A Literature Survey on Removal of Ambiguities in Stereo Images

Simi Thakur, Gianetan S. Sekhon
1,2Dept. of Computer Engineering, YCOE, GuruKashi Campus, Punjabi University
Talwandi Sabo, Punjab, India

Abstract
This paper presents the study of ambiguities presents in Stereo Images. Stereo vision (stereo means “Solid” or “three dimensional” and vision means “appearance” or “sight”) is the impression of depth that is perceived when a scene is viewed with both eyes. There are several ambiguities are there, when any algorithm matches the two images present in image set, these ambiguities are related to either noise, speed, efficiency or reliability. The study of various techniques of stereo matching is considered in this work. To resolve the ambiguity problem in stereo matching many methods also have been studied in this work. Define different ambiguities present in stereo images and comparative analysis of different algorithms is the major contribution of this survey paper.

Keywords
Stereo Image Set, Filter, Disparity Map

I. Introduction
Artificial intelligence is defined as “the art creating machine that perform functions that require intelligence when performed by people.” (Kurzweil, 1990) Or in other words we can say that “it is a field of study that seeks to explain and emulate intelligent behavior in terms of computational processes.” (Schalkoff, 1990) [12]. Pattern recognition and intelligent system is the major part of the artificial intelligence. Pattern recognition is concerned with estimating density function in a high-dimensional space and dividing the space into regions of categories and classes. Computer vision is the application of Pattern recognition; it is the combination of concepts, technique and ideas from: digital image processing, pattern recognitions, artificial intelligence, and computer graphics.

Now a day’s stereo matching is one of the most active areas in computer vision because of the accurate acquiring of the image depth that is related to many applications such as robotics, autonomous system, human computer interaction and scene reconstruction etc. Although many stereo matching algorithms have been proposed, especially for generating disparity maps or depth maps, stereovision is the one of methods that can yield depth information of the scene. It uses stereo image pair from two cameras to produce disparity maps that can be easily turn into depth maps. But still many challenging works needed to be done which caused by textureless, noise, occlusions and discontinuity. Stereo matching calculates disparity or depth with image set using pixel correspondence. Local and global are the categories in which generally any stereo matching method is divided.

There are many ambiguities problems are present while performing the stereo matching, for this purpose many methods have been proposed from decades. Early approaches use local and window-based methods, and employed a local “winner-takes-all” (WTA) strategy in depth estimation at each pixel. Later on, several global methods are proposed, which formulate the depth estimation as an energy-minimization problem. The local matching algorithm has some advantages such as low computation cost and high efficiency. However, its high error rate and low resolution makes it difficult to meet the requirement of practical applications. In the local algorithm, the disparity at a given pixel depends only on color values within a window [8]. Area based stereo matching approaches are applied that solves correspondence problem on the basis of area correlation. But this approach is completely local, to overcome this feature based approach is applied that is better and more accurate than the area based stereo matching but it can produce global matching results and thus efficiently avoid searching blindly in a wide range and computational complexity is another disadvantage for this method. Then the deficiencies of area and feature based methods are overcome by Hierarchical hybrid method. It produces better disparity maps on the edge of the scenes as well as on the occluded areas, and generates robust and dense disparity fields. The hybrid matching reduces the dependency on the quantity and quality of features so that it obtains still good results in sparse feature fields. It is computational efficient and the complexity of the HHM algorithm is acceptable [1]. The graph cut based algorithm which combines window-based local matching into a global optimization framework, that it preserves discontinuities both during the local as well as global matching phase [2]. The region based progressive algorithm is the one in which previous region-based and progressive approaches are combined together and this algorithm is among the state-of-the-art both in accuracy and efficiency.

Recently, segment-based methods have attracted attention due to their good performance. They are based on the assumption that the scene structure can be approximated by a set of non-overlapping planes in the disparity space and that each plane is coincident with at least one homogeneous color segment in the reference image. 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 good match [6] and it also verifies that this algorithm is very efficient under point ambiguity. The algorithm disparity distribution models (DDMs) is also based on segment-to-plane. The key difference from existing segment-based methods is that no global optimal is used [9]. In this a correct disparity plane assign to a segment directly. A region adaptive local stereo matching algorithm is the one which focuses on two methods that are census transform and a fast segmentation based method that is for depth discontinuity and other low texture area respectively. In this paper we are not exactly concern with converting 2-D object into 3-D, we can only compare the different algorithms that obtaining the exact depth of the object. Removal of ambiguities, it can be of any type either related to hardware or software and comparative analysis with different algorithms are key issues.

II. Related Work
Guoping Li et al. [1] presented algorithm based on the combination of feature and area-based matching, called Hierarchical Hybrid Matching Algorithm (HHM). It provides the efficient disparity
The resulting dissimilarity measure is given by:
\[
\text{disparity} = (1-\omega) \times \text{CSEG}(x, y, d) + \omega \times \text{SAD}(x, y, d)
\]

The optimal disparity to outliers. Instead of assigning a disparity value to each pixel, a score that maximizes the number of reliable correspondences. This algorithm runs as-

1. Arrange all UNMATCHED regions in descending order of confidence. Set \(A_{\text{amb}}\) empty.
2. For each region \(R\) with confidence\((R) \neq 0\)
   (a) If ambiguity\((R) \leq \lambda_{\text{amb}}\), label \(R\) as MATCHED and assign it the disparity \(d_{\text{cost}}\). Update disparity functions \(D, D1\) accordingly and repeat Step 2.
   (b) If \(R\) contains GCPs with more than one disparity, split \(R\), label new regions as UNMATCHED and repeat Step 2.
   (c) If there is an entry in \(A_{\text{amb}}\) empty or with a value larger than ambiguity\((R)\), set the value of the entry to ambiguity\((R)\).
3. If there are new regions labeled as MATCHED, repeat Step 2.
4. If there are UNMATCHED regions, set \(\lambda_{\text{amb}}\) to the maximum values in \(A_{\text{amb}}\), go to Step 1; or exit when a pre-defined condition is met in case that dense matching is not desired.

Andreas Klaus et al. [5] presented an algorithm that utilizes color segmentation on the reference image and a self-adapting matching score that maximizes the number of reliable correspondences. The scene structure is modeled by a set of planar surface patches which are estimated using a new technique that is more robust to outliers. Instead of assigning a disparity value to each pixel, a disparity plane is assigned to each segment. The optimal disparity plane labeling is approximated by applying belief propagation. The resulting dissimilarity measure is given by:
\[
C_{\text{SAD}}(x, y, d) = (1-\omega) \times \text{CSEG}(x, y, d) + \omega \times \text{SAD}(x, y, d)
\]

And
\[
C_{\text{GRAD}}(x, y, d) = \sum_{i} N(x, y) \left[ x_{I1}(i, j) - x_{I2}(i+d, j) + \sum_{(i, j)} N(y, x) \left[ y_{I1}(i, j) - y_{I2}(i+d, j) \right] \right]
\]

Where \(N(x, y)\) is a 3 x 3 surrounding window at position \((x, y)\), \(N(x, y)\) a surrounding window without the rightmost column, \(N(y, x)\) a surrounding window without the lowest row, \(Vx\) the forward gradient to the right and \(Vy\) the forward gradient to the bottom. Color images are taken into account by summing up the dissimilarity measures for all channels.

Lin Chen et al. [10] presented a Region-Adaptive stereo matching algorithm which focuses on different cost aggregation strategies for different regions that can apply by census transform method along depth discontinuities and a fast segmentation-based method for other low texture areas.

Total aggregation cost function is defined as:
\[
C_{\text{TOT}}(p, q, d) = \omega \times \frac{C_{\text{SEG}}(p, q, d)}{n(Bp) + n(Bq)} + (1 - \omega) \times \frac{C_{\text{LR}}(p, q, d)}{2}
\]

It contains two parts: segmentation part and window part. In segmentation part, we adopt both the block images of reference and target image. CSEG is defined as:
\[
C_{\text{SEG}}(p, q; d) = C_{\text{SEG}}(p; d) + C_{\text{SR}}(p; q; d)
\]

Where
\[
C_{\text{SEG}}(p; q; d) = \sum_{i \neq d} \min(\delta(p, q) \oplus d, Tr)
\]
\[
C_{\text{SR}}(p; q; d) = \sum_{i \neq d} \min(\delta(p, q) \ominus d, Tr)
\]

Bp is the segment block of p and Bq is the segment block of q. The difference in the intensity values is related to the dissimilarity between image points, which is both closely related to the local appearances of image points; the distinctive nature of an image point is related to the probability of a mismatch while the dissimilarity is related to the probability of a good match. Baris Baykant [7] presented an algorithm named region based stereo matching under this there are two techniques that are Global Error Energy Minimization by Smoothing Functions and Line Growing Based Stereo Matching, both are together used for extraction of depth information for generating 3D image by eliminating noise with the help of averaging filter to generate reliable disparity maps. Error energy is expressed as:
\[
o(i, j, d) = \frac{1}{3 \times m \times n} \sum_{x=i}^{i+n} \sum_{y=j}^{j+m} \sum_{k=1}^{3} \left( L(x, y, d, k) - R(x, y, d, k) \right)^2
\]
disparity value of the CDI (Continuous disparity interval) to fit the disparity plane and denote the plane as P.

2. A matching cost for P is calculated by

\[ C(S, P) = \frac{1}{N_s} \sum_{(x,y) \in S} C(x, y, d) \]

Where \( d \) is the disparity value defined by P at the pixel \((x, y)\) in the segment \( S \). \( N_s \) is the number of the pixels in \( S \).

3. Add GCPs corresponding to the next disparity value to the old ones, fit the disparity plane once more and calculate the matching cost of the new plane.

4. If the GCPs that have the largest disparity value of the CDI are added, go to step 5. Otherwise, repeat step 3.

5. Choose the plane that has the minimum matching cost as the disparity plane of the segment.

Zi-wei Zhou et al. [11] presents another algorithm based on graph cuts stereo matching but the difference is the previous methods are not used for the real time application, but this can increase the accelerate the matching speed. Firstly, the segments of the image are calculated using the Gaussians pyramid (Dogs) computing method, secondly calculate the disparity range of two corresponding image segments in image pair using the color feature, and every segment pair is matched based on graph cuts algorithm according to the offset of disparity range in parallel time, finally all the segments are assembled into a whole disparity image.

### III. Comparative Analysis

<table>
<thead>
<tr>
<th>Ref No.</th>
<th>Author(s)</th>
<th>Year</th>
<th>Title</th>
<th>Technique</th>
<th>Results</th>
<th>Limitation(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[2]</td>
<td>Motilal Agrawal, Larry S. Davis</td>
<td>2004</td>
<td>WINDOW-BASED, DISCONTINUITY PRESERVING STEREO</td>
<td>Graph cut based algorithm for combining window-based local stereo into a global optimization framework</td>
<td>Preserves discontinuities both during the local as well as global matching phase.</td>
<td>Computational speedup, Selection of window size, Incorporating occlusion &amp; uniqueness constraint.</td>
</tr>
<tr>
<td>[4]</td>
<td>Li Hong, George Chen</td>
<td>2004</td>
<td>SEGMENT-BASED STEREO MATCHING USING GRAPH CUTS</td>
<td>Segment-based stereo matching algorithm using graph cuts</td>
<td>Good results in difficult areas such as textureless regions, disparity Discontinuous boundaries and occluded portions.</td>
<td>Not be able to handle the situation if there are disparity boundaries appearing inside the initial color segments</td>
</tr>
<tr>
<td>[5]</td>
<td>Andreas Klaus, Mario Sormann, Konrad Karner</td>
<td>2006</td>
<td>SEGMENT-BASED STEREO MATCHING USING BELIEF PROPAGATION AND A SELF-ADAPTING DISSIMILARITY MEASURE</td>
<td>Segment based matching algorithm using BP and Self-Adapting</td>
<td>Give good results under fixed parameters and less time consuming.</td>
<td>Over segmentation is performed</td>
</tr>
</tbody>
</table>
In this paper, different techniques related to removal of ambiguities in stereo images have been studied. Each technique has its own benefits and limitations and each technique has its own application area in computer vision. Some of the algorithms are related to local methods and some are related to global ones. Some of the algorithms are very good in the state-of-the-art both in efficiency and accuracy. Other are performed very well in the difficult areas like textureless regions, disparity discontinuous boundaries and occluded portions. But still some research is required especially in segmentation quality, reduced computational time, high accuracy, textureless regions, disparity discontinuous boundaries and occluded areas.

**References**


Simi Thakur received her B.Tech degree in computer science engineering from Adesh Institute of Engineering and Technology (Faridkot) under Punjab Technical University in 2011 and pursuing M Tech. (Regular) degree in computer engineering from Yadavindra College of Engineering Punjabi University Guru Kashi Campus Talwandi Sabo (Bathinda), batch 2011-2013. Her research interests include computer vision. At present she has guided many students regarding the research study in computer vision.