Disparity Search Range Estimation Based on Dense Stereo Matching

Rueihung Li, Bumsub Ham, Changjae Oh, and Kwanghoon Sohn
School of Electrical and Electronic Engineering
Yonsei University
Seoul, Korea
Email: khsohn@yonsei.ac.kr

Abstract—This paper presents a scheme for estimating disparity range based on hierarchical stereo matching. It is important to specify a proper range of search space, since it prevents the solution from being trapped in local minima, and saves a lot of time for estimating disparity maps. Conventionally, it has been estimated by finding a sparse set of the correspondences via feature matching techniques. Instead, we address this problem by considering how the dense correspondences, the ultimate goal of estimating search range, can be estimated without search range. First, we estimate the dense correspondences by adapting a simple local stereo matching technique with an arbitrary search range. The hierarchical scheme is leveraged for reducing computational costs and memory usages. Then, reliable checking techniques are performed for eliminating unreliable correspondences. Finally, the range of search space is estimated by observing the distribution of the reliable correspondences. For the quantitative evaluation, a new error metric, biased root mean squared error (B-RMSE), is proposed, which differentiates the disparity candidates in image sequences. It means that the experimental results show that the proposed method gives more accurate search range compared to the conventional method.

I. INTRODUCTION

In the field of three-dimensional (3-D) computer vision, depth information is useful in various applications such as 3-D scene reconstruction, virtual view rendering, and object tracking. Since the performance of the applications depends on the quality of depth information, it is important to obtain accurate depth information given stereo and/or multiview sequences. Many researchers have made an effort to estimate reliable depth information, one of which is to estimate the correspondence, i.e., disparity, between image sequences [1].

Disparity search range is defined as the space for all disparity candidates in image sequences. It means that the inaccurate search range (e.g., invalid disparity candidates) influences the performance of the disparity estimation in terms of the complexity and the accuracy. Generally, the disparity has been estimated with an assumption that an appropriate search range is already known. Although most datasets provide this parameter [10], in practice, any image sequences do not provide the information about the search range, which indicates that the search range should be measured manually on each sequence.

One alternative is to automatically estimate the search range. Cyganek and Borgosz estimated the maximum disparity by a statistical analysis, but only positive disparity can be calculated [2]. Kostikova and Sara proposed an automatic search range estimation algorithm [3] based on a confidently stable matching (CSM) [4], which gives a sparse matching result. In [5], Min et al. determined the search range based on a feature matching algorithm such as the speeded up robust features (SURF) [6]. It enforces temporal consistency between nearby frames for enhancing an accuracy of the results. The CSM and SURF fail in finding the correspondence at homogeneous region. That is, the feature-based approach is not appropriate to estimate disparity search range since homogeneous region can have the biggest or the smallest disparity which cannot be detected by the approach.

Instead, we consider that how we can get dense correspondences without an exact search range, such that the disparity search range can be measured from the estimated disparity map. Furthermore, it can be used to estimate a refined disparity map in a recursive manner. For this, we leverage a simple local stereo matching based on the sum-of-absolute-difference (SAD) and the winner-takes-all (WTA) method with the hierarchical scheme for reducing the complexity and enhancing memory efficiency.

This paper is organized as follows. In section II, we give the detailed explanation of proposed method. In section III, we present a new error metric, biased root mean squared error (B-RMSE), to evaluate the performance of an estimated search range. The experimental results are presented in section IV. Finally, we conclude the paper in section V.

II. THE PROPOSED SEARCH RANGE ESTIMATION METHOD

Fig. 1 shows the overall framework of the proposed method: the stereo matching is performed followed by observing the distribution of the matching result for estimating search range, which is done in a recursive manner. The disparity is hierarchically estimated for reducing the complexity and memory usage. Although the simple local stereo matching is leveraged for reducing the complexity, it generally gives an inaccurate disparity, which prevents from detecting an accurate search range. Thus, disparity reliability checking techniques are performed. It is worthy of noting that the stereo matching also needs a search range in initial for setting bounds for a possible search space. Thus, an arbitrary search range is set initially which should be wider enough than expected one.

A. Hierarchical Stereo Matching

The Gaussian pyramid is used for sub-sampling the stereo sequences to ‘N’ levels with the sampling ratio between
consecutive levels being 0.5: the lowest level (1-level) has the original image size, whereas the highest level (N-level) has the smallest image size. From the highest level to the lowest level, the stereo matching is performed at each level. The number of pyramids $N$ is set according to the original image size, i.e., we empirically found that a proper width of the image at the highest level is in 100~200 pixels for accuracy and speed. Thus, the image size is decreased by half until the width of the original image is lower than 200 pixels.

Let $R_n$ be the search range of n-level. The estimated disparity information is then used to refine search range $R_N$. In the next level, the search range widen from $R_N$ to $R_{N-1}$ which used as the probable search range for (N-1)-level. For robustness, the marginal range $\alpha$ is added. For example, assume that $R_N$ is refined as $[a, b]$, then $R_{N-1}$ is set to $[2a - \alpha, 2b + \alpha]$.

These procedures are performed recursively until reaching the 1-level and estimating $R_1$. Note that the initial search range $R_N$ at N-level is not estimated, but just set to some fixed value, since we empirically found that the maximum/minimum disparity of the most stereo sequences does not exceed twenty per cent of an image width. This coarse-to-fine process provides following two advantages: (1) the stereo matching on the sub-sampled sequences is faster than on the original sequences, (2) the range refined in previous level makes stereo matching faster in next level.

B. Disparity Reliability checking

Since the outliers prevent from detecting accurate search range, disparity reliability checking techniques are performed to eliminate these outliers: (1) cross-checking [7], [8] and (2) matching confidence [9].

The cross-checking is well-known method for detecting occluded region. In the occluded region, the pixels in one image do not have corresponding pixels in another image, resulting in inaccurate depth information. The cross checking discriminates occluded regions from visible regions by checking whether a disparity in one image is same as that of the corresponding point in another image. Let $(i, j)$ be a 2D spatial coordinate. The occluded region $r^C_L$ and $r^C_R$ can be found as follows:

$$r^C_L(i, j) = \begin{cases} 0, & D_L(i, j) = D_L(i - D_L(i, j), j) \\ 1, & \text{otherwise} \end{cases}$$

$$r^C_R(i, j) = \begin{cases} 0, & D_R(i, j) = D_R(i + D_R(i), j) \\ 1, & \text{otherwise} \end{cases}$$

where $D_L(i, j)$ and $D_R(i, j)$ represent the disparity of the pixel of $(i, j)$ in the left and right images, respectively. The $r^C_L(i, j)$ and $r^C_R(i, j)$ are set ‘1’ if the pixel of $(i, j)$ might be in the occlusion region, i.e., the disparity is unreliable.

Matching confidence denotes the distinguishability of the minimum matching cost of a pixel. It is assumed that the minimum cost value would be not far from the cost value of other disparity candidates if a selected disparity is unreliable, i.e., the proportion of the minimum cost to the second minimum cost has higher proportion. The matching confidence of pixel of $(i, j)$, $w(i, j)$ can be defined as follows:

$$w(i, j) = 1 - \frac{c_{i,j}^{2\text{nd}}}{c_{i,j}^{1\text{st}}}$$

where $c_{i,j}^{1\text{st}}$ and $c_{i,j}^{2\text{nd}}$ represent first and second minimum matching cost at pixel of $(i, j)$ respectively. $w(i, j) \in [0, 1]$ is always satisfied since $c_{i,j}^{2\text{nd}}$ is always bigger than $c_{i,j}^{1\text{st}}$. The high $w(i, j)$ indicates that the disparity at the pixel of $(i, j)$ is reliable and vice versa. Thus, the disparities can be classified into two cases, “reliable” or “unreliable” as follows.

$$r^R(i, j) = \begin{cases} 0, & w(i, j) \geq T \\ 1, & \text{otherwise} \end{cases}$$

where $T \in [0, 1]$ is a threshold value. The final reliability checking for left and right images is then defined as follows:

$$r^R_L(i, j) = r^C_L(i, j) \mid r^R_L(i, j),$$

$$r^R_R(i, j) = r^C_R(i, j) \mid r^R_R(i, j)$$
where \( r^L(i, j) \) and \( r^R(i, j) \) are matching confidence map for left and right images, respectively. ‘||’ denotes the logical OR operator of which result is ‘1’ if either or both operand is ‘1’, otherwise, ‘0’. The search range is then refined by using \( r^L(i, j) \) and \( r^R(i, j) \) as follows:

\[
R = R_L \cup R_R
\]

where,

\[
R_L = [\min(D_L(i, j)), \max(D_L(i, j))] \quad \text{for } \forall r^L(i, j)
\]

\[
R_R = [\min(D_R(i, j)), \max(D_R(i, j))] \quad \text{for } \forall r^R(i, j)
\]

Fig. 2 shows the result of disparity reliable checking. We can see that the final reliability checking results are similar to the matching confidence maps, since it include the most region of the cross checking result. Nevertheless, the cross checking is still needed since matching confidence map cannot cover the whole occluded region which have inaccurate disparity. Table I shows the estimated search range for stereo images available at [10], by reliability checking differently: using cross checking only, using matching confidence only, and using both of them.

The detailed overall process of proposed disparity search range estimation method is described in Table II.

### III. Biased Root Mean Squared Error

The error of estimated search range can be classified into two cases whether the search range is wider or narrower than the true range. Although both cases cause severe influences on disparity estimation, the latter case is more crucial, since the true disparity cannot be detected in the estimated range. Fig. 3 shows estimated disparity maps for the 'Tsukuba' image [10] and corresponding bad matching maps obtained by setting different search ranges. It shows that the result with a narrow search range lost most depth information (Fig. 3(b)), although the number of errors in search range is same, i.e., the ground truth range is [5, 14], and the errors in the search range between the Fig. (a) and (c) has same quantity. The bad matching percentage of the disparity maps in Fig. (d)-(f) are 21.7%, 2.88%, and 4.40%, respectively [10].

However, the conventional root mean squared error (RMSE) cannot differentiate these cases, i.e., the RMSE do not consider whether the estimated search is narrower or wider than true disparity range. Thus, a new evaluation metric, the B-RMSE, is proposed as follows.

\[
e = \frac{\sum_{m=1}^{M} \sqrt{e_{\min}[m] + e_{\max}[m]}}{M}
\]

where \( M \) is the number of frames in stereo sequences. The squared errors of m-th frame, \( e_{\min}[m] \) and \( e_{\max}[m] \), are determined as follows:

### TABLE I. Search Range Estimated by Reliability Checking Differently

<table>
<thead>
<tr>
<th></th>
<th>Tsukuba</th>
<th>Venus</th>
<th>Cone</th>
<th>Teddys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>min</td>
<td>max</td>
<td>min</td>
<td>max</td>
</tr>
<tr>
<td>Rgal</td>
<td>4</td>
<td>14</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>CC only</td>
<td>-4</td>
<td>17</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>MC only</td>
<td>-78</td>
<td>16</td>
<td>-85</td>
<td>20</td>
</tr>
<tr>
<td>CC &amp; MC</td>
<td>-4</td>
<td>16</td>
<td>2</td>
<td>20</td>
</tr>
</tbody>
</table>

where \( \beta \) denotes a weight parameter which should be larger than 1. \( R_{\text{est}}^\text{min}[m] \) and \( R_{\text{true}}^\text{min}[m] \) represent the minimum value of estimated and true search range in \( m \)-th frame, respectively. Similarly, \( R_{\text{est}}^\text{max}[m] \) and \( R_{\text{true}}^\text{max}[m] \) represent the maximum value of estimated and true search range of \( m \)-th frame respectively. The B-RMSE gives more weight if the estimated range is within the true disparity range.

### IV. Experimental Results

In order to verify the performance, the proposed scheme was compared with the conventional feature-based method [5] with the stereo images [10] and video [11] were tested. All parameter was fixed in the experiments \(( \alpha=2, T=0.2)\).

The search range estimation results with image sequence as shown in Fig. 4 are shown in Table III. The average error shows that the proposed method provides more accurate result than the conventional method [5]. The average error inside the true range indicates that the estimated search range is in the true disparity range, which should be prevented. As expected, the feature-based method [5] shows a poor performance especially at featureless images such as ‘Bowling1’, ‘Lampshade1’, and ‘Plastic’, the method [5]. In contrast, the proposed method provides more reliable result for these sequences.

The stereo video sequences for experiment and SURF feature matching results of them are shown in Fig. 5. In the figure, we also marked the regions which have max and min disparity. We can see that the features are rare in the foreground and background respectively at ‘Skull Rock’ and ‘Knight’ sequences. In ‘Heidelberg’ sequences, the features of suddenly appeared object is not detected well in some frames, which cause the increment in the error of estimated search range in feature-based method [5]. Fig. 6 shows the estimated disparity search ranges for the stereo video sequences, obtained by the conventional method [5] and the proposed method.
### TABLE III. SEARCH RANGE ESTIMATION RESULTS OF STEREO IMAGE SETS

<table>
<thead>
<tr>
<th>Original Range</th>
<th>Conventional Method</th>
<th>Proposed Method</th>
<th>Conventional Method</th>
<th>Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>max</td>
<td>min</td>
<td>max</td>
<td>min</td>
<td># of err</td>
</tr>
<tr>
<td>Baby1</td>
<td>68</td>
<td>26</td>
<td>66</td>
<td>25</td>
</tr>
<tr>
<td>Bowling1</td>
<td>115</td>
<td>6</td>
<td>80</td>
<td>4</td>
</tr>
<tr>
<td>Cloth1</td>
<td>84</td>
<td>20</td>
<td>80</td>
<td>18</td>
</tr>
<tr>
<td>Cone</td>
<td>55</td>
<td>16</td>
<td>52</td>
<td>18</td>
</tr>
<tr>
<td>Lampshade1</td>
<td>97</td>
<td>26</td>
<td>66</td>
<td>46</td>
</tr>
<tr>
<td>Moebius</td>
<td>107</td>
<td>34</td>
<td>101</td>
<td>39</td>
</tr>
<tr>
<td>Plastic</td>
<td>98</td>
<td>27</td>
<td>66</td>
<td>60</td>
</tr>
<tr>
<td>Tsukuba</td>
<td>14</td>
<td>5</td>
<td>17</td>
<td>-3</td>
</tr>
<tr>
<td>Venus</td>
<td>19</td>
<td>3</td>
<td>24</td>
<td>-3</td>
</tr>
<tr>
<td>Average</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

# of err: The number of error between the estimated search range with original range
# of insider err: The number of error only counted if it is inside of the original range

The upper bound of the search range estimated by [5] for ‘Skull Rock’ does not contain most true disparity candidates. For ‘Heidelberg’, the conventional method [5] cannot estimate the disparity candidates when the object suddenly appears, which can be observed between 10 and 15 frames. In this experiment, it is shown that the lack of features in foreground and/or background degrades the performance of feature-based method. On the other hand, the proposed scheme gives more proper search range for the stereo video sequences.

The comparison of RMSE and B-RMSE are summarized in the Table IV. B-RMSE is calculated with constant weight $\beta = 2$. It shows that the search range estimated by proposed method is more similar to the true disparity range than one obtained by [5]. Note that the difference between RMSE and B-RMSE is high when the estimated search range is narrower than true disparity range. By observing the difference between RMSE and B-RMSE, most true search range can be covered by the proposed method. That is, the proposed method preserves the true disparity candidates better than the conventional method [5].

### TABLE IV. RESULT OF RMSE AND B-RMSE

<table>
<thead>
<tr>
<th></th>
<th>Conventional</th>
<th>Proposed</th>
<th>B-RMSE</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skull Rock</td>
<td>15.7</td>
<td>3.17</td>
<td>22.14</td>
<td>3.25</td>
</tr>
<tr>
<td>Knight</td>
<td>9.53</td>
<td>6.03</td>
<td>12.55</td>
<td>6.11</td>
</tr>
<tr>
<td>Heidelberg</td>
<td>13.04</td>
<td>9.34</td>
<td>21.52</td>
<td>9.04</td>
</tr>
</tbody>
</table>

### V. CONCLUSION

In this paper, we have presented a scheme for estimating search range. Conventional feature-based methods give inappropriate results for stereo sequences which include homogeneous region in foreground or background. To solve this problem, we leveraged hierarchical stereo matching and disparity reliability checking techniques. The proposed method shows high performance regardless of homogeneity of stereo sequences. In order to evaluate the performance of search range quantitative, B-RMSE was proposed, which differently considered the estimated search range whether it is narrower or wider than true disparity range. The experimental results confirmed the performance of proposed method by comparing with conventional feature-based method. In the future, we will further accelerate proposed method by considering the per-pixel search range, which drastically reduces the complexity of the conventional stereo matching techniques.

---

Fig. 5. Test stereo video sequences available at [10]. Top: original images, Bottom: SURF feature matching results and region of min/max disparity

The Region which includes max disparity.

The Region which includes min disparity.

---

2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA) 757
Fig. 6. Search range estimation result for Test stereo video sequences.
REFERENCES


