Stereo Matching with Global Edge Constraint and Graph Cuts

Hong Zhang, Feiyang Cheng, Ding Yuan, Yuecheng Li, Mingui Sun

Abstract

We propose a novel method on stereo matching based on the Global Edge Constraint (GEC) and Graph Cuts. Firstly, the GEC composed of particular image edges is employed to generate the initial disparity maps. And then the reliable disparity maps consistent with the observed data are extracted to construct the data term of the energy function. Finally, we incorporate the GEC as a soft constraint into our global optimization framework, and the optimal solution could be approximated via graph cuts. Experimental results demonstrated the good performance of our proposed approach.

1. Introduction

Stereo matching in computer vision has been heavily investigated for decades because of its crucial role in numerous modern applications, such as 3D reconstruction, image-based rendering etc. Global algorithms have been conducted and dramatic results have been obtained over the last few years [1, 2] due to powerful new optimization algorithms such as graph cuts [3, 4] and loopy believe propagation (LBP) [5]. However, the energy function defined in the traditional global algorithms usually has higher cost near the occlusions and the disparity discontinuities, as they have to make explicit smoothness assumptions. Hence, one solution is to incorporate some prior constraints into the global framework to approximate the minimal energy by locating these particular regions. According to the Middlebury benchmark, segment-based prior has been widely explored due to its first-rank performance [1, 6]. These methods usually adopt the color-based segmentation to segment the reference image as a pre-processing procedure. The matching results greatly depend on how the segmentation results are coincident with the locations of the occlusions and disparity discontinuities. However, the coincidence could hardly be achieved when tackling the highly-textured areas and the non-planar surfaces with uniform color in the images.

Motivated by the problems as what we have discussed, we propose a new regularization prior in which we explore the characteristics of different kinds of edges in both stereo pair of images. Inspired by some research works [7] which concluded that both disparity discontinuities and occlusions will just appear near the object boundaries and its neighboring edges, we examined the characteristics of different kinds of edges and proposed the GEC [8] to locate the occlusions and the disparity discontinuities. Different from the segment-based methods, our algorithm only concern the disparity space instead of the color information of the images.

We have proposed a method to establish initial disparity maps based on the GEC [8]. In this paper, the reliable disparity maps consistent with the observed data are extracted to construct the data term of the energy function at first. Then the prior GEC is incorporated as a soft constraint into our global optimization framework. We employ the α-β-swap algorithm to minimize the energy function [3].

2. Stereo matching algorithm

2.1 Definition of the GEC

In our previous work [8], we proposed a method to classify the edges into several groups based on the variance of neighboring pixels. Firstly, we categorize the detected edges of both images into two groups: the
match-edges represent the edges on which the pixels do have correspondences in the second image, and the unmatch-edges represent the edges where the pixels have no correspondences. Moreover, the match-edges could be further categorized into two groups: pull-edges and internal-edges. The pull-edges are the possible candidates of disparity discontinuities (i.e., object boundaries). And the internal-edges are the other kind of match-edges which have approximately constant disparities within some specific regions. The GEC is then defined as the combination of the unmatch-edges and the pull-edges. We utilize the GEC to locate the occlusions and the disparity discontinuities. The estimated GECs of both images of the stereo pair named Cones are shown in Figure 1.

Based on the GEC, we employed variable correlation windows and propagation strategies [8] to estimate the initial disparity maps. The initial disparity map of Cones is shown in Figure 2(a).

![Figure 1. The categorization of image edges. (a) Stereo image pair: left and right images; (b) Different groups of edges: internal-edges (green), pull-edges (red), unmatch-edges (blue); (c) The GECs extracted from (b).](image)

### 2.2 Extracting of the observed data-based reliable disparity maps

Traditionally, the energy function is composed of data term and smooth term. The data term penalizes solutions that are inconsistent with the observed data (i.e., pixel dissimilarity) and the smooth term enforces the piecewise smoothing assumption [3]. We consider the energy function:

$$E(d) = E_{\text{data}}(d) + E_{\text{smooth}}(d) = \sum_{x \in I} D(d_x) + \sum_{(x,y) \in \mathcal{N}} V_{xy}(d_x, d_y)$$  \hspace{1cm} (1)

where $I$ represents the reference image and $\mathcal{N}$ is the set of interacting pairs of pixels. $d_x$ and $d_y$ represent the disparities of pixel $x$ and $y$ respectively.

In (1), $D(d_x)$ (here $d_x \in \text{Label}$, and Label = 0, 1, ..., $d_{\text{max}}$), usually consistent with the observed data, is computed to estimate the optimal disparity of pixel $x$ without any prior disparity candidates in most of the global optimization algorithms, and it is not always unique. Thus the estimation of the disparity of pixel $x$ is not accurate by using the traditional methods. Hence, we prefer extracting reliable disparity maps which are consistent with the observed data at first, and constructing the data term subsequently by searching a few possible disparity candidates for pixel $x$ in the reliable disparity maps. In this way, more reliable information can be provided to the global optimization algorithm.

The technique mentioned in [9] is employed to extract the reliable disparity maps, and could be expressed in:

$$P^d(x, x') = \min_{q \in \lfloor x^* - 1/2, x^* + 1/2 \rfloor} |L(x) - R^d(q)|$$  \hspace{1cm} (2)

where the capitals $L$ and $R$ symbolize the 1st image and the 2nd image respectively. $L(x)$ and $R(x^*)$ represent the intensities of pixel $x$ and its correspondence $x^*$ respectively. We use $L'(q)$ and $R'(q)$ to represent the linear interpolated functions of the sample points on the scan lines. $P^d(x, x')$ measures how well $L(x)$ fits into the intensities $[R(x^* - 1/2), R(x^*), R(x^* + 1/2)]$.

And similarly we define $P^b(x, x')$ for the 2nd image,

$$P^b(x, x') = \min_{q \in \lfloor x^* - 1/2, x^* + 1/2 \rfloor} |L'(q) - R'(x^*)|$$  \hspace{1cm} (3)

A symmetric measure of pixels’ dissimilarity between $x$ and $x^*$ is defined as:

$$P(x, x^*) = \min \left[ P^d(x, x^*), P^b(x, x^*) \right]$$  \hspace{1cm} (4)

$P(x, x^*)$ has been proved to be insensitive to image sampling in [9]. In the initial disparity maps, we compute $P(x, x^*)$ and eliminate the disparity of pixel $x$ if $P(x, x^*) > t$. Actually $x$ and $x^*$ may not be of the same intensity even if they are exactly the corresponding pair, due to the noise, illumination conditions etc. Therefore, it is reasonable to let $t > 0$. Experiments showed that $t = 4$ is a good compromise to have an accurate extracting of the observed data-based reliable disparity maps. The corresponding result of Cones is shown in Figure 2(b).

![Figure 2. Results of Cones (a) the initial disparity map; (b) the observed data-based reliable disparity map.](image)
maps. It is noticed that most of these pixels distribute in the occluded areas, except that a small portion of pixels have no disparities due to various uncertain reasons such as noise etc.

In the reliable disparity maps, we use notation \( f_{null} \) to represent the pixels having no disparities, and \( f_d \) to represent the pixels who have initial disparities. For the two groups of pixels, we define data term as (5):

\[
D(d_x) = \begin{cases} P(x, x'), & x \in f_d, d_x \in L_w \\ 0, & x \in f_{null}, d_x \in L_w \\ 255, & x \in f_d \cup f_{null}, d_x \in others \end{cases}
\]

where \( W \) is a finite window centered at pixel \( x \). \( L_w \) is the set of the disparities within window \( W \). In this work, \( W \) is squared and of size \( 11 \times 11 \), so as to cover sufficient disparity candidates.

For any pixel \( x \in f_d \), the disparities, which are neighboring \( x \), are considered as the disparity candidates, and the optimal disparity could be determined by computing \( D(d_x) \), where \( d_x \in L_w \), in the global optimization algorithm. As to the pixels in set \( f_{null} \), most of them possibly locate in the occluded areas as what we have mentioned above. Thus for any pixel \( x \in f_{null} \), \( D(d_x) = P(x, x') \) results in a large data term, even though \( d_x \) is the correct disparity of pixel \( x \). Aiming at this problem, for any pixel \( x \in f_{null} \), if \( d_x \in L_w \), we let \( D(d_x) = 0 \) to avoid the data term going into effect.

2.4 The smooth term

In energy function (1), smooth term \( V_{xy} \) is given by Potts model:

\[
V_{xy} = u_{\{x,y\}} \cdot T(d_x \neq d_y), \quad T = \begin{cases} 1, & d_x \neq d_y \\ 0, & d_x = d_y \end{cases}
\]

where \( x \) and \( y \) are two neighboring pixels in the observed data-based reliable disparity maps.

Suppose \( I_x \) and \( I_y \) are the intensities of two neighboring pixels \( x \) and \( y \) in the first image respectively. It is reasonable to assume that \( x \) and \( y \) share the same disparity, if their intensities satisfy \( I_x \approx I_y \). Boykov and Veksler et al [3] incorporated the above contextual information into their framework by allowing \( u_{\{x,y\}} \) to vary according to the intensities \( I_x \) and \( I_y \). And term \( u_{\{x,y\}} \) is defined as:

\[
u_{\{x,y\}} = U(\mid I_x - I_y \mid) = \begin{cases} 2K, & \text{if } \mid I_x - I_y \mid \leq 5 \\ K, & \text{if } \mid I_x - I_y \mid > 5 \end{cases}
\]

We let Potts model parameter \( K = 20 \) as used in [3].

The prior GEC is incorporated into our framework as a soft constraint based on the above model. We hope that the disparity discontinuities are coincident with the pull-edges by giving lower smooth penalties to the pull-edges in the smooth term. Motivated by this purpose, we set \( u_{\{x,y\}} \) in the following form:

\[
u_{\{x,y\}} = \begin{cases} 2K, & \text{if edges do not exist at pixel } x \text{ or } y \\ K, & \text{if edges do exist at pixel } x \text{ or } y \\ 0.5K, & \text{if pull-edges do exist at pixel } x \text{ or } y \end{cases}
\]

where \( u_{\{x,y\}} \) will be smaller if the potential disparity discontinuity is detected according to the GEC.

Finally, in the last step the disparity maps are improved by applying median filter. The window of median filter could be determined according to the GEC, as illustrated in Figure 3. By using this method, the smoothing of disparity discontinuities can be effectively avoided.

![Figure 3. Suppose region A and C belong to the same object surface, region B is adjacent to A, and C is the occluded region between A and B in the disparity map. Edges marked in green represent the GEC. For any pixel \( x \in C \), the window of median filter is not permitted to straddle the GEC, therefore the smoothing of disparity discontinuities can be avoided. However for any pixel \( y \in A \), we adopt square window centered at \( y \) to filter the disparity map. We let the window be a \( 5 \times 5 \) square window to improve the matching results locally. Since the window covers a small region of the image and the disparity map within the window varies smoothly, median filter could be utilized as a smoothing function.]

3. Experimental results

We tested our algorithm on the standard Middlebury stereo pairs [10] to prove the good performance of our algorithm. Middlebury benchmark which is commonly accepted in the field of stereo vision was adopted to evaluate our matching results. The comparison to other stereo matching algorithms is shown in Table I. Our algorithm currently takes rank 17 with the error threshold=1 among approximate 115 methods. It proved that our method is one of the top-performing algorithms which improved the ill-posed stereo matching problem. Figure 4 shows the corresponding disparity maps established by using our algorithm. All the test images are the standard stereo pairs downloaded from the webpage of Middlebury. Compared to the other methods, our algorithm established more smooth disparity maps within object
surfaces and obtained the best disparity map of the stereo pair named Cones.

4. Conclusions

A novel algorithm on stereo matching is proposed. The proposed GEC is incorporated into the global framework and excellent results can be obtained. Superior performance on the Middlebury test bed demonstrates that our algorithm is a competitive one.

Further work should concentrate on improving the accuracy of the estimation of the GEC. Furthermore, more effective incorporation method is needed to develop robust global optimization framework.

References


![Figure 4. The disparity maps obtained by using our algorithm: the 1st row-Tsukuba, the 2nd row-Venus, the 3rd row-Teddy, and the 4th row-Cones; (a) reference images; (b) ground truth; (c) corresponding disparity maps of our algorithm.](image)

**Table I:** The rank on the Middlebury Stereo test bed[10].

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Avg. Rank</th>
<th>Tsukuba</th>
<th>Venus</th>
<th>Teddy</th>
<th>Cones</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>nonocc</td>
<td>all</td>
<td>disc</td>
<td>nonocc</td>
</tr>
<tr>
<td>Info+ Permeable</td>
<td>24.3</td>
<td>1.06</td>
<td>1.53</td>
<td>1.11</td>
<td>5.64</td>
</tr>
<tr>
<td>CostFilter</td>
<td>25.3</td>
<td>1.51</td>
<td>1.85</td>
<td>5.44</td>
<td>7.61</td>
</tr>
<tr>
<td>GlobalGCP</td>
<td>25.6</td>
<td>0.87</td>
<td>2.54</td>
<td>52</td>
<td>4.69</td>
</tr>
<tr>
<td>Our Method</td>
<td>26.6</td>
<td>1.53</td>
<td>1.75</td>
<td>25</td>
<td>8.07</td>
</tr>
<tr>
<td>FeatureGC</td>
<td>26.7</td>
<td>1.08</td>
<td>1.47</td>
<td>1.19</td>
<td>5.82</td>
</tr>
<tr>
<td>AdaptOvrSegBP</td>
<td>26.8</td>
<td>1.69</td>
<td>2.04</td>
<td>43</td>
<td>5.64</td>
</tr>
<tr>
<td>P-LinearS</td>
<td>27.5</td>
<td>1.10</td>
<td>1.67</td>
<td>19</td>
<td>5.92</td>
</tr>
</tbody>
</table>

The first number in each column shows the error rate of the method, and the second number denotes the rank.