential application areas including the robot navigation, 3D modelling, and image based rendering. In the stereo matching problem, we are given more than two images of the same scene. Then the goal of stereo matching is to compute the disparity map for the reference image. A disparity describes the difference in the positions of two corresponding pixels. Therefore, to get the disparity map, we have to solve the correspondence problem for each pixel. Generally, in binocular stereo, we assume that two input images are calibrated and rectified in advance, so that the epipolar line becomes horizontal. However despite those constraints, due to the ill-posed nature of the stereo matching problem, determination of accurate disparities is still a hard problem, especially in the occluded and textureless areas. To resolve this problem, many stereo matching algorithms have been proposed, and a detailed review of those algorithms can be found in [1].

In general, stereo algorithms can be classified into the local or global approaches. Local algorithms often use a finite-size window to increase the discrimination power of correspondence. Corresponding points can be found by

P.J. Narayanan et al. (Eds.): ACCV 2006, LNCS 3851, pp. 31–40, 2006.

comparing the intensity values of the local windows with various matching metrics like SSD, SAD, NCC, and Birchfield measure [2]. Local algorithms are very efficient, but they are sensitive to locally ambiguous regions (e.g., occlusion regions or regions with uniform texture) and disparity discontinuous boundaries.

Global algorithms use the smoothness constraint in order to resolve the ill-posed problem of stereo matching. By using this, the problem of textureless regions can be handled successfully. However, the discontinuous features of the disparity map usually cannot be recovered by the simple linear or quadratic smoothness constraint. Thus, the discontinuity preserving smoothness constraint such as Potts model has been employed for the stereo model, and the energy function including such a smoothness constraint is minimized through various minimization techniques. Among them, graph cuts [3, 4] and belief propagation [5] have attracted much attention due to their excellent performances. Nevertheless, since many global stereo algorithms still do not consider the occlusion problem explicitly, eventually reconstruction errors dominate in the occluded regions.

Recently, stereo matching algorithms using color segmentation has received a lot of attention [6, 7, 8, 9]. These algorithms are based on the assumption that there are no large disparity discontinuities inside homogeneous color segments. In general, we can get much sharper intensity boundaries by using color segmentation. Therefore color segmentation based stereo matching algorithms produce better performance on disparity boundaries. Tao et al. [6], Ernst et al. [7], and Hong and Chen [8] made an assumption that pixels inside each color segment produced by a color segmentation algorithm have the same disparity value. Under this assumption, the stereo matching problem can be formulated as an energy minimization problem in the segment domain instead of the pixel domain. Specifically, the energy function contains two parts; the data energy term and the smoothness term. The data energy measures the disagreement of corresponding segments given disparity value. The smoothness energy measures how smooth the disparities of neighboring segments are. In order to minimize both the data energy and the smoothness energy, Tao et al. [6] used a local greedy search algorithm, Ernst et al. [7] used the relaxation algorithm, and Hong and Chen [8] used the graph cuts technique. However these methods depend largely on the initial color segmentation result. Consequently, these methods usually get in trouble when there exist disparity boundaries inside the initial color segments.

In this paper, we present a new segmentation-based stereo matching algorithm using an iterated graph cuts and mean shift filtering technique. In contrast to most conventional segmentation based stereo matching methods that exploit only the color segmentation information or the disparity segmentation information independently, our proposed method considers the segmentation using both the color and disparity information simultaneously in the color-disparity space. Through the mean shift filtering [10] in the color-disparity space, the proposed method corrects the current disparity map coherently with the disparity distribution information as well as the color information. In order to reduce the effect of outliers and to obtain more reliable disparities, the disparity crosschecking

(left-right checking) is performed before the mean-shift filtering. Thus, through the crosschecking and mean shift filtering, we obtain a RDM (Reliable Disparity Map) from the current disparity map, that is sparse but contains reliable disparities (of ground control points). Such a RDM is then used to guide more correct and dense disparity map through a constrained energy minimization framework that can handle the occlusion. The reliable disparity constrained energy minimization is solved via graph cuts, and it makes the proposed algorithm more robust to the occlusion problem.

The rest of the paper is organized as follows. First we present the constrained stereo matching method by the reliable disparities in Section 2. Then we explain how to compute the RDM through the crosschecking and mean shift filtering procedures in Section 3. And we describe the structure of the overall algorithm in Section 4. Experimental results on various data sets are shown in section 5, and finally, conclusions are drawn in Section 6.

## 2 Stereo Matching with the RDM

In this section, we present the first part of the proposed algorithm, that is, the stereo matching with the RDM. Firstly, we introduce the conventional energy-based stereo model. Then, we explain how to formulate and solve the constrained stereo model with a given RDM.

### 2.1 Energy-Based Stereo Matching Model

Let $L$ and $R$ be the sets of pixels in the left and right images, respectively. The goal of stereo matching is to determine a label $f_p$ for each pixel $p$ in the left image, which denotes a disparity value for that pixel. Then, the stereo matching can be modelled as the following energy minimization problem,

$$E(f) = E_{data}(f) + E_{smooth}(f). \tag{1}$$

The data term, $E_{data}(f)$, measures how consistent the disparity function $f$ agrees with the input images, and can be written as

$$E_{data}(f) = \sum_{p \in L} D_p(f_p), \tag{2}$$

where $D_p(f_p)$ is a penalty function of the pixel $p$ having the disparity $f_p$. This penalty function can be the usual SSD, SAD or normalized correlation. However, in this paper, we use the pixel dissimilarity measure proposed in [2], since it is known to be insensitive to the image sampling noise. The smoothness term, $E_{smooth}(f)$, encodes the smoothness assumption imposed by the algorithm, and can be written as

$$E_{smooth}(f) = \sum_{p,q \in N} V_{p,q} \cdot T(f_p \neq f_q), \tag{3}$$

where $N$ is a neighborhood system for the pixels of the left image, $V_{p,q}$ is a function to control the level of smoothness, and $T(\cdot)$ is 1 if its argument is true and 0 otherwise. This is called the Potts energy model, and we adopt this smoothness model for its discontinuity preserving feature.

### 2.2 A New Modified Stereo Model

By employing the Potts energy model for the smoothness constraint, we can remedy the problems of disparity discontinuous boundaries as well as the textureless regions. However, the conventional energy-based stereo models still lack proper consideration of the occlusion problem. A simple example is shown in figure 1. The arrows indicate the true correspondences between pixels in two images. The true disparity value of white pixels is 0, and that of gray pixels is 1. According to the conventional stereo model, the data term for these true correspondences becomes $E_{data}(f) = D_p(0) + D_q(0) + D_r(1) + D_s(1)$. However, note that the pixel $q$ is occluded by the pixel $r$, and true corresponding pixel does not exist in the right image. Thus, minimizing the penalty term of the occluded pixel $q$, $D_q$ is meaningless, and produces false matching.

Therefore, in order to make the penalty term of each pixel in the left image contribute to the data term properly, we have to check the visibility of each pixel in the right image. For that purpose, we introduce a function $Vis_p$ that indicates whether the occlusion is occurred or not for pixel $p$. When the pixel $p$ is occluded, $Vis_p$ is 0, otherwise, $Vis_p$ is 1. Note that, in general, the occlusion of a pixel $p$ depends not only $f_p$, the disparity at $p$, but also the disparities of the neighboring pixels that can occlude it. So, $Vis_p$ should be a function of $f_p$ and $f$. Now, the data term modified by the visibility function $Vis_p$ can be written by

$$E'_{data}(f) = \sum_{p \in L} Vis_p(f_p, f) \cdot D_p(f_p). \tag{4}$$

Because of the dependency of the visibility function $Vis_p$ on $f$, minimizing the total energy function $E(f)$ becomes a nontrivial problem. Actually, we can

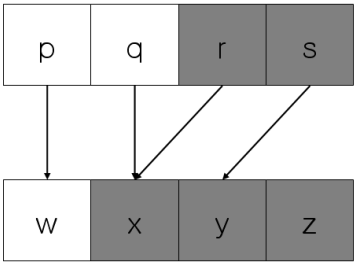


**Fig. 1.** An example of stereo matching with occlusion: $L = \{p, q, r, s\}$ and $R = \{w, x, y, z\}$. Arrows indicate the true correspondences between pixels in two images.

prove that the new energy function does not satisfy the regularity condition [11]. And, according to [11], the regularity condition is a necessary and sufficient condition for minimizing the energy function via graph cuts. Thus, the energy function involving the modified data term in (4) can not be solved by the graph cuts directly.

In order to minimize the modified energy function via graph cuts efficiently, we introduce a RDM, $r$ in which each element $r_p$ can have a label of reliable disparity value, or the $UD$ label for the undetermined or invalid disparity. Thus, the RDM provides the information on each pixel whether its disparity has been already determined reliably or to be further estimated. By using a given RDM, we can modify (4) by

$$E''_{data}(f) = \sum_{p \in L} D'_p(f_p), \tag{5}$$

$$D'_p(f_p) = \begin{cases} Vis_p(f_p, r) \cdot D_p(f_p), & \text{if } r_p = UD; \\ 0, & \text{if } r_p \neq UD \text{ and } f_p = r_p; \\ \infty, & \text{if } r_p \neq UD \text{ and } f_p \neq r_p, \end{cases} \tag{6}$$

where $D'_p(f_p)$ is a modified data penalty term by the RDM constraint. For the pixels that have reliable disparities ($r_p \neq UD$), we do not change the current disparity values. While, for the pixels that need new disparity estimation ($r_p = UD$), by using the visibility function $Vis_p(f_p, r)$ constrained by the reliable disparities, we can eliminate the effect of the occluded pixels efficiently. Thus, by employing this new data term, we can resolve the occlusion problem effectively.

Now, the proposed energy function consists of the modified data term with given $r$ in (5) and (6), and the traditional Potts energy model in (3) given by

$$E(f) = E''_{data}(f) + E_{smooth}(f). \tag{7}$$

Note that the modified data term is a summation of the new penalty terms that depend only on the disparity $f_p$ of each pixel $p$, and has the same formation as the conventional data term in equation (2). Therefore the proposed energy function can be minimized via graph cuts. In this paper, we use the $\alpha$-expansion algorithm [3].

## 3 Computing the RDM

In this section, we explain the second part of the proposed algorithm, that is, how to construct the RDM from the disparity map estimated in the first step. The RDM consists of pixels with reliable disparity values and pixels of which disparity values are invalid. For estimating whether given disparity values are reliable or not, we use the conventional crosschecking technique followed by clustering in the color-disparity space. Through the crosschecking of left and right disparity maps, only the disparity values that are consistent in both maps are survived as the reliable disparities, and the others are assigned by $UD$ label that means undetermined disparity. Next, as in many other works [6, 7, 8, 9] that successfully

applied the color segmentation information to stereo matching, we also use the color information to refine and correct the crosschecked disparity map. We adopt the mean shift algorithm [10] for this purpose.

### 3.1 Crosschecking Technique

Let $f_p$ and $f'_{p'}$ be the disparity values of the corresponding pixels $p$ and $p'$ in the left and right images, respectively. Then, if $f_p = f'_{p'}$, we consider the disparity $f_p$ at $p$ as a reliable disparity value, otherwise an invalid one.

### 3.2 Mean Shift Algorithm

The mean shift algorithm is a nonparametric density estimation-based method for feature space analysis, proposed by Comaniciu and Meer [10]. It assumes that the feature space can be regarded as an empirical probability density function (p.d.f) of the represented parameter. Dense regions in the feature space correspond to local maxima of the p.d.f., that is, the modes of the unknown density. Once the location of a mode is determined, the cluster associated with it can be delineated based on the local structure of the feature space. Thus, the mode detection is an important part for the feature space analysis. In the mean shift algorithm, such a mode detection process is based on the mean shift procedure.

According to the work of Comaniciu and Meer [10], the mean shift procedure that is the successive iteration of the following two steps;

- computation of the mean shift vector $m_{h,G}(x)$,
- translation of the kernel $G(x)$ by $m_{h,G}(x)$,

is guaranteed to converge at a nearby point where the density estimator has zero gradient, that is, a mode. Here, $m_{h,G}(x)$ is the mean shift vector defined by

$$m_{h,G}(x) = \frac{\sum_{i=1}^{n} x_i g(\|\frac{x-x_i}{h}\|^2)}{\sum_{i=1}^{n} g(\|\frac{x-x_i}{h}\|^2)} - x, \tag{8}$$

where $x_i$, $i = 1, ..., n$, are data points, and the function $g(x)$ is the profile of the kernel. The set of all locations that converge to the same mode defines the basin of attraction of that mode. Thus, the delineation of the clusters is a natural outcome of the above mode detection process. After convergence, the basin of attraction of a mode, i.e., the data points visited by all the mean shift procedures converging to that mode, automatically delineates a cluster of arbitrary shape.

### 3.3 Mean Shift Filtering in Color-Disparity Space

Most conventional color segmentation based stereo matching algorithms use the segmentation information in the color space. However, the color segmentation algorithm cannot produce correct scene segmentation results, because it does not consider the disparity (or depth) information. Therefore, in this paper, we

incorporate the disparity information with the color and spatial coordinates information through the mean shift algorithm. We compose the CSD (Color-Spatial coordinates-Disparity) space by adding the disparity component $x^d$ to the conventional color-spatial coordinates space. We use the crosschecked disparity as the disparity component $x_p^d$ at the pixel $p$. Then, the feature vector of each pixel $p$ in the input image can be represented by $x_p = (x_p^c, x_p^s, x_p^d)$, a point in the 6-D CSD space. In this expression, $x^c$ and $x^s$ are the color and spatial coordinates part of the feature vector, respectively. We apply the mean shift procedure to such feature points in the CSD space repeatedly until it converges, and replace the disparity value of each pixel by that of the corresponding point of convergence. The mean shift filtering algorithm in the CSD space can be summarized as follows. Let $x_p$ and $z_p$, $p = 1, ..., n$, be the input and filtered feature vector of a pixel $p$ in the CSD domain, respectively. For each pixel,

1. Initialize $i = 1$ and $y_{p,1} = x_p$.
2. Compute $y_{p,i+1}$ according to $y_{p,i+1} = y_{p,i} + m_{h,G}(y_{p,i})$ until convergence. $m_{h,G}(y_{p,i})$ is the mean shift vector at the point $y_{p,i}$. Let $y_{p,c}$ be the converging point.
3. Assign $z_p = (x_p^c, x_p^s, y_{p,c}^d)$.

After convergence, we define a reliable disparity map $r$ as $r_p = z_p^d$.

In order to perform the above mean shift clustering algorithm, we have to compute the mean shift vector $m_{h,G}(x)$. However, because of the characteristic of the disparity space different from the color and spatial coordinates space, we have to set a new definition of the mean shift vector.

**Distance in Disparity Space.** Let $x$ be a point in the CSD space. Then, the mean shift vector at the point $x$ can be computed by (8). In order to compute the mean shift vector, we have to compute the distance between the point $x$ and the data points in the input image. We can compute the distance by the sum of the distances of each component normalized by the bandwidth in its domain. For the color and spatial component, we use the Euclidean distance. However, it is not appropriate for the disparity space, since we assume the piecewise constant constraint among local disparities by Potts model as in (3). Moreover, the $UD$ label makes the Euclidean distance unusable. By the piecewise constant assumption, we enforce the same cost for the neighboring pixels with unequal disparities, regardless of the magnitude of the disparity difference. Thus, following this notion, let us define the distance in the disparity domain as follows:

$$\| \frac{x^d - x_i^d}{h_d} \| = \begin{cases} 0, \text{ if } x^d = x_i^d; \\ k, \text{ otherwise,} \end{cases} \tag{9}$$

where $k$ is some constant.

**Mean Shift Vector in Disparity Space.** In the mean shift procedure, the position of the kernel is translated by the mean shift vector. The mean shift vector (8) implies the difference between the weighted means with the weighting

kernel $G$. Therefore, by the mean shift procedure, the kernel is moved to the mean of data points that belong to the kernel $G$. However, for the disparity space, the arithmetic mean of disparity values is meaningless. Therefore, instead of the arithmetic mean, we define the mean value in the disparity domain as the most frequent disparity value (mode) among disparity values of points in the kernel:

$$m_{h,G}(x)^d = \arg\max_{j \in D} \sum_{x_i^d = j} g(\| \frac{x - x_i}{h} \|^2). \tag{10}$$

In this equation, $D$ is the set of all possible disparity values.

## 4 Experimental Results

For the quantitative evaluation and comparison of different stereo algorithms, Scharstein and Szeliski [1] have proposed a test bed along with ground truths which is available at their website (http://www.middlebury.edu/stereo). We have evaluated the proposed algorithm on these test data sets. The evaluation metric is the percentage of bad pixels, of which disparity are different from the true values more than 1 pixel. This measure is calculated in three different parts of an input image including the entire image (all), untextured (untex), and discontinuity (disc) regions. And, only non-occluded pixels are considered in all three cases.

Our algorithm has four parameters; one parameter that controls the level of smoothness $V_{p,q}$ in the stereo matching part, and three parameters, $h_c$, $h_s$, and $k$ for the mean shift filtering part. In this paper, following other researchers' works [1, 3], we employed the gradient-dependent smoothness cost for the smoothness control, given by

$$V_{p,q} = \begin{cases} 2\lambda, \text{ if } |I_p - I_q| \leq 5; \\ \lambda, \quad \text{otherwise,} \end{cases} \tag{11}$$

where $I_p$ and $I_q$ are intensity values of pixel $p$ and $q$, respectively.

All the parameters were fixed for all the test sets, and the best results were obtained when $\lambda = 10$, $h_c = 6.5$, $h_s = 7$, and $k = 0.7$. The proposed algorithm has been implemented on a Pentium IV 3.0GHz PC. Typically, after few iterations, the RDM converged, and the final dense disparity map was computed within few minutes (e.g. Tsukuba data, 3 iterations, 95 seconds).

Figure 2 reports the detailed intermediate results on the Tsukuba data. We can see that through the crosschecking and mean shift filtering process, reliable disparities coherent with color information have been extracted from the given disparity map. And through the updating stereo matching process guided by those ground control points with reliable disparities, more undetermined pixels become fixed and the reliable disparity range expands.

Table 1 presents the overall performance of our algorithm, where it summarizes the quantitative evaluation results. The proposed algorithm performs quite well, and our overall rank is $4^{th}$ out of about 30 algorithms. From the extracted disparity maps, we can observe that especially good performances have been

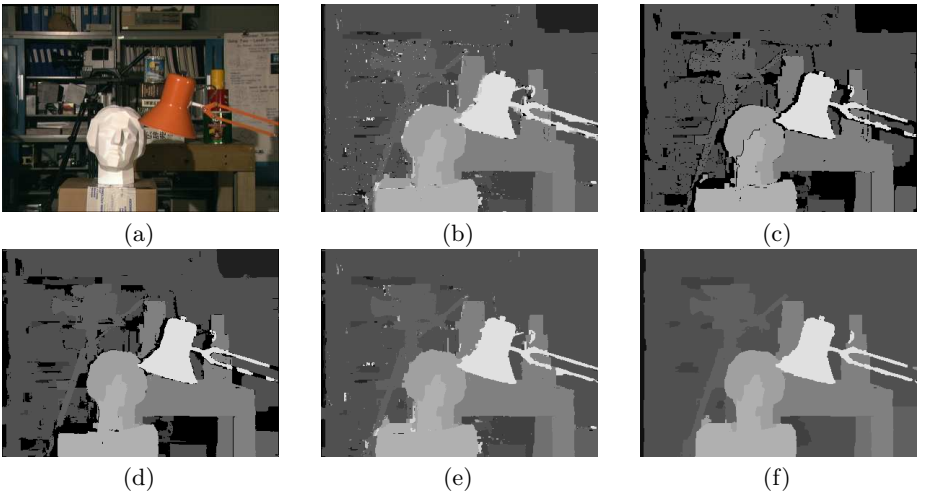

**Fig. 2.** Detailed results on the Tsukuba data. (a) Reference image. (b) disparity map in the first iteration, (c) RDM after crosschecking, (d) RDM after mean shift filtering, (e) disparity map in the second iteration, (f) final disparity map after convergence.

**Table 1.** Evaluation table of different stereo algorithms

| Algorithms | Tsukuba | | | Sawtooth | | | Venus | | | Map | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | all | untex | disc | all | untex | disc | all | untex | disc | all | disc |
| Sym.BP+occl. | 0.97 | 0.28 | 5.45 | 0.19 | 0.00 | 2.09 | 0.16 | 0.02 | 2.77 | 0.16 | 2.20 |
| Segm.-based GC [8] | 1.23 | 0.29 | 6.94 | 0.30 | 0.00 | 3.24 | 0.08 | 0.01 | 1.39 | 1.49 | 15.46 |
| Graph+segm. | 1.39 | 0.28 | 7.17 | 0.25 | 0.00 | 2.56 | 0.11 | 0.02 | 2.04 | 2.35 | 20.87 |
| **Our method** | 1.13 | 0.48 | 6.38 | 1.14 | 0.06 | 3.34 | 0.77 | 0.70 | 3.61 | 0.95 | 12.83 |
| Segm.+glob.vis. | 1.30 | 0.48 | 7.50 | 0.20 | 0.00 | 2.30 | 0.79 | 0.81 | 6.37 | 1.63 | 16.07 |
| Layered | 1.58 | 1.06 | 8.82 | 0.34 | 0.00 | 3.35 | 1.52 | 2.96 | 2.62 | 0.37 | 5.24 |
| Belief prop. [5] | 1.15 | 0.42 | 6.31 | 0.98 | 0.30 | 4.83 | 1.00 | 0.76 | 9.13 | 0.84 | 5.27 |
| MultiCam GC | 1.85 | 1.94 | 6.99 | 0.62 | 0.00 | 6.86 | 1.21 | 1.96 | 5.71 | 0.31 | 4.34 |
| 2-pass DP | 1.53 | 0.66 | 8.25 | 0.61 | 0.02 | 5.25 | 0.94 | 0.95 | 5.72 | 0.70 | 9.32 |
| GC+occl. | 1.19 | 0.23 | 6.71 | 0.73 | 0.11 | 5.71 | 1.64 | 2.75 | 5.41 | 0.61 | 6.05 |

achieved in the areas such as disparity discontinuous boundaries and occluded regions, where the conventional stereo algorithms usually suffer.

## 5 Conclusion

In this paper, we presented a new stereo matching algorithm based on iterated constrained graph cuts with reliable disparities obtained by the mean shift filtering in the CSD space. Through the mean shift filtering in the CSD space, a RDM coherent with disparity information as well as color information is ob-

tained. And computing the solution of a new constrained stereo energy model with given RDM enables the proposed algorithm to be more robust to the occlusion. Evaluation and comparison result shows that our algorithm is one of the state-of-the-arts.

## Acknowledgements

This work has been supported in part by the ITRC (Information Technology Research Center) support program of Korean government and IIRC (Image Information Research Center) by Agency of Defense Development, Korea.

## References

1. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. IJCV **47** (2002) 7–42
2. Birchfield, S., Tomasi, C.: A pixel dissimilarity measure that is insensitive to image sampling. PAMI **20** (1998) 401–406
3. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. PAMI **23** (2001) 1222–1239
4. Kolmogorov, V., Zabih, R.: Computing visual correspondence with occlusions using graph cuts. In: ICCV01. (2001) 508–515
5. Sun, J., Zheng, N.N., Shum, H.Y.: Stereo matching using belief propagation. PAMI **25** (2003) 787–800
6. Tao, H., Sawhney, H.: A global matching framework for stereo computation. In: ICCV01. (2001) I: 532–539
7. Ernst, F., Wilinski, P., Overveld, K.V.: Dense structure-from-motion: An approach based on segment matching. In: ECCV02. (2002) II: 217–231
8. Hong, L., Chen, G.: Segment-based stereo matching using graph cuts. In: CVPR04. (2004) I: 74–81
9. Wei, Y., Quan, L.: Region-based progressive stereo matching. In: CVPR04. (2004) I: 106–113
10. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. PAMI **24** (2002) 1–18
11. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts. PAMI **26** (2004) 147–159