Stereo Matching with the Distinctive Similarity Measure

Kuk-Jin Yoon
Perception Team
INRIA Rhône-Alpes, France
Kuk-Jin.Yoon@inrialpes.fr

In So Kweon
Robotics and Computer Vision Lab.
KAIST, Korea
iskweon@ee.kaist.ac.kr

Abstract

The point ambiguity owing to the ambiguous local appearances of image points is one of the main causes making the stereo problem difficult. Under the point ambiguity, local similarity measures are easy to be ambiguous and this results in false matches in ambiguous regions.

In this paper, we present the new similarity measure to resolve the point ambiguity problem based on the idea that the distinctiveness, not the interest, is the appropriate criterion for the feature selection under the point ambiguity. The proposed similarity measure named the Distinctive Similarity Measure (DSM) is essentially based on the distinctiveness of image points and the dissimilarity between them, which are both closely related to the local appearances of image points; the distinctiveness of an image point is related to the probability of a mismatch while the dissimilarity is related to the probability of a good match. We verify the efficiency of the proposed DSM by using testbed image sets. Experimental results show that the proposed DSM is very effective and can be easily used for improving the performance of existing stereo methods under the point ambiguity.

1. Introduction

Stereo vision has been a long lasting research topic in computer vision. Here, as pointed out in [1], the crux of stereo matching problem is how to deal with the inherent point ambiguity that results from the ambiguous local appearances of image points. If the local structures of neighboring image points are quite similar as in textureless or repetitive-textured regions, it may be very difficult to find their correspondences in other images without any proper global reasoning.

To resolve the point ambiguity problem in stereo matching, many methods have been proposed for decades. Feature-based methods match only a few points proper for matching [7, 12, 13, 15] while filtering out ambiguous points. In general, they first extract local features such as corners, edges, and line segments from input images, of which the neighborhood contains enough information to facilitate matching. Those features are then matched using some appropriate local feature descriptors. As a result, feature-based methods yield sparse disparity maps. This approach is comparatively robust to the point ambiguity and produces accurate results rapidly in general. However, sparse disparity maps are not sufficient in many applications and some local features can be also ambiguous in some regions.

On the other hand, area-based methods try to yield a dense disparity map while handling the point ambiguity locally or globally [17]. Area-based local methods [5, 8, 9, 22, 24] typically use some kinds of statistical correlation among color or intensity patterns in local support windows to deal with the ambiguity. In this approach, it is implicitly assumed that all points in a support window are from the same depth in the scene. However, support windows that are located on depth discontinuities represent points at different depth, and this results in the foreground fattening phenomenon. Moreover, because this approach generally selects the disparities of image points locally using the winner-takes-all (WTA) method, most local methods still have difficulty in dealing with the point ambiguity owing to insufficient or repetitive texture. Therefore, an appropriate support window should be selected for each point to obtain more accurate results under the point ambiguity.

Unlike the local methods, global methods seek a disparity surface minimizing a global cost function defined by making an explicit smoothness assumption. In this approach, it is assumed that disparity values vary smoothly in an image and, therefore, ambiguous points get assigned disparity values inferred by propagating the disparity values of neighboring points. Some global stereo methods have recently achieved good results by modeling a disparity surface as a Markov random field to deal with the point ambiguity. For this end, various optimization techniques such as the cooperative algorithm [15, 25], dynamic programming [3], non-linear diffusion [16], graph cut [4, 9, 11], and belief propagation [18, 19] are used. Therefore, global
methods deal with the inherent image ambiguity more effectively than local methods. However, they mainly focus on how to efficiently minimize the conventional cost in spite of the fact that lower cost solutions do not always correspond to better performance as pointed out in [20]. Therefore, it is more important to properly measure the point (dis-)similarity used in cost functions. Nevertheless, there is a relatively small amount of work on it.

In this paper, we present a new similarity measure to resolve the point ambiguity problem in the stereo matching, inspired by the work of Manduci and Tomasi [14]. They showed that the distinctiveness, not the interest, is the appropriate criterion for the feature selection. For instance, interesting points such as intensity edges and corners can be also highly ambiguous when they are in repetitive, periodic textured regions.

The proposed similarity measure named the Distinctive Similarity Measure (DSM) is essentially based on the distinctiveness of image points and the dissimilarity between them, which are both closely related to the local appearances of image points; the distinctiveness of an image point is related to the probability of a mismatch while the dissimilarity is related to the probability of a good match. Therefore, it is possible to get the more reliable similarity measure under the inherent point ambiguity by properly combining the distinctiveness and the dissimilarity. Here, it is also worthy of notice that we take both the reference image and the target image into account in defining the similarity measure, aiming at improving performance without modeling half-occluded points explicitly and/or using color segmentation, which are also difficult problems. We verify the efficiency of the proposed DSM using testbed image sets.

The rest of the paper is organized as follows. We first present an ideal model for the similarity measure in the sense of distinctiveness in Section 2 and propose the simple DSM in Section 3. We then present simple local methods using the DSM in Section 4. Experimental results are shown in Section 5 and we finally conclude the work in Section 6.

2. Similarity Measure Model

Suppose that two sets of image points, $\mathbb{I}$, $\mathbb{R}$, are given.\footnote{In this paper, we consider the similarity measure for only two points. However, it is straightforward to generalize it for the $N$-point similarity measure.} Basically, the similarity measure between two points $p_i \in \mathbb{I}$ and $p_r \in \mathbb{R}$, $S(p_i, p_r)$, should be in proportion to the probability of the event $O_{p_i}$\footnote{Note that $O_{p_i}^p = O_{p_r}^p$.} saying that "$p_i \Rightarrow p_r$ is true" as

\[
S(p_i, p_r) \propto P(O_{p_i}^p | (A_{p_i}, A_{p_r})) ,
\]

where $A_{p_i}$ and $A_{p_r}$ are the image appearances of $p_i$ and $p_r$ respectively. By the Bayes' rule, we get

\[
P(O_{p_i}^p | (A_{p_i}, A_{p_r})) = \frac{P((A_{p_i}, A_{p_r}) | O_{p_i}^p) P(O_{p_i}^p)}{P(A_{p_i}, A_{p_r})} ,
\]

where $P((A_{p_i}, A_{p_r}) | O_{p_i}^p)$ is a likelihood related to the matching cost for $O_{p_i}^p$ and $P(O_{p_i}^p)$ is a prior on the event $O_{p_i}^p$. Here, $P(O_{p_i}^p)$ is concerned with the prior on the disparity field such as the smoothness assumption in global methods, while it is also related to the interest of image points in local methods. For example, feature-based local methods extract interesting features having large $P(O_{p_i}^p)$ to reduce false matches. On the other hand, $P(A_{p_i}, A_{p_r})$ is a prior on image appearances $A_{p_i}$ and $A_{p_r}$. When assuming two events $A_{p_i}$ and $A_{p_r}$ are independent, it can be expressed as $P(A_{p_i}, A_{p_r}) = P(A_{p_i}) P(A_{p_r})$. While $P(O_{p_i}^p)$ is related to the interest of image points, $P(A_{p_i}, A_{p_r})$ is related to the distinctiveness of image points. If a point $p$ is distinctive, then $p$ has small $P(A_{p_i})$, and vice versa.

Most local methods focus on the $P((A_{p_i}, A_{p_r}) | O_{p_i}^p)$ and $P(O_{p_i}^p)$ but pay little attention on $P(A_{p_i}, A_{p_r})$. In this work, we focus on $P(A_{p_i}, A_{p_r})$ not $P(O_{p_i}^p)$ to reduce false matches owing to the point ambiguity. In this case, we have the following model from Eq. (2).

\[
S(p_i, p_r) \propto P(O_{p_i}^p | (A_{p_i}, A_{p_r})) \propto \frac{P((A_{p_i}, A_{p_r}) | O_{p_i}^p)}{P(A_{p_i}) P(A_{p_r})} .
\]

This model is analogous to the ratio of the between-class correlation and the within-class correlation, which is commonly used in pattern recognition [10]; $P((A_{p_i}, A_{p_r}) | O_{p_i}^p)$ is corresponding to the between-class correlation and $P(A_{p_i})$ and $P(A_{p_r})$ are corresponding to the within-class correlation.

3. Distinctive Similarity Measure

There may be many ways to define similarity measures satisfying Eq. (3). Here, we propose a new simple similarity measure named the Distinctive Similarity Measure (DSM), which has a specific form satisfying the condition shown in Eq. (3). The proposed DSM is essentially based on both the image appearances of points ($P(A_{p_i})$ and $P(A_{p_r})$) themselves and their matching cost $P(O_{p_i}^p | (A_{p_i}, A_{p_r}))$.

3.1. Point distinctiveness

In defining the DSM, the distinctiveness of an image point is considered for $P(A_{p_i})$. The distinctiveness of an image point is defined in [14] as the difference with the most similar other point in a given search region. When some points in a given search region look similar with the point of interest, we can say that the point of interest is ambiguous
and its distinctiveness is small. Therefore, the distinctiveness is the appropriate criterion for the selection of reliable match candidates. The distinctiveness of an image point can be formally defined as Definition 1.

**Definition 1. Distinctiveness of a point:** The distinctiveness of an image point \( p \), \( D(p) \), is defined as

\[
D(p) = \min_{q \in W_p, q \neq p} f(p, q),
\]

where \( f(p, q) \) denotes the difference between the appearances of two image points \( p \) and \( q \). \( \square \)

Here, \( W_p \) is a set of points that can possibly correspond to the correspondence of \( p \) in the other image. When images are rectified so that correspondences are restricted to lie along a scanline, and the maximum and the minimum disparity values are \( d_{\text{max}} \) and \( d_{\text{min}} \), respectively, \( W_p \) is then a set of points like

\[
W_p = \{ p + d | d_{\text{min}} \leq d \leq d_{\text{max}} - d_{\text{min}} \}
\]

as shown in Fig. 1, where \( p + d \) is the point with coordinates of \( p \) shifted by \( d \) in the same image.

### 3.2. Distinctive similarity measure

In this paper, to define the DSM, we simply set by using \( f \) and the point distinctiveness as

\[
P(A_p) \propto \frac{1}{D(p)},
\]

\[
P((A_{p1}, A_{p2}) | O_{p1}^{p2}) \propto \frac{1}{f(p_1, p_r)},
\]

and this leads

\[
S(p_1, p_r) \propto \frac{P((A_{p1}, A_{p2}) | O_{p1}^{p2})}{P(A_{p1})P(A_{p2})} \propto \frac{D(p_1) \times D(p_r)}{f(p_1, p_r)}
\]

As a result, we define the DSM of a match candidate (a pair of image points) as follow.

**Definition 2. Distinct Similarity Measure:** The Distinct Similarity Measure (DSM) of a match candidate, \((p_1, p_r)\), is defined as

\[
S(p_1, p_r) = \frac{D(p_1) \times D(p_r)}{f(p_1, p_r)}.
\]

which is analogous to the model shown in Eq. (3). \( \square \)

The proposed DSM can be thought as a simple example among all possible similarity measures satisfying Eq. (3). However, the definition of the DSM is very intuitive and reasonable from the stereo matching point of view. The more similar two points are to each other and the more distinctive they are in images, the larger probability of a correct match they have. In other words, when two points (i.e., the appearances of two points) look alike and they are distinctive in each image, they are likely to be a correct match pair. On the other hand, even two points look alike, they may be an incorrect match pair if their appearances are ambiguous in each image as points in textureless or repeatitive-textured regions. In addition, even if two points are distinctive, they may be an incorrect match pair if their matching cost is large as points in half-occluded regions. Therefore, the DSM can be easily used to select correct matches among all possible match candidates robust against the image ambiguity.

Here, the proposed DSM can be generalized for the \( N \)-view matching problem as

\[
S(p_1, p_2, \ldots, p_N) = \frac{\prod_{i=1}^{N} D(p_i)}{g(p_1, p_2, \ldots, p_N)},
\]

where \( g(p_1, p_2, \ldots, p_N) \) represents the matching cost when \((p_1, p_2, \ldots, p_N)\) are matched.

### 3.3. Dissimilarity measure for the DSM

As shown in Definition 2, the proposed DSM is at once a new similarity measure and the function of the dissimilarity measure \( f \). Therefore, to get the reliable DSM value, it is important to accurately measure the dissimilarity between two image points. We can use any statistical correlation measures using color or intensity patterns in local support windows, such as the sum of absolute differences (SAD), the sum of squared differences (SSD), the normalized cross correlation (NCC), and the adaptive-weight dissimilarity measure [24]. Here, the dissimilarity measure should be robust against image noise and work well near arbitrary-shaped depth discontinuities. Figure 2 shows the point distinctiveness map of the ‘Tsukuba’ image when using the adaptive-weight dissimilarity measure. The brighter a point is, the larger distinctiveness a point has. We can see that the point distinctiveness is highly correlated with the point ambiguity as we expected.

### 4. Stereo Matching with the DSM

Since the DSM is a good similarity measure in itself, it can be used with a simple WTA method. In addition, it can be combined easily with existent stereo matching methods. In this paper, we present simple area-based local stereo methods that use the proposed DSM, which are not new but proper to verify the effectiveness of the proposed DSM.
4.1. Semi-dense local stereo matching

Due to the nature of area-based local methods, the accuracy of area-based local methods is generally worse than that of feature-based methods, while the resulting information of area-based local methods is much richer than that of feature-based methods. To retain the accuracy of feature-based methods while matching as many pixels as possible, Veksler [21] presented a semi-dense algorithm using dense features that are a connected set of pixels in the reference image and a corresponding set of pixels in the target image. However, the dense feature extraction is more or less complicated.

The semi-dense matching can be achieved simply and efficiently by using the DSM. Unlike previous feature-based methods, we do not need to find good points to match in the pre-processing stage, but reliable feature extraction and semi-dense matching are accomplished simultaneously in the matching stage. The semi-dense matching is achieved by finding match candidates having large DSM values with the WTA-based relaxation method — the DSM values of all possible match candidates are computed first and good matches having DSM values larger than the pre-determined threshold are then selected. A large threshold will produce accurate results while a small threshold will result in dense results. The semi-dense method using the DSM is given in Algorithm 1. \( M(p_l) \) represents the correspondence of a pixel \( p_l \) and \( D_{lh} \) is a threshold value used for selecting good match candidates. As shown in Algorithm 1, the proposed method is very simple and can be done rapidly.

4.2. Dense local stereo matching

It is possible to get accurate dense matching results by using simple WTA based methods with the proposed DSM without any complicated processes. As in the semi-dense method, the DSM values of all possible match candidates \((p_l, p_r)\) are calculated first. Then, the WTA method or the WTA-based relaxation scheme can be applied to select correct matches. When using the WTA-based relaxation method, the DSM of unmatched points are updated. Any complicated pre-/post-processes to select correct matches or to minimize some kinds of cost functions are not needed because the proposed DSM provides the large discriminative power for match candidates by itself. Although correspondences are determined locally, we can get accurate results even in ambiguous image regions. The simple dense methods using the DSM are also given in Algorithm 1.

5. Experimental Results

We verified the efficiency of the proposed DSM by using testbed images with ground truth, which are often used for performance comparison of various methods as in [17]. For computing dissimilarity between points, we adopted the adaptive-weight dissimilarity measure proposed in [24], which provides large discriminative power and also works very well at depth discontinuities. We used the same parameters ((35 x 35) local windows, \( \gamma_c = 5, \gamma_p = 17.5 \)) as in [24] to fairly compare the performance. Here, the total execution time is dominated by the dissimilarity computation because other computations are not time-consuming at all.

5.1. Semi-dense local stereo matching

The results of the semi-dense method using the proposed DSM are shown in Fig. 3 and the performance according to the threshold value \( D_{lh} \) for the “Tsukuba” image is summarized in Table 1. As mentioned before, the error rate and %
Table 1. Performance of the semi-dense method according to the threshold value for the ‘Tsukuba’ data set

<table>
<thead>
<tr>
<th>$D_{th}$</th>
<th>% of matched</th>
<th>% of incorrect</th>
<th>RMS error</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>100.00</td>
<td>1.16</td>
<td>0.63</td>
</tr>
<tr>
<td>20</td>
<td>99.99</td>
<td>1.15</td>
<td>0.63</td>
</tr>
<tr>
<td>30</td>
<td>98.00</td>
<td>1.12</td>
<td>0.64</td>
</tr>
<tr>
<td>40</td>
<td>89.99</td>
<td>0.97</td>
<td>0.58</td>
</tr>
<tr>
<td>50</td>
<td>78.40</td>
<td>0.70</td>
<td>0.51</td>
</tr>
<tr>
<td>60</td>
<td>68.00</td>
<td>0.44</td>
<td>0.43</td>
</tr>
<tr>
<td>70</td>
<td>58.70</td>
<td>0.30</td>
<td>0.37</td>
</tr>
<tr>
<td>80</td>
<td>49.80</td>
<td>0.25</td>
<td>0.33</td>
</tr>
<tr>
<td>90</td>
<td>41.10</td>
<td>0.18</td>
<td>0.28</td>
</tr>
<tr>
<td>100</td>
<td>34.10</td>
<td>0.16</td>
<td>0.25</td>
</tr>
</tbody>
</table>

of matched pixels decrease as $D_{th}$ increases.

Here, it is worthy of notice that resulting disparity maps provide sufficient information with low error rates while leaving ambiguous pixels unmatched automatically, and half-occluded pixels are also automatically excluded in the semi-dense matching. In addition, dense results obtained by using the small threshold values are roughly the best among the state-of-the-arts local methods. From the results, we can see that the proposed DSM is very efficient for the reliable local feature selection and matching.

5.2. Dense local stereo matching

We then applied the proposed DSM for dense stereo matching. To fairly verify the performance improvement by the proposed DSM, we did not perform any complicated optimization process, but we adopted the simplest winner-takes-all (WTA) method (method 2 in Algorithm 1) as

$$d_p = \arg \min_{d_{\in C_d}} S(p, p + d),$$  \hspace{1cm} (11)

where $d_p$ is the disparity of $p$ and $C_d = \{d_{min}, \cdots, d_{max}\}$ is the set of all possible disparities, and performed the simple left-right consistency check as in [24].

The results of dense stereo method for testbed images are given in Fig. 4. As shown in Fig. 4, the proposed method yields very accurate results for all testbed images. The performance of the proposed method for the testbed images is summarized in Table 2 to compare the performance with other state-of-the-art area-based methods. We can see that the proposed DSM really improves the performance of the adaptive method (except the ‘Cone’ image set)$^3$ and is better than any other state-of-the-art local methods. Especially, the performance in non-textured regions and half-occluded regions, which are both ambiguous regions, is much better than other state-of-the-art local methods and comparable with state-of-the-art global methods without explicitly modeling half-occluded pixels, without using color segmentation, and without any complicated, time-consuming processes. This effect can be explained by using Fig. 5. Figure 5 shows the DSM values of matched pixels in the ‘Tsukuba’ data set. As we expected, half-occluded pixels have very small DSM values as well as the pixels in homogeneous regions, even though they are self-distinctive. This demonstrates that large point distinctiveness does not guarantee the large DSM value, and also proves the effectiveness of the proposed DSM as a similarity measure under the inherent point ambiguity.

6. Conclusion

In this paper, a new and simple similarity measure named the DSM has been proposed to resolve the point ambiguity problem in stereo matching, focusing on the distinctiveness. The DSM is defined in terms of both the point distinctiveness of image points and the dissimilarity between image points. It provides large discriminative power even under

$^3$In the case of the ‘Cone’ image set, the performance improvement/degradation are marginal because the adaptive support method works very well for the ‘Cone’ image set by itself.
the inherent point ambiguity.

We verified the efficiency of the proposed DSM using testbed image sets. Experimental results showed that the proposed DSM is very effective and can be easily used for improving the performance of existing stereo methods under the point ambiguity, although it has a simple form. We expect that more effective similarity measures can be proposed based on the ideal similarity measure model.

References