Stereo Matching: An Outlier Confidence Approach

Li Xu and Jiaya Jia

Department of Computer Science and Engineering
The Chinese University of Hong Kong
{xuli,leojia}@cse.cuhk.edu.hk

Abstract. One of the major challenges in stereo matching is to handle partial occlusions. In this paper, we introduce the Outlier Confidence (OC) which dynamically measures how likely one pixel is occluded. Then the occlusion information is softly incorporated into our model. A global optimization is applied to robustly estimating the disparities for both the occluded and non-occluded pixels. Compared to color segmentation with plane fitting which globally partitions the image, our OC model locally infers the possible disparity values for the outlier pixels using a reliable color sample refinement scheme. Experiments on the Middlebury dataset show that the proposed two-frame stereo matching method performs satisfactorily on the stereo images.

1 Introduction

One useful technique to reduce the matching ambiguity for stereo images is to incorporate the color segmentation into optimization [1,2,3,4,5,6]. Global segmentations improve the disparity estimation in textureless regions; but most of them do not necessarily preserve accurate boundaries. We have experimented that, when taking the ground truth occlusion information into optimization, very accurate disparity estimation can be achieved. This shows that partial occlusion is one major source of matching errors. The main challenge of solving the stereo problems now is the appropriate outlier detection and handling.

In this paper, we propose a new stereo matching algorithm aiming to improve the disparity estimation. Our algorithm does not assign each pixel a binary visibility value indicating whether this pixel is partially occluded or not [7,4,8], but rather introduces soft Outlier Confidence (OC) values to reflect how confident we regard one pixel as an outlier. The OC values, in our method, are used as weights balancing two ways to infer the disparities. The final energy function is globally optimized using Belief Propagation (BP). Without directly labeling each pixel as “occlusion” or “non-occlusion”, our model has considerable tolerance of errors produced in the occlusion detection process.

Another main contribution of our algorithm is the local disparity inference for outlier pixels, complementary to the global segmentation. Our method defines the disparity similarity according to the color distance between pixels and naturally transforms color sample selection to a general foreground or background color inference problem using image matting. It effectively reduces errors caused by inaccurate global color segmentation and gives rise to a reliable inference of the unknown disparity of the occluded pixels.

D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part IV, LNCS 5305, pp. 775–787, 2008.

We also enforce the inter-frame disparity consistency and use BP to simultaneously estimate the disparities of two views. Experimental results on the Middlebury dataset [9] show that our OC model effectively reduces the erroneous disparity estimate due to outliers.

2 Related Work

A comprehensive survey of the dense two-frame stereo matching algorithms was given in [10]. Evaluations of almost all stereo matching algorithms can be found in [9]. Here we review previous work dealing with outliers because, essentially, the difficulty of stereo matching is to handle the ambiguities.

Efforts of dealing with outliers are usually put in three stages in stereo matching – that is, the cost aggregation, the disparity optimization, and the disparity refinement. Most approaches use outlier truncation or other robust functions for cost computation in order to reduce the influence of outliers [2,11].

Window-based methods aggregate matching cost by summing the color differences over a support region. These methods [12,13] prevent depth estimation from aggregating information across different depth layers using the color information. Yoon and Kweon [14] adjusted the support-weight of a pixel in a given window based on the CIELab color similarity and its spatial distance to the center of the support window. Zitnick *et al.* [12] partitioned the input image and grouped the matching cost in each color segment. Lei *et al.* [15] used segmentation to form small regions in a region-tree for further optimization.

In disparity optimization, outliers are handled in two ways in general. One is to explicitly detect occlusions and model visibility [7,4,8]. Sun *et al.* [4] introduced the visibility constraint by penalizing the occlusions and breaking the smoothness between the occluded and non-occluded regions. In [8], Strecha *et al.* modeled the occlusion as a random outlier process and iteratively estimated the depth and visibility in an EM framework in multi-view stereo. Another kind of methods suppresses outliers using extra information, such as pixel colors, in optimization. In [16,6], a color weighted smoothness term was used to control the message passing in BP. Hirschmuller [17] took color difference as the weight to penalize large disparity differences and optimized the disparities using a semi-global approach.

Post-process was also introduced to handle the remaining outliers after the global or local optimization. Occluded pixels can be detected using a consistency check, which validates the disparity correspondences in two views [10,4,17,6]. Disparity interpolation [18] infers the disparities for the occluded pixels from the non-occluded ones by setting the disparities of the mis-matched pixels to that of the background. In [1,3,4,5,6], color segmentation was employed to partition images into segments, each of which is refined by fitting a 3D disparity plane. Optimization such as BP can be further applied after plane fitting [4,5,6] to reduce the possible errors.

Several disparity refinement schemes have been proposed for novel-view synthesis. Sub-pixel refinement [19] enhances details for synthesizing a new view. In [12] and [20], boundary matting for producing seamless view interpolation was introduced. These methods only aim to synthesize natural and seamless novel-views, and cannot be directly used in stereo matching to detect or suppress outliers.

3 Our Model

Denoting the input stereo images as I_l and I_r, and the corresponding disparity maps as $\mathcal{D}_l$ and $\mathcal{D}_r$ respectively, we define the matching energy as

$$E(\mathcal{D}_l, \mathcal{D}_r; I_l, I_r) = E_d(\mathcal{D}_l; I_l, I_r) + E_d(\mathcal{D}_r; I_l, I_r) + E_s(\mathcal{D}_l, \mathcal{D}_r), \quad (1)$$

where $E_d(\mathcal{D}_l; I_l, I_r) + E_d(\mathcal{D}_r; I_l, I_r)$ is the data term and $E_s(\mathcal{D}_l, \mathcal{D}_r)$ defines the smoothness term that is constructed on the disparity maps. In our algorithm, we not only consider the spatial smoothness within one disparity map, but also model the consistency of disparities between frames.

As the occluded pixels influence the disparity estimation, they should not be used in stereo matching. In our algorithm, we do not distinguish between occlusion and image noise, but rather treat all problematic pixels as outliers. *Outlier Confidences* (OCs) are computed on these pixels, indicating how confident we regard one pixel as an outlier. The outlier confidence maps U_l and U_r are constructed on the input image pair. The confidence $U_l(x)$ or $U_r(x)$ on pixel x is a continuous variable with value between 0 and 1. Larger value indicates higher confidence that one pixel is an outlier, and vice versa.

Our model combines an initial disparity map and an OC map for two views. In the following, we first introduce our data and smoothness terms. The construction of the OC map will be described in Section 4.2.

3.1 Data Term

In the stereo configuration, pixel x in I_l corresponds to pixel $x - d_l$ in I_r by disparity d_l. Similarly, x in I_r corresponds to $x + d_r$ in I_l. All possible disparity values for d_l and d_r are uniformly denoted as set Ψ, containing integers between 0 and N, where N is the maximum positive disparity value. The color of pixel x in I_l (or I_r) is denoted as $I_l(x)$ (or $I_r(x)$). We define the data term $E_d(\mathcal{D}_l; I_l, I_r)$ on the left image as

$$E_d(\mathcal{D}_l; I_l, I_r) = \sum_x [(1 - U_l(x))(\frac{f_0(x, d_l; I_l, I_r)}{\alpha}) + U_l(x)(\frac{f_1(x, d_l; I_l)}{\beta})], \quad (2)$$

where α and β are weights. $f_0(x, d; I_l, I_r)$ denotes the color dissimilarity cost between two views. $f_1(x, d; I_l)$ is the term defined as the local color and disparity discontinuity cost in one view. $E_d(\mathcal{D}_r; I_l, I_r)$ on the right image can be defined in a similar way.

The above two terms, balanced by the outlier confidence $U_l(x)$, model respectively two types of processes in disparity computation. Compared to setting $U_l(x)$ as a binary value and assigning pixels to either outliers or inliers, our cost terms are softly combined, tolerating possible errors in pixel classification.

For result comparison, we give two definitions of $f_0(x, d_l; I_l, I_r)$ respectively corresponding to whether the segmentation is incorporated or not. The first is to use the color and distance weighted local window [14,6,19] to aggregate color difference between conjugate pixels:

$$f_0^{(1)}(x, d_l; I_l, I_r) = \min(g(\|I_l(x) - I_r(x - d_l)\|_1), \varphi), \quad (3)$$

where $g(\cdot)$ is the aggregate function defined similarly to Equation (2) in [6]. We use the default parameter values (local window size 33×33, $\beta_{cw} = 10$ for normalizing color differences, $\gamma_{cw} = 21$ for normalizing spatial distances). φ determines the maximum cost for each pixel, whose value is set as the average intensity of pixels in the correlation volume.

The second definition is given by incorporating the segmentation information. Specifically, we use the Mean-shift color segmentation [21] with default parameters (spatial bandwidth 7, color bandwidth 6.5, minimum region size 20) to generate color segments. A plane fitting algorithm using RANSAC (similar to that in [6]) is then applied to producing the regularized disparity map d_{pf}. We define

$$f_0^{(2)}(x, d_l; I_l, I_r) = (1-\kappa) f_0^{(1)}(x, d_l) + \kappa\alpha|d - d_{pf}|, \tag{4}$$

where κ is a weight balancing two terms.

$f_1(x, d_l; I_l)$ is defined as the cost of assigning local disparity when one pixel has chance to be an outlier.

$$f_1(x, d_l; I_l) = \sum_{i \in \Psi} \omega_i(x; I_l)\delta(d_l - i), \tag{5}$$

where $\delta(\cdot)$ is the Dirac function, Ψ denotes the set of all disparity values between 0 and N and $\omega_i(x; I_l)$ is a weight function for measuring how disparity d_l is likely to be i. We omit subscript l in the following discussion of $\omega_i(x; I_l)$ since both the left and right views can use the similar definitions.

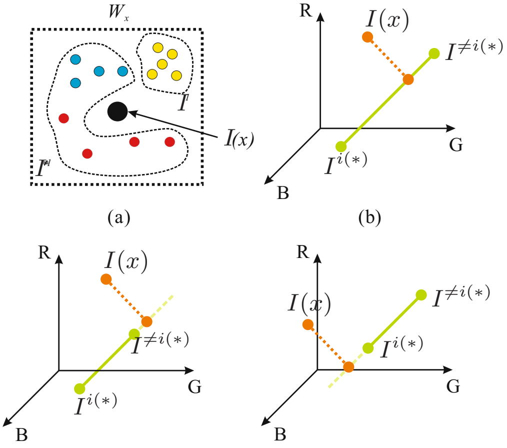
For ease of explanation, we first give a general definition of weight $\omega'_i(x; I)$, which, in the following descriptions, will be slightly modified to handle two extreme situations with values 0 and 1. We define

$$\omega'_i(x; I) = 1 - \frac{\mathcal{L}(I(x), \mathbf{I}^i(W_x))}{\mathcal{L}(I(x), \mathbf{I}^i(W_x)) + \mathcal{L}(I(x), \mathbf{I}^{\neq i}(W_x))}, \tag{6}$$

where $I(x)$ denotes the color of pixel x and W_x is a window centered at x. Suppose after initialization, we have collected a set of pixels x' detected as inliers within each W_x (i.e., $U(x') = 0$), and have computed disparities for these inliers. We denote by $\mathbf{I}^i$ the set of inliers whose disparity values are computed as i. Similarly, $\mathbf{I}^{\neq i}$ are the inliers with the corresponding disparity values not equal to i. $\mathcal{L}$ is a metric measuring the color difference between $I(x)$ and its neighboring pixels $\mathbf{I}^i(W_x)$ and $\mathbf{I}^{\neq i}(W_x)$. One example is shown in Figure 1(a) where a window W_x is centered at an outlier pixel x. Within W_x, inlier pixels are clustered into $\mathbf{I}^1$ and $\mathbf{I}^{\neq 1}$. $\omega'_1(x; I)$ is computed according to the color similarity between x and other pixels in the two clusters.

(6) is a function to assign an outlier pixel x a disparity value, constrained by the color similarity between x and the clustered neighboring pixels. By and large, if the color distance between x and its inlier neighbors with disparity i is small enough compared to the color distance to other inliers, $\omega'_i(x; I)$ should have a large value, indicating high chance to let $d_l = i$ in (5).

Now the problem is on how to compute a metric $\mathcal{L}$ that appropriately measures the color distance between pixels. In our method, we abstract color sets $\mathbf{I}^i(W_x)$ and



(a) (b) (c) (d)

Fig. 1. Computing disparity weight ω'. (a) Within a neighborhood window W_x, inlier pixels are clustered into $\mathbf{I}^1$ and $\mathbf{I}^{\neq 1}$. (b)-(d) illustrate the color projection. (b) The projection of $I(x)$ on vector $I^{i(*)} - I^{\neq i(*)}$ is between two ends. (c-d) The projections of $I(x)$ are out of range, thereby are considered as extreme situations.

$\mathbf{I}^{\neq i}(W_x)$ by two representatives $I^{i(*)}$ and $I^{\neq i(*)}$ respectively. Then $\mathcal{L}$ is simplified to a color metric between pixels. We adopt the color projection distance along vector $I^{i(*)} - I^{\neq i(*)}$ and define

$$\mathcal{L}(I(x), c) = \|\langle I(x) - c, I^{i(*)} - I^{\neq i(*)}\rangle\|, \tag{7}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product of two color vectors and c can be either $I^{i(*)}$ or $I^{\neq i(*)}$. We regard $I^{i(*)} - I^{\neq i(*)}$ as a projection vector because it measures the absolute difference between two representative colors, or, equivalently, the distance between sets $\mathbf{I}^i(W_x)$ and $\mathbf{I}^{\neq i}(W_x)$.

Projecting $I(x)$ to vector $I^{i(*)} - I^{\neq i(*)}$ also makes the assignment of two extreme values 0 and 1 to $\omega_i(x; I)$ easy. Taking Figure 1 as an example, if the projection of $I(x)$ on vector $I^{i(*)} - I^{\neq i(*)}$ is between two ends, its value is obviously between 0 and 1, as shown in Figure 1 (b). If the projection of $I(x)$ is out of one end point, its value should be 0 if it is close to $I^{i(*)}$ or 1 otherwise (Figure 1 (c) and (d)). To handle the extreme cases, we define the final $\omega_i(x; I)$ as

$$\omega_i(x; I) = \begin{cases} 0 & \text{if } \langle I - I^{\neq i(*)}, I^{i(*)} - I^{\neq i(*)}\rangle < 0 \\ 1 & \text{if } \langle I^{i(*)} - I, I^{i(*)} - I^{\neq i(*)}\rangle < 0 \\ \omega'_i(x; I) & \text{Otherwise} \end{cases}$$

which is further expressed as

$$\omega_i = \mathcal{T}\left(\frac{(I - I^{\neq i(*)})^T (I^{i(*)} - I^{\neq i(*)})}{\|I^{i(*)} - I^{\neq i(*)}\|_2^2}\right), \tag{8}$$

where

$$\mathcal{T}(x) = \begin{cases} 0 & x < 0 \\ 1 & x > 1 \\ x & \text{otherwise} \end{cases} \tag{9}$$

Note that term $\frac{(I - I^{\neq i(*)})^T (I^{i(*)} - I^{\neq i(*)})}{\|I^{i(*)} - I^{\neq i(*)}\|_2^2}$ defined in (8) is quite similar to an alpha matte model used in image matting [22,23] where the representative colors $I^{i(*)}$ and $I^{\neq i(*)}$ are analogous to the unknown foreground and background colors. The image matting problem is solved by color sample collection and optimization. In our problem, the color samples are those clustered neighboring pixels $\mathbf{I}^i(W_x)$ and $\mathbf{I}^{\neq i}(W_x)$.

With the above analysis, computing the weight ω_i is naturally transformed to an image matting problem where the representative color selection is handled by applying an optimization algorithm. In our method, we employ the robust matting with optimal color sample selection approach [23]. In principle, $I^{i(*)}$ and $I^{\neq i(*)}$ are respectively selected from $\mathbf{I}^i(W_x)$ and $\mathbf{I}^{\neq i}(W_x)$ based on a *sample confidence* measure combining two criteria. First, either $I^{i(*)}$ or $I^{\neq i(*)}$ should be similar to the color of the outlier pixel I, which makes weight ω_i approach either 0 or 1 and the weight distribution hardly uniform. Second, I is also expected to be a linear combination of $I^{i(*)}$ and $I^{\neq i(*)}$. This is useful for modeling color blending since outlier pixels have chance to be the interpolation of color samples, especially for those on the region boundary.

Using the sample confidence definition, we get two weights and a neighborhood term, similar to those in [23]. Then we apply the Random Walk method [24] to compute weight ω_i. This process is repeated for all ω_i's, where $i = 0, \cdots, N$. The main benefit that we employ this matting method is that it provides an optimal way to select representative colors while maintaining spatial smoothness.

3.2 Smoothness Term

Term $E_s(\mathcal{D}_l, \mathcal{D}_r)$ contains two parts, representing intra-frame disparity smoothness and inter-frame disparity consistency:

$$\begin{aligned} E_s(\mathcal{D}_l, \mathcal{D}_r) = \sum_x [& \sum_{x' \in \mathcal{N}_1(x)} (\frac{f_3(x, x', d_l, d_r)}{\lambda}) + \sum_{x' \in \mathcal{N}_2(x)} (\frac{f_2(x, x', d_l)}{\gamma}) + \\ & \sum_{x' \in \mathcal{N}_1(x)} (\frac{f_3(x, x', d_r, d_l)}{\lambda}) + \sum_{x' \in \mathcal{N}_2(x)} (\frac{f_2(x, x', d_r)}{\gamma})], \end{aligned} \tag{10}$$

where $\mathcal{N}_1(x)$ represents the N possible corresponding pixels of x in the other view and $\mathcal{N}_2(x)$ denotes the 4-neighborhood of x in the image space. f_2 is defined as

$$f_2(x, x', d_i) = \min(|d_i(x) - d_i(x'))|, \tau), \quad i \in \{l, r\}, \tag{11}$$

where τ is a threshold set as 2. To define (11), we have also experimented with using color weighted smoothness and observed that the results are not improved.

We define $f_3(\cdot)$ as the disparity correlations between two views:

$$\begin{aligned} & f_3(x, x', d_l, d_r) = \min(|d_l(x) - d_r(x')|, \zeta) \text{ and} \\ & f_3(x, x', d_r, d_l) = \min(|d_r(x) - d_l(x')|, \zeta) \quad , \end{aligned} \tag{12}$$

where ζ is a truncation threshold with value 1. We do not define a unique x' corresponding to x because x' is unknown in the beginning. The other reason is that both f_2 and f_3 are the costs for disparity smoothness. In f_2, all neighboring pixels are encoded in $\mathcal{N}_2$ though $d_i(x)$ is not necessarily similar to all $d_i(x')$. So we introduce f_3 with the similar thought for reducing the disparity noise in global optimization considering the inter-frame consistency.

4 Implementation

The overview of our framework is given in Algorithm 1, which consists of an initialization step and a global optimization step. In the first step, we initialize the disparity maps by minimizing an energy with the simplified data and smoothness terms. Then we compute the Outlier Confidence (OC) maps. In the second step, we globally refine the disparities by incorporating the OC maps.

Algorithm 1. Overview of our approach

```
1. Initialization:
   1.1 Initialize disparity map $\mathcal{D}$ by setting $U = 0$ for all pixels.
   1.2 Estimate Outlier Confidence map $U$.

2. Global Optimization:
   2.1 Compute data terms using the estimated outlier confidence maps.
   2.2 Global optimization using BP.
```

4.1 Disparity Initialization

To initialize disparities, we simply set all values in U_l and U_r to zeros and optimize the objective function combining (2) and (10):

$$(\sum_x \frac{f_0(x, d_l) + f_0(x, d_r)}{\alpha}) + E_s(\mathcal{D}_l, \mathcal{D}_r). \tag{13}$$

Because of introducing the inter-frame disparity consistency in (12), our Markov Random Field (MRF) based on the defined energy is slightly different from the regular-grid MRFs proposed in other stereo approaches [2,25]. In our two-frame configuration, the MRF is built on two images with $(4 + N)$ neighboring sites for each node. N is the total number of the disparity levels. One illustration is given in Figure 2 where a pixel x in I_l not only connects to its 4 neighbors in the image space, but also connects to all possible corresponding pixels in I_r.

We minimize the energy defined in (13) using Belief Propagation. The inter-frame consistency constraint makes the estimated disparity maps contain less noise in two frames. We show in Figure 3(a) the initialized disparity result using the standard 4-connected MRF without defining f_3 in (10). (b) shows the result using our $(4 + N)$-connected MRF. The background disparity noise is reduced.

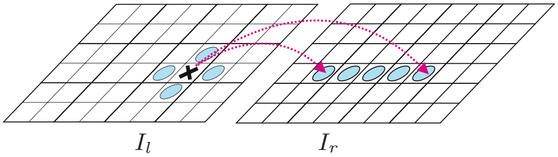


Fig. 2. In our dual view configuration, x (marked with the cross) is not only connected to 4 neighbors in one image, but also related to N possible corresponding pixels in the other image. The total number of neighbors of x is $4 + N$.

Depending on using $f_0^{(1)}$ in (3) or $f_0^{(2)}$ in (4) in the data term definition, we obtain two sets of initializations using and without using global color segmentation. We shall compare in the results how applying our OC models in the following global optimization improves both of the disparity maps.

4.2 Outlier Confidence Estimation

We estimate the outlier confidence map U on the initial disparity maps. Our following discussion focuses on estimating U_l on the left view. The right view can be handled in a similar way. The outlier confidences, in our algorithm, are defined as

$$U_l(x) = \begin{cases} 1 & |d_l(x) - d_r(x - d_l(x))| \geq 1 \\ \mathcal{T}(\frac{b_x(d^*) - b_{min}}{\|b_o - b_{min}\|}) & b_x(d^*) > t \wedge |d_l(x) - d_r(x - d_l(x))| = 0 \\ 0 & \text{Otherwise} \end{cases} \tag{14}$$

considering 2 cases.

Case 1: Our MRF enforces the disparity consistency between two views. After disparity initialization, the remaining pixels with inconsistent disparities are likely to be occlusions. So we first set the outlier confidence $U_l(x) = 1$ for pixel x if the inter-frame consistency is violated, i.e., $|d_l(x) - d_r(x - d_l(x))| \geq 1$.

Case 2: Besides the disparity inconsistency, pixel matching with large matching cost is also unreliable. In our method, since we use BP to initialize the disparity maps, the matching cost is embedded in the output disparity belief $b_x(d)$ for each pixel x. Here, we introduce some simple operations to manipulate it. First, we extract $b_x(d^*)$, i.e., the smallest belief, for each pixel x. If $b_x(d^*) < t$, where t is a threshold, the pixel should be regarded as an inlier given the small matching cost. Second, a variable b_o is computed as the average of the minimal beliefs regarding all occluded pixels detected in Case 1, i.e., $b_o = \sum_{U_l(x)=1} b_x(d^*)/K$ where K is the total number of the occluded pixels. Finally, we compute b_{min} as the average of top $n\%$ minimal beliefs among all pixels. n is set to 10 in our experiments.

Using the computed $b_x(d^*)$, b_o, and b_{min}, we estimate $U_l(\widetilde{x})$ for pixels neither detected as occlusions nor treated as inliers by setting

$$U_l(\widetilde{x}) = \mathcal{T}\left(\frac{b_{\widetilde{x}}(d^*) - b_{min}}{\|b_o - b_{min}\|}\right), \tag{15}$$

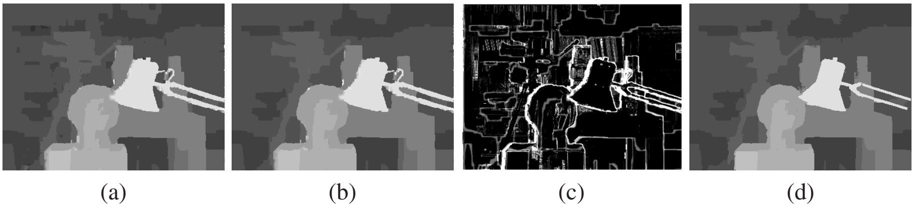


Fig. 3. Intermediate results for the "Tsukuba" example. (a) and (b) show our initial disparity maps by the 4-connected and $(4 + N)$-connected MRFs respectively without using segmentation. The disparity noise in (b) is reduced for the background. (c) Our estimated OC map. (d) A disparity map constructed by combining the inlier and outlier information. The disparities for the outlier pixels are set as the maximum weight ω_i. The inlier pixels are with initially computed disparity values.

where $\mathcal{T}$ is the function defined in (9), making the confidence value in range $[0, 1]$. (15) indicates if the smallest belief $b_x(d^*)$ of pixel x is equal to or larger than the average smallest belief of the occluded pixels detected in Case 1, the outlier confidence of x will be high, and vice versa.

Figure 3(c) shows the estimated outlier coefficient map for the "tsukuba" example. The pure black pixels represent inliers where $U_l(x) = 0$. Generally, the region consisting of pixels with $U_l(x) > 0$ is wider than the ground truth occluded region. This is allowed in our algorithm because $U_l(x)$ is only a weight balancing pixel matching and color smoothness. Even if pixel x is mistakenly labeled as an outlier, the disparity estimation in our algorithm will not be largely influenced because large $U_l(x)$ only makes the disparity estimation of x rely more on neighboring pixel information, by which $d(x)$ still has a large chance to be correctly inferred.

To illustrate the efficacy of our OC scheme, we show in Figure 3(d) a disparity map directly constructed with the following setting. Each inlier pixel is with initially computed disparity value and each outlier pixel is with the disparity i corresponding to the maximum weight ω_i among all ω_j's, where $j = 0, \cdots, N$. It can be observed that even without any further global optimization, this simple maximum-weight disparity calculation already makes the object boundary smooth and natural.

4.3 Global Optimization

With the estimated OC maps, we are ready to use global optimization to compute the final disparity maps combining costs (2) and (10) in (1). Two forms of $f_0(\cdot)$ ((3) and (4)) are independently applied in our experiments for result comparison.

The computation of $f_1(x, d; I)$ in (5) is based on the estimated OC maps and the initial disparities for the inlier pixels, which are obtained in the aforementioned steps. To compute ω_i for outlier pixel x with $U_l(x) > 0$, robust matting [23] is performed as described in Section 3.1 for each disparity level. The involved color sampling is performed in each local window with size 60×60. Finally, the smoothness terms are embedded in the message passing of BP. An acceleration using distance transform [25] is adopted to construct the messages.

5 Experiments

In experiments, we compare the results using and without using the Outlier Confidence maps. The performance is evaluated using the Middlebury dataset [10]. All parameters used in implementation are listed in Table 1 where α, β and κ are the weights defined in the data term. γ and λ are for intra-frame smoothness and inter-frame consistency respectively. φ, τ, and ζ are the truncation thresholds for different energy terms. t is the threshold for selecting possible outliers. As we normalize the messages after each message passing iteration by subtracting the mean of the messages, the belief b_{min} is negative, making $t = 0.9b_{min} > b_{min}$.

A comparison of the state-of-the-art stereo matching algorithms is shown in Table 2 extracted from the Middlebury website [9]. In the following, we give detailed explanations.

Table 1. The parameter values used in our experiments. N is the number of the disparity levels. $\overline{c}$ is the average of the correlation volume. b_{min} is introduced in (15).

Parameters	α	β	κ	γ	λ	φ	τ	ζ	t
value	φ	0.8	0.3	5.0	$5N$	$\overline{c}$	2.0	1.0	$0.9b_{min}$

Table 2. Algorithm evaluation on the Midellbury data set. Our method achieves overall rank 2 at the time of data submission.

	Avg.	Tsukuba			Venus			Teddy			Cones		
Algorithm	Rank	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
Adap.BP [5]	2.3	1.11	1.37	5.79	**0.10**	**0.21**	**1.44**	4.22	7.06	11.8	**2.48**	7.92	7.32
Our method	3.6	**0.88**	1.43	**4.74**	0.18	0.26	2.40	5.01	9.12	12.8	2.78	8.57	**6.99**
DoubleBP [6]	3.7	0.88	**1.29**	4.76	0.14	0.60	2.00	3.55	8.71	**9.70**	2.90	9.24	7.80
SPDou.BP [19]	4.6	1.24	1.76	5.98	0.12	0.46	1.74	**3.45**	8.38	10.0	2.93	8.73	7.91
SymBP+occ [4]	8.8	0.97	1.75	5.09	0.16	0.33	2.19	6.47	10.7	17.0	4.79	10.7	10.9

Table 3. Result comparison on the Middlebury dataset using (1st and 3rd rows) and without using (2nd and 4th rows) OC Maps. The segmentation information has been incorporated for the last two rows.

	Overall	Tsukuba			Venus			Teddy			Cones		
Algorithm	Rank	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
COLOR	16	1.12	3.29	5.92	0.49	1.48	6.78	10.5	16.9	21.1	3.42	12.1	8.26
COLOR+OC	5	**0.83**	1.41	**4.45**	0.25	0.31	3.22	10.1	14.6	19.9	3.22	9.82	7.40
SEG	4	0.97	1.75	5.23	0.30	0.70	3.98	5.56	9.99	13.6	3.04	8.90	7.60
SEG+OC	2	**0.88**	1.43	**4.74**	0.18	0.26	2.40	5.01	9.12	12.8	2.78	8.57	**6.99**

5.1 Results without Using Segmentation

In the first part of our experiments, we do not use the segmentation information. So data term $f_0^{(1)}$ defined in (3) is used in our depth estimation.

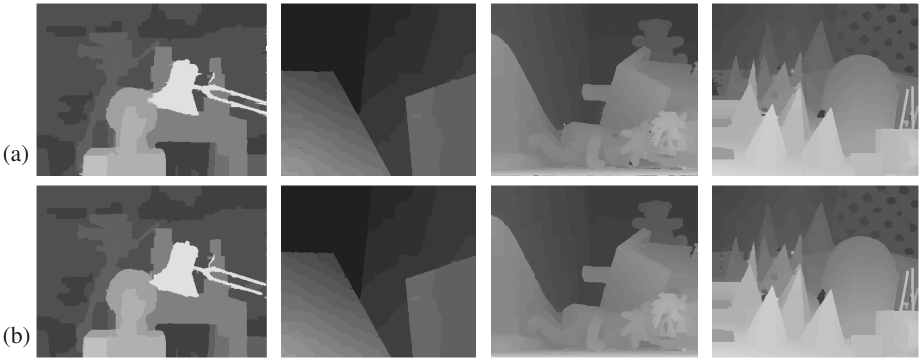


Fig. 4. Disparity result comparison. (a) Disparity results of "SEG" (b) Our final disparity results using the Outlier Confidence model ("SEG+OC").

We show in the first row of Table 3 (denoted as "COLOR") the statistics of the initial disparities. The algorithm is detailed in Section 4.1. We set $U(x) = 0$ for all x's and minimize the energy defined in (13). Then we estimate the OC maps based on the initial disparities and minimize the energy defined in (1). We denote the final results as "COLOR+OC" in the second row of Table 3.

Comparing the two sets of results, one can observe that incorporating the outlier information significantly improves the quality of the estimated disparity maps. The overall rank jumps from initial No. 16 to No. 5, which is the highest position for all results produced by the stereo matching algorithms without incorporating segmentation.

In analysis, for the "Teddy" example, however, our final disparity estimate does not gain large improvement over the initial one. It is because that the remaining errors are mostly caused by matching large textureless regions, which can be addressed by color segmentation.

5.2 Results Using Segmentation

In this part of the experiments, we incorporate the segmentation information by using the data term $f_0^{(2)}$ defined in (4). Our initial disparities are denoted as "SEG". Our final results obtained by applying the global optimization incorporating the Outlier Confidences are denoted as "SEG+OC". We show in the third and forth rows of Table 3 the error statistics of the initial disparity maps and our refined results. The average rank rises from 6.9 to 3.6 and the overall rank jumps from No. 4 to No. 2. The improvement validates the effectiveness of our approach in handling outliers and its nature of complementarity to color segmentation.

The computed disparity maps are shown in Figure 4, where (a) and (b) respectively show the results of "SEG" and "SEG+OC". A comparison of disparity errors is demonstrated in Figure 5 using the "Cones" example. The magnified patches extracted from the error maps are shown in (b). The comparison shows that our approach can primarily improve the disparity estimation for outlier pixels.

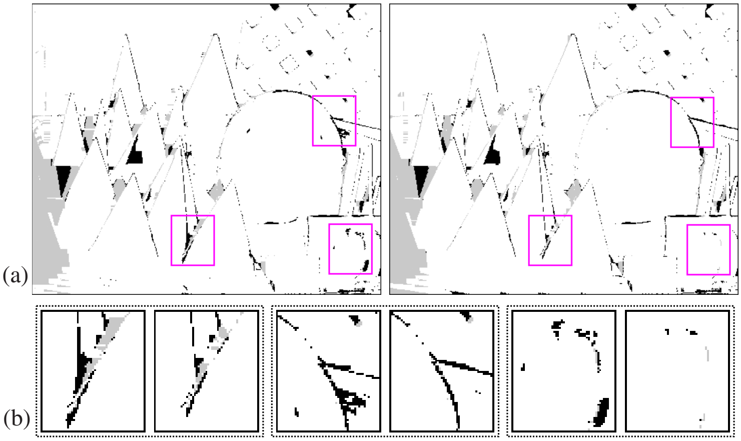


Fig. 5. Error comparison on the “Cones” example. (a) shows the disparity error maps for “SEG” and “SEG+OC” respectively. (b) Comparison of three magnified patches extracted from (a). The “SEG+OC” results are shown on the right of each patch pair.

Finally, the framework of our algorithm is general. Many other existing stereo matching methods can be incorporated into the outlier confidence scheme by changing f_0 to other energy functions.

6 Conclusion

In this paper, we have proposed an Outlier-Confidence-based stereo matching algorithm. In this algorithm, the Outlier Confidence is introduced to measure how likely that one pixel is an outlier. A model using the local color information is proposed for inferring the disparities of possible outliers and is softly combined with other data terms to dynamically adjust the disparity estimate. Complementary to global color segmentation, our algorithm locally gathers color samples and optimizes them using the matting techniques in order to reliably measure how one outlier pixel can be assigned a disparity value. Experimental results on the Middlebury data set show that our proposed method is rather effective in disparity estimation.

Acknowledgements

This work was fully supported by a grant from the Research Grants Council of Hong Kong (Project No. 412708) and is affiliated with the Microsoft–CUHK Joint Laboratory.

References

1. Tao, H., Sawhney, H.S., Kumar, R.: A global matching framework for stereo computation. In: ICCV, pp. 532–539 (2001)
2. Sun, J., Zheng, N.N., Shum, H.Y.: Stereo matching using belief propagation. IEEE Trans. Pattern Anal. Mach. Intell. 25(7), 787–800 (2003)

3. Hong, L., Chen, G.: Segment-based stereo matching using graph cuts. In: CVPR (1), pp. 74–81 (2004)
4. Sun, J., Li, Y., Kang, S.B.: Symmetric stereo matching for occlusion handling. In: CVPR (2), pp. 399–406 (2005)
5. Klaus, A., Sormann, M., Karner, K.F.: Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In: ICPR (3), pp. 15–18 (2006)
6. Yang, Q., Wang, L., Yang, R., Stewénius, H., Nistér, D.: Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling. In: CVPR (2), pp. 2347–2354 (2006)
7. Kang, S.B., Szeliski, R.: Extracting view-dependent depth maps from a collection of images. International Journal of Computer Vision 58(2), 139–163 (2004)
8. Strecha, C., Fransens, R., Van Gool, L.J.: Combined depth and outlier estimation in multi-view stereo. In: CVPR (2), pp. 2394–2401 (2006)
9. Scharstein, D., Szeliski, R.: http://vision.middlebury.edu/stereo/eval/
10. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47(1-3), 7–42 (2002)
11. Zhang, L., Seitz, S.M.: Parameter estimation for mrf stereo. In: CVPR (2), pp. 288–295 (2005)
12. Zitnick, C.L., Kang, S.B., Uyttendaele, M., Winder, S.A.J., Szeliski, R.: High-quality video view interpolation using a layered representation. ACM Trans. Graph. 23(3), 600–608 (2004)
13. Yoon, K.J., Kweon, I.S.: Stereo matching with the distinctive similarity measure. In: ICCV (2007)
14. Yoon, K.J., Kweon, I.S.: Adaptive support-weight approach for correspondence search. IEEE Trans. Pattern Anal. Mach. Intell. 28(4), 650–656 (2006)
15. Lei, C., Selzer, J.M., Yang, Y.H.: Region-tree based stereo using dynamic programming optimization. In: CVPR (2), pp. 2378–2385 (2006)
16. Strecha, C., Fransens, R., Gool, L.J.V.: Wide-baseline stereo from multiple views: A probabilistic account. In: CVPR (1), pp. 552–559 (2004)
17. Hirschmüller, H.: Accurate and efficient stereo processing by semi-global matching and mutual information. In: CVPR (2), pp. 807–814 (2005)
18. Hirschmüller, H., Scharstein, D.: Evaluation of cost functions for stereo matching. In: CVPR (2007)
19. Yang, Q., Yang, R., Davis, J., Nistér, D.: Spatial-depth super resolution for range images. In: CVPR (2007)
20. Hasinoff, S.W., Kang, S.B., Szeliski, R.: Boundary matting for view synthesis. Computer Vision and Image Understanding 103(1), 22–32 (2006)
21. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24(5), 603–619 (2002)
22. Chuang, Y.Y., Curless, B., Salesin, D., Szeliski, R.: A bayesian approach to digital matting. In: CVPR (2), pp. 264–271 (2001)
23. Wang, J., Cohen, M.F.: Optimized color sampling for robust matting. In: CVPR (2007)
24. Grady, L.: Random walks for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 28(11), 1768–1783 (2006)
25. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient belief propagation for early vision. In: CVPR (1), pp. 261–268 (2004)