An Optimized Mean Shift Filtering Technique to Image Representation Through Disparity Map for Large Scale Stereo Images

Kavitha¹ and Balakrishnan²

¹ Research and Development Centre, Bharathiar University, Coimbatore, Tamil Nadu, India 641 046, Kavitha0876@gmail.com
² Department of Computer Science Engineering IndraGanesan College of Engineering, Tiruchirappalli, Tamilnadu, India – 620012, eusuffali2016@gmail.com

Received: 4-January-2017; Accepted: 20-January-2017; Published: 22-January-2017

Abstract: Density estimation in the Stereo Images is carried out through disparity map using many conventional algorithms, which are very expensive and inconvenient in estimating disparity in the occlusion, discontinuities and texture less regions in the images. In order to overcome these issues, in this paper, we propose a stereo matching algorithm based on the optimized mean shift image filtering to compute the depth estimation in the 3D information’s along sparse measurement. Mean shift is a procedure for locating the maxima of a density region from given discrete data sampled from that region or Image to compute disparity between the two stereo images and disparity refinement. It is used to detect the modes of density. It determines the weight of nearby points for re-estimation of the mean. For pixels corresponding to different depths, an adaptive iterative algorithm is proposed to choose optimal frames for stereo matching, which can take advantage of continuously pose-changing imaging and save the time consumption amazingly too. The experimental results on Middlebury data’s with ground truth disparities to demonstrate that proposed method with quantitative results in order to produces high quality disparity map with less computation time and high matching accuracy along complexity Q1-reduction.

Keywords: Stereo Matching, Mean Shift Filtering, Disparity Map, Disparity estimation, Depth Estimation, Image reconstruction

1. Introduction

Image reconstruction from the stereo images has received worldwide attention over the past decade and now it is widely used in augmented reality, 3D modeling and intelligent surveillance etc. Conventional depth estimation algorithms are mainly based on image matching between two or more stereo images [1], [2]. There are mainly two types of stereo matching techniques which are heavily application dependent: sparse and dense. The sparse algorithms generally involve four steps: (i) detection of a sparse set of features, (ii) feature description, (iii) initial matching or tracking, (iv) The outlying rejection via robust estimators.

The dense algorithms aims at finding a correspondence for each pixel of a reference image based on the use of suitable cost functions and optimization strategies in order to maximize the smoothness of the output (i.e. disparity map). Under such conditions, the color of objects and their textures might present significant variations which can severely degrade the performance of stereo matching.
algorithms [3]. A variety of techniques have been proposed in the literature to tackle this issue. Among those disparities fitting algorithm is employed to perform the task of disparity refinement for images. Since they reduce the labeling of the features, it may further lead to false disparity map generation. As features of the disparity region is poorly extracted, it will be lead to poor image reconstruction. Hence in order to solve those issues, In this Paper , we propose a stereo matching algorithm based on the mean shift image filtering to compute the depth estimation in the 3D information's along sparse measurement[4].

Mean shift is a procedure for locating the maxima of a density region from given discrete data sampled from that region or Image to compute disparity between the two stereo images and disparity refinement. It is used to detect the modes of density. It determines the weight of nearby points for re-estimation of the mean [5]. For pixels corresponding to different depths, an adaptive iterative algorithm is proposed to choose optimal frames for stereo matching, which can take advantage of continuously pose-changing imaging and save the time consumption amazingly too [6].

Figure 1: Mean Shift Vector Representation of the stereo images in terms of pixel

The proposed model is simple and efficient. The process is non parametric one. Initially mean Shift algorithm classifies the reference images into regions [7]. Then local window based matching is used to estimate the disparity in each image pixel. Finally, the scene structure is modeled by a set of planar surface patches, which are needed to be stored for each segment rather than each pixel as shown in Figure 1. Instead of assigning a disparity value to each pixel, a disparity plane is assigned to each segment by which it can be reduce complexity of computing. Also in order to obtain a more reasonable disparity map, iterative optimization between adjacent regions has been employed to minimize the matching cost of the all regions of interest in the image.

Saxamato and cox et al (2013) represents multi window algorithm technique in which nine different windows are used for calculating disparity of a single pixel [8]. The window which gives the maximum correlation is used for disparity calculations.

Dense stereo matching algorithms discussed by jeon and kim et al (2010) which deal with illumination invariance are generally classified into local and global approaches depending on how the effect of illumination variations is handled [9]. Local approaches tackle the effect of variations due to light with a local window and are categorized into parametric and non-parametric methods.

The Sum of Absolute or Squared Differences (SAD/SSD) and Normalized Cross-Correlation (NCC) are the most common parametric methods. In contrast to SAD and SSD, NCC is invariant to linear brightness and contrast variations [10]. The color formation model and an Adaptive
Normalized Cross Correlation (ANCC) similarity measure were used to tackle the fattening effect of NCC [11].

Jianguo Liu et al (2016) paper discusses the occlusion problem for analyzing the stereo matching system combines post processing algorithms based on overall understanding and performance oriented stereo matching algorithm [12].

Our proposed paper is organized as introduction part in the first section and in second Section it describes stereo image for visualization in which image acquisition and mean shift algorithms were detailed in the next section we provide the experimental and result analysis and discuss their performance against the existing approaches. In the last section we provide conclusion for our stereo image processing.

2. Stereo image for visualization

The stereo image is represented in the 3D discrete disparity space as each element denotes the correspondence in order to determine the depth of the two stereo images for image visualization [13]. Visualization is the process by which the depth estimates from the stereo matching for projecting their developments. 3-Dimensional information for the above projections can be represented in following ways:
- Orthographic projections
- Perspective Projections

2.1. Image acquisition

The Stereo Image should be obtained through a Camera model [14]. The Coordinates system of both cameras is perfectly aligned.

Analysis depth of the images provides the disparity between the right and left image.

\[ \text{Depth} = \frac{1}{\text{Disparity}} \]

Depth Estimation through coordinate system is given by
Where, $\lambda$ is the point of the intersection, $W$ represents the cross ratio property, where $x_1 - x_2 = \text{Disparity}$.

### 2.2. Feature initialization

The feature of the stereo images is extracted Scale Invariant Feature transform which detect the key points in an image that are highly distinct from the neighbors, difference is determined through Gaussian function which is Invariant to scale and orientation, and placed separately in the feature vector [15]. Distortion of the image from its neighbors is calculated based on Mean Shift Segmentation Algorithm.

### 2.3. Mean shift algorithm

The mean shift algorithm is a nonparametric density estimation-based method for feature space analysis [16]. It assumes that the feature space can be regarded as an empirical probability density function of the determined parameters. Dense regions in the aspect of space corresponding to local maxima have modes of the unknown density [17]. Once the location of a mode is determined, the cluster associated with it can be delineated based on the local structure of the feature space [18]. Thus, the mode detection is an important part for the feature space analysis. In the mean shift algorithm, such a mode detection process is based on the mean shift procedure to determine the kernel.
Figure 4  Disparity Estimation using optimized Mean Shift Segmentation based Stereo matching Algorithm

Mean Shift is determined by two factors, they are as follows,
1. Computing of Mean Shift Vector
2. Estimating Multivariate kernel Density -Translation of the kernel

Multivariate kernel Density

\[ f(x) = \frac{1}{n h^d} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - x_i^d}{h}\right) \]

Guassian Kernel is given by

\[ K_N = (2\pi)^{-d/2} \exp\left(-\frac{1}{2} \|x\|^2\right) \]

Algorithm for Disparity estimation

for each row, k
for j = D to w
  cmin = ∞
for d = 0 to D  // check each possible disparity
  c(d) = f (I1(j,k), I2(j-d,k))
  if c(d) < cmin then
    dbest = d
    cmin = c(d)
  disp(j,k) = dbest  // Save best d value
  f (I1(j,k), I2(j-d,k)) = | I1 (j,k) - I2 (j-d,k) |
Figure 5 Examples of the mean Shift Segmentation (a) Left Image, (b) initial disparity, (c) ground truth data.

After determining the kernel, the disparity of the stereo images can be computed. If the image is separated by huge distance, it leads to the more noise in the image reconstruction mechanism, so we can use distance minimization principle to represent the image, the procedure is as follows,

\[
\rho(\hat{p}(y), \hat{q}) \approx \frac{1}{2} \sum_{u=1}^{m} \sqrt{\hat{p}_u(y_0) \hat{q}_u} + \frac{1}{2} \sum_{u=1}^{m} \hat{p}_u(y) \sqrt{\frac{\hat{q}_u}{\hat{p}_u(y_0)}} \\
= \frac{1}{2} \sum_{u=1}^{m} \sqrt{\hat{p}_u(y_0) \hat{q}_u} + C \frac{1}{2} \sum_{i=1}^{n} w_i K \left( \frac{y - x_i}{h} \right)
\]

Where Weight applied to the distance minimization is given by,

\[
w_i = \sum_{u=1}^{m} \delta [b(x_i) - u] \sqrt{\frac{\hat{q}_u}{\hat{p}_u(y_0)}}
\]

The depth estimation pipeline consists of two main phases: 1) choose a view which is adjacent to the k-th view, and then estimate a coarse depth map using this narrow-baseline pair. 2) Iteratively refine the depth map using kernel density for mean shift vector values [8].

3. Experimental evaluation on the real dataset

We design and implement the disparity map using Mean shift Segmentation in the experimental evaluation. Since there are so many alternatives possible for computing the Disparity Space, that it is better to sample the disparity depth at fractional disparities and to interpolate the resulting surface when looking for local minima [19]. However, real images have noise and other artifacts such as aliasing and depth discontinuities. We therefore evaluate our new techniques using the Cones and Tsukuba, and Venus stereo test sequences with ground truth from the references [20]. Two of these sequences are shown in Figure 4. We should note that the Cones data set have high-quality sub pixel accurate ground-truth estimates, while the Tsukuba ground truth has only integer disparities.

For Middlebury datasets, it is undesirable because their image size is too small to apply 3-level or more. For the Tsukuba image in the 3-level pyramid, the coarsest image size becomes 96x72 while the support window size for stereo matching is fixed to along scale levels. Moreover, the disparity range becomes indistinguishable. Sub-pixel algorithm should show robustness to various surface types.
The effect of different matching costs in texture less areas is harder to evaluate since the results depend strongly on the aggregation or Local optimization algorithm. We therefore restrict our analysis to textured areas and use a simple window-based correspondence algorithm. Un-textured areas can be handled filled using aggregation with successively larger windows.

For our analysis, we select textured pixels by computing the Multivariate kernel Density at each pixel (averaging the left and right values to remain symmetrical). These values are then averaged in a $3 \times 3$ neighborhoods and threshold, using a threshold of 6 gray levels squared which is described in figure4.

![Figure 4a– First iteration](image1)

![Figure 4b- Second Iteration](image2)

![Figure 4c- Third Iteration](image3)

![Figure 4d- Fourth Iteration](image4)

**Figure 6**: The disparity computation (depth less) results of Tsukuba stereo pair for each iteration.

The depth estimation pipeline consists of two main phases: 1) choose a view which is adjacent to the k-th view and then estimate a coarse depth map using this narrow-baseline pair. 2) Iteratively refine the depth map using kernel density of the mean Shift Algorithm.

<table>
<thead>
<tr>
<th>Image</th>
<th>Depth Estimation (m)</th>
<th>Disparity Estimation (m)</th>
<th>Window Size</th>
<th>Max. Disparity</th>
<th>Matching Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cones</td>
<td>9.05</td>
<td>11.34</td>
<td>21</td>
<td>11</td>
<td>1008ms</td>
</tr>
<tr>
<td>Tsukuba</td>
<td>6.52</td>
<td>15.32</td>
<td>11</td>
<td>15</td>
<td>196ms</td>
</tr>
</tbody>
</table>

**Table 1** matching Performance of the Middlebury dataset images using our Proposed Method

From Table 1 the numerical results of some of our experiments are detailed. There is no single setting that consistently outperforms the Others but our new cost variants optimization generally
does better than the original Costs in terms of squared difference and matching cost. Table 2 represents Performance Comparison of the optimized mean Shift segmentation against the seed growing algorithm.

<table>
<thead>
<tr>
<th>Method</th>
<th>Tsukuba Error Rate</th>
<th>Cones Error Rate</th>
<th>Venus Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Shift Segmentation</td>
<td>0.88</td>
<td>2.9</td>
<td>0.13</td>
</tr>
<tr>
<td>Optimized mean Shift Segmentation</td>
<td>0.83</td>
<td>1.7</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Table 2 Error rate Classification in the Image Restoration

Table 2 shows the performance comparison of optimized mean shift segmentation against the error rate with mean shift segmentation. Interval differences outperform squared differences on the Tsukuba and Venus data sets. Significant reduction of errors in high-frequency image regions can be carried out using window size, as predicted by our theoretical analysis. But the precision is limited by quantization, and the memory consumption grows linearly as quantization resolution increases.

In figure 5, it is clearly depicted about the distance variation between the cones images. Since variation model is sensitive to initial solution and a good initial matching is very difficult to find in wide baseline systems, Iterative method using weight injection yields better results in the matching performance. This is most apparent for the Venus images, which contain many such regions. Errors are also reduced in other areas affected by aliasing, such as strong intensity discontinuities or near-horizontal edges. Other errors, however, are not a direct result of the matching cost and can obscure the numerical results.

The Tsukuba images in particular contain fewer high-frequency regions, but several areas with repetitive patterns and fine disparity variations that are challenging for a window-based method and, thus, result in spurious errors that are not directly a function of the matching cost used. Using linear interpolation, reconstruction of images can be given clearly for stereo images. Use of multiple fitting produces a smooth disparity surface along less computational cost.
4. Conclusion

In our proposed paper, we designed and implemented stereo matching algorithm based on the mean shift image filtering to compute the depth estimation in the 3D information's along sparse measurement. Mean shift is a procedure for locating the maxima of a density region from given discrete data sampled from that region or Image to compute disparity between the two stereo images and disparity refinement. It is used to detect the modes of density. The weight of immediate points considered for re-estimation of the mean. Various pixels of different depths in accordance with adaptive iterative algorithm can choose optimal stereo matching frames. It will grant advantages of incessant pose changing images that has adequate time consuming and cost effective. The experimental and result analysis has grounds for demonstrating the disparities of truth over Middlebury data that can intend methods with quantitative techniques for providing disparity map with high quality and clarity. It takes only less computational time thus achieves high accuracy for matching them along with the complexity of Q1 reduction.

References


© 2016 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).