Robust local stereo matching under varying radiometric conditions

Yufu Qu¹, Jixiang Jiang¹, Xiangjin Deng², Yanhong Zheng²

¹School of Instrument Science and Opto-Electronics Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, People’s Republic of China
²Beijing Institute of Spacecraft System Engineering, Beijing 100094, People’s Republic of China
E-mail: yufugu@gmail.com

Abstract: The authors present a local stereo matching algorithm whose performance is insensitive to changes in radiometric conditions between the input images. First, a prior on the disparities is built by combining the DAISY descriptor and Census filtering. Then, a Census-based cost aggregation with a self-adaptive window is performed. Finally, the maximum a-posteriori estimation is carried out to compute the disparity. The authors’ algorithm is compared with both local and global stereo matching algorithms (NLCA, ELAS, ANCC, AdaptWeight and CSBP) by using Middlebury datasets. The results show that the proposed algorithm achieves high-accuracy dense disparity estimations and is more robust to radiometric differences between input images than other algorithms.

1 Introduction

Dense stereo matching is an active area of research in the field of computer vision. In visual applications, such as autonomous driving and virtual reality, dense stereo matching is an important and vital issue [1]. An accurate disparity map can help unmanned vehicle to navigate in real environment. In virtual reality, the dense disparity map is necessary for proper reconstruction of object boundaries. Most dense stereo matching methods are based on the smoothness assumption. This assumes that the corresponding pixels have similar intensities or colour values. In reality, however, the smoothness assumption does not hold and radiometric differences are often found between input images. Variation in illumination and exposure are the two radiometric differences that can cause the most problems in dense stereo matching [2]. The performance of most dense stereo matching algorithms can be severely degraded under these conditions. To avoid a high matching error, most dense stereo matching algorithms are limited to images photographed with the same camera settings and under the same lighting conditions.

Dense stereo matching methods generally consist of four main steps: (1) matching cost computation, (2) cost aggregation, (3) disparity computation/optimisation and (4) disparity refinement. Dense stereo matching methods can be separated into two broad classes, local and global methods [3], depending on whether step (2) is included. For local algorithms the matching costs are aggregated over fixed or varying support windows, depending on the image intensities or colour values, and usually make implicit smoothness assumptions by aggregation. A winner-takes-all (WTA) strategy is then adopted to select the minimum cost disparity. In global methods, explicit smoothness assumptions are used. The disparity at one pixel is estimated using the estimated disparity at all other pixels [4], which involve complex global energy functions. The main distinction between global algorithms is the minimisation procedure used. Local methods are typically fast, but do not reliably match areas with little texture or discontinuities in depth. Global methods, on the other hand, can achieve high-accuracy stereo matching, but are computationally expensive and require a large memory space. To obtain high-accuracy, fast, stereo matching, Geiger et al. [5] proposed combining local and global methods by using a Bayesian dense stereo matching algorithm. Their method builds a prior over the disparity space by forming a triangulation on a set of support points to disambiguate the matching problem. This improves efficiency by restricting the search to plausible regions. Experimental results show that this method has high precision and is fast. For instance, computing the left and right disparity maps for a 1 Megapixel image pair takes about 1 s on a single CPU core. However, their method, like other state-of-the-art local methods, is sensitive to radiometric variations that may occur between input images. Fig. 1 shows that dense and high-accuracy matching can be obtained using the method of Geiger et al. [5] only when the left and right images are captured under the same radiometric conditions. As the variation in radiometric conditions between the images increase, the results substantially deteriorate.

In this paper, we present a method that is able to achieve high-accuracy local stereo matching regardless of radiometric variations between the left and right images. Inspired by Geiger et al. [5], our method relies on the
maximum a-posteriori energy function to compute the disparities. The matching cost computation and cost aggregation are also redesigned. The method consists of five main steps: (i) prior estimation of disparities, (ii) matching cost computation, (iii) cost aggregation, (iv) disparity computation and (v) disparity refinement.

Our method has two main innovations over previous methods. First, in order to improve accuracy and robustness, our method builds a prior over the disparity space by combining the DAISY [6] descriptor and Census [7] filtering. This efficiently restricts the disparity search to plausible regions only. Secondly, to properly handle areas with low texture and disparity boundaries of input greyscale images, our method includes our newly developed Census-based cost aggregation with a self-adaptive window strategy.

2 Related works

The first key step for all dense stereo matching algorithms is matching cost computation [8]. Traditionally, matching cost computing methods are pixel-based, for example, absolute intensity difference (AD) [9] and squared intensity differences [10], or window-based, for example, sum of absolute difference, sum of squared differences and normalised cross-correlation (NCC) [11], or common filter-based, for example, Laplacian of Gaussian [9] filtering, mean filtering and bilateral filtering [12], or involve non-parametric-based matching, for example, Rank and Census [7]. With the increasing use of stereo matching algorithms for real-world scenarios, the focus has shifted towards matching cost methods that are not sensitive to radiometric differences between the input images. Although a number of different matching cost methods have been developed, most of them only achieve satisfactory results when radiometric variations between the left and right images are small.

Hirschmuller and Scharstein [13] compared the performance of 15 different matching cost methods on images with simulated and real radiometric differences. They varied the light intensity, exposure and noise. The Census algorithm was found to perform the best performing and be the most robust. Recently, Heo et al. [14] presented a new radiometric invariant matching cost algorithm named adaptive NCC (ANCC). This method employs the colour formation model and extracts colour invariant information from the input images taken under different radiometric conditions. Although robust and dense matching can be obtained using this method, the performance and speed of this approach is highly dependent on window size. To compensate for radiometric differences between images, it was suggested that a mutual information matching cost method could be used [15]. The use of a colour transform-based matching cost method has also been proposed [16], where the colours of the stereo images are transformed adaptively so that matching points have the same colours. However, this method does not take into account the effects of occluded pixels. AD-Census matching cost methods were shown in [17]. However, because of the use of colour information in the matching cost step, these only work when the input images have very minor radiometric differences.

The second key step for local stereo matching methods is cost aggregation. Cost aggregation methods are traditionally performed by summing or averaging the matching cost over pre-defined windows using various smoothness assumptions. As a result, the cost aggregation step can significantly influence the computation speed. Viola and Jones [18] carried out cost aggregation using an un-normalised box-filter, where the speed is related to the size of the support window. To improve efficiency, Min et al. [19] reduced the search range using a subset of informative disparity hypotheses. However, this method cannot obtain accurate results at depth discontinuities as the aggregation windows located on depth edges represent pixels from different depths. To solve these problems, Veksler [20] developed a cost aggregation method with size-adaptive windows. Kang et al. [21], on the other hand, adopted a multiple-window strategy that selects the optimal

![Fig. 1](image-url)
aggregation window from a set of pre-defined windows of the same size that are located at different positions. Yoon and Kweon [22] demonstrated an adaptive support-weight aggregation approach that improved the performance at depth discontinuities. The support weight of the pixels, in a fixed-size window, was calculated using colour similarity and geometric proximity. Although the approach developed by Yoon and Kweon [22] can accurately preserve depth edges, it is much slower than fixed-window methods; this limits the size of the input images that can be used. A non-local cost aggregation matching method using a minimum spanning tree [4] has recently been shown to be the top performing local stereo matching algorithm. However, the performance of this algorithm, along with other state-of-the-art local matching algorithms, is strongly dependent on the image colour information.

The colour-based aggregation approaches mentioned above all perform reasonably well, however, choosing the appropriate support window is not always easy. To accurately estimate disparities in homogeneous regions, the support window must be large enough to include as many points at the same, unknown, depth as possible. At the same time, differently shaped supports window are needed near object boundaries to avoid crossing depth discontinuous areas. Although the different methods discussed all provide dense and accurate results, they are slow and therefore not suitable for practical applications. In addition, experimental results show that most of these methods only work for colour images taken under consistent radiometric conditions. Therefore a simple and efficient cost aggregation strategy for greyscale images is still needed.

In local stereo matching methods, emphasis is placed on the matching cost computation and cost aggregation. The final disparity is computed using a local WTA optimisation at each pixel. The drawback of a WTA optimisation is that it minimises a global cost function, which makes explicit smoothness assumptions. The algorithms are formulated using an energy minimisation framework. The disparity solution is chosen such that it minimises a global cost function, which includes data and smoothness terms [8]. Examples of energy minimisation, disparity optimisation, routines include graph cut [23], belief propagation [24] and top-rank algorithms of the Middlebury benchmark [25, 26].

State-of-the-art stereo matching can be achieved using global matching methods. However, the energy minimisation techniques used for disparity optimisation generally require large computational efforts with high memory capacities. To improve the efficiency of global stereo matching methods, Veksler [27] reduced the search space using local stereo matching methods. The common graph cut technique was then used to minimise the energy function. The method used by Veksler [27] successfully reduced the computer memory required, but did not increase the computation speed. Introducing physical constraints into the simulated annealing algorithm enables good stereo matching, but does not solve depth discontinuities [28].

Geiger et al. [5] used a global approach to calculate disparities in local stereo matching methods via a novel dense stereo matching algorithm based on a generative probabilistic model. They achieved fast, state-of-the-art performance by building a prior on the disparities. The prior approach is similar to global matching methods. However, this algorithm is more similar to local matching methods because of the traditional fixed window used in the cost aggregation step. Unlike other local matching methods, this approach uses the maximum a-posteriori estimation to compute the final disparity. This gives accurate results without loss of speed. The biggest drawback in the method developed by Geiger et al. [5] is its sensitivity to radiometric variations [13]. Here, we present a dense local algorithm that is invariant to radiometric variations. This algorithm is based on that of Geiger et al. [5], but with redesigned, high-performance matching cost computation and cost aggregation steps.

3 Our approach

In our approach, the left and right greyscale images are defined as the reference and target images, respectively. Our algorithm contains the following steps: (i) a prior on the disparities is built by combining the DAISY descriptor and Census filtering. (ii) Census-based cost aggregation with a self-adaptive window is performed. (iii) The maximum a-posteriori estimation is performed to compute the disparity. (iv) A post-processing step is carried out to optimise the computed disparities.

3.1 Prior estimation of disparities

The first step of our method, prior estimation of disparities, searches the reference image for a set of robust matching points. These matching points are used to construct piecewise-linear disparity planes and provide a prior on the disparities of the remaining points. Finding robust matching points is therefore key to obtaining the correct disparity estimation. Edge points are ideal robust matching points as the image edge can be matched with high certainty. Points other than edge points can also be robustly matched. We therefore also consider regular grid sampling points as candidate robust matching points.

To enable stereo matching that performs well even when there are radiometric variations between the input images, the DAISY descriptor is used to compute the descriptor at every point in the image transformed by Census filtering. The robust matching points are then matched using the DAISY descriptor directly. The half width of the input image is set as the initial disparity reach range $D_r$. The terms $I_{RCensus}$ and $I_{TCensus}$ correspond, respectively, to the reference and target images transformed by Census filtering. $DAISY(u,v)$ is the descriptor for the pixel $p(u,v)$ in $I_{RCensus}$ and $d$ is the corresponding disparity level. The corresponding matching point and descriptor in $I_{TCensus}$ for $p(u,v)$ are given by $p(u-d,v)$ and $DAISY(u-d,v)$. The probability $P(d)$ that $p(u,v)$ has a disparity $d$ is

$$P(d) = \frac{\exp\left(-\frac{1}{\delta} \left\| DAISY(u,v) - DAISY(u-d,v) \right\|^2\right)}{\sum_{d \in D_r} \left\| DAISY(u,v) - DAISY(u-d,v) \right\|^2}$$

(1)

where $\delta$ indicates the sharpness of the Gaussian distribution.
The final disparity is found using a WTA strategy such that

\[
d(u, v) = \arg \max d \in D_r P(d)
\]

To avoid ambiguous matching, the ratio of the first two probability maxima is calculated. Only ratios above a fixed threshold are considered to represent robustly matched points. In all the experiments, the threshold is set to 0.97. To remove occluded areas, correspondences are retained only if they pass cross-checking (comparison of left-to-right and right-to-left disparity maps). Fig. 2 shows the performance of computing the support points using the DAISY descriptor for radiometrically varying images. Visual comparison in Figs. 2 shows that the DAISY descriptor remains stable regardless of radiometric variations between the left and right images.

After establishing a set of robustly matched support points, a prior on the disparities for the remaining points is built. Fig. 3 shows the mapping relationship between the reference image and disparity space. Let \(d_s(u, v)\), \(s \in \{i, j, k\}\) be the disparity of a robustly matched point \(p_s(u, v)\) in the reference image. The disparity plane can be found using the disparities of the three matched points \(p_s(u, v), p_j(u, v), p_k(u, v)\), such that

\[
d_{\text{plane}}(u, v) = au + bv + c
\]

where \((a, b, c)\) are the parameters of the disparity plane, and \(d_{\text{plane}}(u, v)\) is the disparity of the pixel \(p(u, v)\) in the reference image. For an unmatched point \(p_s(u, v)\), a prior estimate of \(d_{\text{plane}}\) can be obtained from (3) using neighbouring matched points. There are often more than three matched points around each unmatched point. In the experiments we only consider matched points contained in a \(17 \times 17\) pixel window. Let \(D_{\text{plane}}(u, v) = \{d_{\text{plane}1}, d_{\text{plane}2}, \ldots, d_{\text{plane}N}\}\) be a set of prior disparity estimates for \(p_s(u, v)\), where \(N\) is the number of the disparity planes computed in the square window. The prior probability of \(p_s(u, v)\) having a disparity \(d\) is (see (4))

\[
P_{\text{prior}}(d) = \begin{cases} k_1 + \frac{1}{2\pi\sigma^2} \exp \left( -\frac{(d - d_{\text{plane}})^2}{2\sigma^2} \right), & \text{if } |d - d_{\text{plane}}| < 3\sigma \\forall d \in N_s \\ 0, & \text{otherwise} \end{cases}
\]

where \(d_{\text{plane}} \in (\min D_{\text{plane}}(u, v), \max D_{\text{plane}}(u, v)), N_s\) is the set of robust matching points disparities in the \(17 \times 17\) pixel neighbourhood around \(p_s(u, v)\). \(\sigma\) is the variance of the Gaussian Kernels and \(k_1\) is a constant. To accurately calculate disparity discontinuities in places where the linearity assumption might be violated, the condition \(d \in N_s\) is taken into account in (4). From (4) it can be seen that the disparity search range is small for each unmatched point. This is the key feature to obtaining accurate dense matching over short timescales.

### 3.2 Matching cost computation

Census filtering is a non-parametric local transform that encodes pixel intensities into the local image structures. Each pixel of the resulting transformed image contains
and Iq i

The cost value intensity

The 64-bit string ⊗ pixels in size.

window strategy. Inspired by the cost-volume designed a cost aggregation function with a self-adaptive

to obtain satisfactory performance in poorly textured areas

depth.

aggregation window, of similar intensities have the same

method [29], we assume that pixels, in the same

be shown. The corresponding transformed images are

is a powerful tool and can extract the image structure even

poorly lit and highly textured regions. This is crucial for

stereo matching algorithms.

In the reference image, R(qi) is the intensity of point q(u, v),

and N(qi) is a 9 × 7 pixel window surrounding q(u, v). If

I(qi) > I(qr) then η(qi, qr) = 1, otherwise η(qi, qr) = 0.

The 64-bit string R(qi) is then given by

\[ R(q_i) = \circleslash_{[i \in N(q_i)]} \eta(q_i, q_i + [i, j]) \] (5)

where \( \circleslash \) is concatenation. The 64-bit string R(qi) can be obtained for the target image using the same calculation.

The cost value \( C_{\text{census}}(u, v, d) \) is then computed as follows

\[ C_{\text{census}}(u, v, d) = H(R(q_i), R(q_r)) \] (6)

where H is the Hamming distance between R(qi) and R(qr).

3.3 Cost aggregation

To obtain satisfactory performance in poorly textured areas and regions with many depth discontinuities, this paper designed a cost aggregation function with a self-adaptive window strategy. Inspired by the cost-volume filtering method [29], we assume that pixels, in the same aggregation window, of similar intensities have the same depth.

Here, S is the number of pixels in a support window of \( a \times a \) pixels in size. \( I_i(u, v) \) is the intensity of the central pixel \( p_i(u, v) \), and \( I(u_i, v_i) \) is the intensity of pixel \( p_i(u_i, v_i) \). The mean intensity \( u_{\text{mean}} \) of the window is

\[ u_{\text{mean}} = \frac{\sum_{i \in S} I_i(u_i, v_i)}{S} \] (7)

As shown in Fig. 5, for the reference image, \( d \) of \( I_i(u, v) \) is calculated using the binary support weight \( w_i(u_i, v_i) \) of each pixel in the support window, where

\[ w_i(u_i, v_i, d) = \begin{cases} 1, & (I_i(u, v) - u_{\text{mean}})(I_i(u, v) - u_{\text{mean}}) > 0 \\ 0, & \text{otherwise} \end{cases} \] (8)

The cost aggregation function at \( p_i(u_i, v_i) \) is then given by

\[ C_{\text{agg}}(u, v, d) = \sum_{i \in S} w_i(u_i, v_i, d)C_{\text{census}}(u_i, v_i, d) \] (9)

3.4 Disparity computation

The dense disparity map is found using the cost aggregation function defined in (9). Unlike traditional dense local matching methods, where the minimum cost disparity is calculated from a local WTA optimisation at each pixel, our method calculates the disparities using the maximum a-posteriori estimation. In the reference image, \( O(u, v, d) = \{o_1(u_1, v_1, d_1), ..., o_N(u_N, v_N, d_N) \} \) of robustly matched points around \( q_i(u, v) \) for a 17 × 17 pixel support window. The corresponding point in the target image is \( q_i(u - d, v) \). Using the maximum a-posteriori estimation to compute the final disparity of \( q_i(u, v) \) gives

\[ d^* = \arg \max P(d|q_i, q_r, O) \] (10)

The full conditional probability of \( d \) is given by

\[ P(d|q_i, q_r, O) \propto P(d|q_i, O)P(q_i|q_r, d) \] (11)

The prior probability of \( q_i(u, v) \) was given in (4). Replacing \( P(d|q_i, O) \) with \( P_{\text{prior}}(d) \) in (11) gives

\[ P(d|q_i, q_r, O) \propto P_{\text{prior}}(d)P(q_i|q_r, d) \] (12)

Fig. 4 Performance of Census filtering on Aloe images

a–d Aloe images taken under different illumination conditions
e–h Results using Census filtering for (a)–(d)

IET Comput. Vis., 2014, Vol. 8, Iss. 4, pp. 263–276
© The Institution of Engineering and Technology 2014
Using the cost aggregation function shown in (9), the image likelihood $P(q_r|q_l, d)$ is given by a Laplace distribution, such that

$$P(q_r|q_l, d) = \frac{1}{\pi N} \exp \left( - \frac{k_2 \mathcal{C}_{\text{seg}}(u, v, d)}{N} \right)$$  \hspace{1cm} (13)$$

where $k_2$ is a constant. Substituting (4) and (13) into (12) and taking the negative logarithm of both sides gives

$$\log \left( \frac{1}{p(d|q_1, q_r, O)} \right) \propto \log \left( \frac{\pi N}{\pi} + \frac{k_2 \mathcal{C}_{\text{seg}}(u, v, d)}{N} \right)$$

$$- \log \left( k_1 + \frac{1}{2 \pi^2} \exp \left( - \frac{(d - d_{\text{prior}})^2}{2 \sigma^2} \right) \right)$$  \hspace{1cm} (14)$$
Fig. 7 Disparity maps calculated by different algorithms

a NLCA
b ELAS
c ANCC
d AdaptWeight
e CSBP
f Our method
The negative log likelihood is an energy function. To find the most likely solution \( d \), (14) is rewritten in the form

\[
E(d) = \log(p_N) + k_2 C_{\text{agg}}(u, v, d) - \log\left( k_1 + \frac{1}{2\pi\sigma^2} \exp\left( -\frac{(d - d_{\text{prior}})^2}{2\sigma^2} \right) \right) \tag{15}
\]

By minimising (15), we can obtain an initial disparity map. This map contains a few spurious points, some black holes and does not contain the occlusion areas. To rectify these issues and obtain a high-precision dense disparity map, post-processing of the computed disparities must be carried out.

### 3.5 Disparity refinement

The disparity map of the target image is computed in the same way as the reference image. Using the method described in [5], a left/right cross-check is performed to remove any spurious points and to estimate the disparities of the occluded regions. The dense disparity map is obtained by interpolating missing disparities using a piecewise constant function on the smallest valid neighbour in the same image line. A median filter is then applied to clean-up spurious mismatches. Once this step is completed the final dense disparity map is obtained.

<table>
<thead>
<tr>
<th></th>
<th>Tsukuba</th>
<th>Venus</th>
<th>Teddy</th>
<th>Cone</th>
<th>Average percent of bad pixels</th>
<th>Average rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLCA [4]</td>
<td>11.10%</td>
<td>8.64%</td>
<td>13.50%</td>
<td>9.40%</td>
<td>10.7%</td>
<td>58.5</td>
</tr>
<tr>
<td>ELAS [5]</td>
<td>14.90%</td>
<td>9.25%</td>
<td>18.00%</td>
<td>11.90%</td>
<td>13.5%</td>
<td>107.7</td>
</tr>
<tr>
<td>ANCC [14]</td>
<td>28.00%</td>
<td>7.77%</td>
<td>17.60%</td>
<td>14.00%</td>
<td>14.4%</td>
<td>123.8</td>
</tr>
<tr>
<td>AdaptWeight [22]</td>
<td>18.10%</td>
<td>11.50%</td>
<td>20.80%</td>
<td>14.50%</td>
<td>18.60%</td>
<td>90.8</td>
</tr>
<tr>
<td>CSBP [24]</td>
<td>22.00%</td>
<td>7.60%</td>
<td>19.40%</td>
<td>15.10%</td>
<td>16.0%</td>
<td>113.2</td>
</tr>
<tr>
<td>our method</td>
<td>29.50%</td>
<td>5.32%</td>
<td>11.00%</td>
<td>8.18%</td>
<td>13.5%</td>
<td>64.6</td>
</tr>
<tr>
<td>NLCA [4]</td>
<td>1.47%</td>
<td>0.25%</td>
<td>6.01%</td>
<td>2.87%</td>
<td>2.65%</td>
<td>45.7</td>
</tr>
<tr>
<td>ELAS [5]</td>
<td>5.70%</td>
<td>2.85%</td>
<td>12.2%</td>
<td>7.41%</td>
<td>7.07%</td>
<td>134.9</td>
</tr>
<tr>
<td>ANCC [14]</td>
<td>3.28%</td>
<td>3.46%</td>
<td>11.60%</td>
<td>5.53%</td>
<td>5.97%</td>
<td>120.3</td>
</tr>
<tr>
<td>AdaptWeight [22]</td>
<td>1.38%</td>
<td>0.71%</td>
<td>7.88%</td>
<td>3.97%</td>
<td>3.48%</td>
<td>75.2</td>
</tr>
<tr>
<td>CSBP [24]</td>
<td>2.00%</td>
<td>1.48%</td>
<td>11.1%</td>
<td>5.98%</td>
<td>5.15%</td>
<td>121.1</td>
</tr>
<tr>
<td>our method</td>
<td>3.32%</td>
<td>0.92%</td>
<td>6.25%</td>
<td>3.62%</td>
<td>3.53%</td>
<td>90.0</td>
</tr>
</tbody>
</table>

Fig. 8 Middlebury datasets taken under different illumination conditions (from left to right: Aloe, Baby2, Dolls and Reindeer)

a Reference images  
b Target images  
c Ground-truth disparity maps
Fig. 9  Performance of the algorithms under different illumination conditions for the images shown in Fig. 8

a  NLCA  
b  ELAS  
c  ANCC  
d  AdaptWeight  
e  CSBP  
f  Our method
4 Experimental results

The performance of our method is shown for the Middlebury datasets [30] and compared with both local and global stereo matching algorithms NLCA [4], ELAS [5], ANCC [14], AdaptWeight [22] and CSBP [24]. It is worth noting that ANCC uses the colour formation model explicitly and is robust to radiometric conditions between stereo images according to its author. Our method and ELAS are greyscale image algorithms, whereas the other four methods have been developed for use with colour images. In all the experiments the matching parameters in our method are set to $a = 3$, $k_1 = 15$, $k_2 = 0.2$ and $\sigma = 1$. The parameters of NLCA, ELAS, ANCC, AdaptWeight and CSBP are set using the values given in [4, 5, 14, 22] and [24], respectively.

4.1 Constant lights and camera exposure

To demonstrate the performance of our method in the absence of radiometric differences, our method was tested using the standard, radiometrically clean, Tsukuba, Venus, Teddy and Cone images as shown in Fig. 6. The results are shown in Fig. 7. The disparity maps obtained using NLCA, AdaptWeight and CSBP are taken from the Middlebury website [30]. The disparity map obtained using ELAS and ANCC are generated using code provided to us by the authors. Our method does not give border disparities, therefore the ground-truth disparities are used in these regions. Visual comparison shows that our method performs better near object boundaries and regions with low texture.

To quantitatively compare the performance of the algorithms, the differences between the disparity and ground-truth disparity maps are calculated to find the percentage of badly matching pixels in all non-occluded regions. The results are shown in Table 1. The data show that, in the absence of radiometric differences, the performance of our method is comparable with state-of-the-art stereo matching methods.

In addition, we have tested the runtime of the algorithms on a DELL computer with a 3.4 GHz Intel Core i7 CPU and 4 GB memory. No parallelism technique is utilised. The average runtime of the four standard Middlebury datasets (including Tsukuba, Venus, Teddy and Cones) is about 74 ms using our method, which is about 1.18 × faster than NLCA (87 ms), 1.68 × slower than ELAS (44 ms) and 7.92 × faster than ANCC (586 ms). The average runtime of CSBP was also computed using the same computer, and is 1379 ms, which is about 18.6 × slower than our method. As AdaptWeight is almost the slowest local matching algorithm in the Middlebury benchmark, there is no need to compare its average runtime with our method at all.
4.2 Light source changes

The robustness of the methods is checked using the test bed images (Aloe, Baby2, Dolls and Reindeer). The image sets used have three different exposure conditions (exp0, exp1 and exp2) and three different light source configurations (Illum0, Illum1 and Illum2). In the experiments, images with exp2 conditions are used. The light source configurations of the reference and target images are Illum0 and Illum2, respectively. Fig. 8 shows the stereo image pairs and their corresponding ground-truth disparity maps.

The disparity maps obtained for the different stereo matching algorithms are shown in Fig. 9. For NLCA, considerable noise is seen in the resulting disparity maps (Fig. 9a). ELAS does not perform well under light source changes (Fig. 9b). Although ANCC is designed for a robust and accurate correspondence, it cannot handle the multiple illumination conditions well (Fig. 9c). As can be seen in Figs. 9d and e, both AdaptWeight and CSBP are very sensitive to light source changes between left and right images. The results show that our method significantly outperforms the others (Fig. 9f).

Furthermore, for a quantitative comparison, Fig. 10 presents the errors of the estimated disparities of unoccluded pixels under various illumination conditions. The errors are computed against the ground-truth disparity maps in Fig. 8c. It can be seen that our method is very stable and low error rate, whereas other algorithms are very sensitive to the illumination changes. These results indicate that our method can reduce the effect of different lighting conditions efficiently.

4.3 Camera exposure changes

To test the robustness of the algorithms to changes in the camera exposure, the light source illumination conditions are set to Illum1 and the exposure conditions are changed. The exposure conditions of the reference and target images are exp1 and exp2, respectively. The stereo image pairs and their corresponding ground-truth disparity maps are shown in Fig. 11.

Fig. 12 shows the resulting disparity maps obtained for the images in Fig. 11 using the six different algorithms. As can be seen in Fig. 12a, NLCA produced some noise in the resulting disparity map. No noise is observed in the disparity maps obtained using ELAS (Fig. 12b). ANCC is more robust to camera exposure changes than the first two methods (Fig. 12c). The performance of AdaptWeight and CSBP, however, are sensitive to changes in the camera exposure conditions (Fig. 12d and e). For all of the stereo pair images, our method performs consistently with no sensitivity to exposure being observed (Fig. 12f).

To quantitatively compare the performance of the test stereo algorithms, we fixed the illumination to Illum1 and varied only the exposure from exp0 to exp2. Fig. 13 presents the errors of the estimated disparities of unoccluded pixels under various exposure conditions. As preceding introduction, the errors that are computed against the ground-truth disparity maps are shown as Fig. 11c. It can be seen that our method yields more stable performance with different exposure conditions than other stereo algorithms. Therefore it is safe to say that our method can reduce the effect of different exposure conditions efficiently.
Fig. 12 Performance of the algorithms under different exposure conditions for the images shown in Fig. 11

a NLCA
b ELAS
c ANCC
d AdaptWeight
e CSBP
f Our method
5 Conclusions

We have developed and demonstrated a local stereo matching algorithm that is insensitive to changes in radiometric conditions between the input images. In our method, a prior on the disparities is first built by combining the DAISY descriptor and Census filtering. A Census-based cost aggregation with a self-adaptive window strategy is then performed. The maximum a-posteriori estimation is then adopted to compute the final disparity. Our method can be used to obtain dense, high-accuracy disparity maps under different radiometric conditions, including varying light illumination and camera exposure. Experiments on the Middlebury benchmark dataset images show that the performance of our method is comparable with the performance of state-of-the-art stereo matching algorithms on radiometrically clean images. We demonstrate that, for input images with different radiometric conditions, the performance of our algorithm is substantially better. In future experiments, real scene images taken with different radiometric conditions will be used to test the performance of our stereo matching algorithm. To make our method more robust to noise, we plan to implement other cost aggregation strategies.

6 Acknowledgments

This work was supported by the National Natural Science Foundation of China Grant No. 51105027. We thank Prof. D. Scharstein from Middlebury College, and Dr. R. Szeliski from Microsoft Research for providing the images with ground truth. We also thank A. Geiger from Karlsruhe Institute of Technology, Q. Yang from City University of Hong Kong and Y.S. Heo from Seoul National University for providing the computer codes used in this paper.

7 References