ABSTRACT
In this paper we propose a new technique for computing disparity maps. This paper is based on the popular correlation based disparity map computation approach. The task is to estimate a dense disparity map given two stereo images. Instead of computing the full disparity map, we only compute the disparity values for randomly selected epipolar lines. The output of this exercise is a sampled disparity map; from the sampled disparity map, the full map is reconstructed by exploiting the sparsity of the map in Fourier domain using Compressed Sensing techniques. Quantitative results show that our method of reconstructing disparity maps from sub-sampled data is only marginally inferior to the full disparity map computed by correlation based method.

KEY WORDS
Stereo Vision, Disparity Map, Compressive Sampling

1 Introduction

One of the most fundamental problems in stereo vision is the estimation of disparity maps. There are many commercial applications that require depth information; such as automated vehicle navigation, 3D tele-conferencing, 3D rendering etc. All such applications require fast and accurate reconstruction of disparity maps. Unfortunately both the objectives speed and accuracy, are hard to achieve simultaneously.

Broadly there are two approaches to address this problem local/correlation based methods and global methods. In correlation based methods, the disparity at each point is collected independently by matching pixels along the epipolar lines of the right and left images. Global methods frame the disparity map estimation as an optimization problem which in most cases needs to be solved via dynamic programming [1, 2]. Correlation based methods are easy to implement and are fast [3]. Almost all commercial products use such correlation based disparity map computation techniques. Global (optimization based) techniques are better than correlation based techniques in accuracy, but are very slow. Owing to the sluggishness of these techniques, they are not used in commercial applications.

There may be two kinds of disparity maps sparse and dense. In sparse disparity maps, one is interested in estimating the disparities of certain points (corresponding to some objects) in the scene. In automated navigation, it may be enough to know the depth (disparity) of certain points in the object of interest; knowing the disparity map for all the points of the object may not be required. However for video based applications such as 3D tele-conferencing and rendering for 3D videos, one needs to have dense disparity maps. Global methods can only produce dense disparity maps but correlation based methods can yield both dense and sparse maps; correlation based methods are more versatile. Unfortunately in a lot of situations correlation based methods fail to find disparities of certain locations. These remain as holes in the otherwise dense disparity maps. In practice, these holes are later filled in by interpolation during post-processing.

In this paper we propose a novel technique to compute disparity maps. Fundamentally it is a correlation based technique which computes the sparse disparity maps. From these sparse disparity maps, a dense map is computed by framing it as a Compressed Sensing problem. The main advantage of this method is that, the output is a dense disparity map; the unwanted holes (which could not be estimated by correlation) are not filled in. The other advantage is that, the method although slower than correlation based methods is much faster than optimization based (dynamic programming) techniques. This paper is divided into several sections. Section II describes Stereo vision and correlation based algorithms, Section III discussed the compressive sampling, Section IV discusses the algorithm to implement compressive sampling in disparity map computation, and Section V discusses the accuracy of the results obtained for the Venus Image pair [7] and other stereo images pairs from Middlebury evaluation site [9].

2 Stereo Vision- Correlation Based Methods

Stereo vision helps us estimating depth information from two slightly shifted versions of the same scene. Objects which are nearer shift more while objects that are further away in the scene shift less. Mammals have stereo vision; our right and left eyes perceive two slightly shifted versions of the same scene. Our brain processes these to estimate the depth information. In computer vision the physical principle for computing the disparity map is the same.

As mentioned before, there are two broad approaches
for disparity map estimation. We will briefly discuss the basics of correlation based techniques; this is because our implementation is based on this approach. We will not use global approaches for disparity map computation, and hence will not be discussing it further.

In correlation based methods the disparity map is computed pixel-by-pixel. One of the images (either left or the right) is considered as a reference. Each pixel in the reference image is matched along the epipolar line of the other image. The match is based on a correlation score (hence the name); different statistical measures can be used in practice (correlation, Euclidian distance, taxi-cab distance, cosine distance etc.). The pixel is said to match if the score is above a certain defined threshold.

This is the basic technique. It is simple and intuitive. It is fast to implement. Moreover since finding the match for each pixel is an independent operation, it can be parallelized using multi-core CPUs or GPUs. Owing to its speed and simplicity correlation based methods are widely used in commercial applications. In recent years certain variations to the existing baseline method has been proposed [3, 4].

The main problem with this approach is that, in many a times there is no match for a certain pixel in the reference image. In that case, the disparity map at that point cannot be computed. There is a hole in the disparity map. Post-processing interpolation techniques (nearest neighbor, bi-linear, bi-cubic, etc.) are used to fill in these holes. Such filling-in techniques are not very accurate. In general correlation based disparity map computation proceeds in the following steps:

1. Geometry Correction – corrects the distortions in input images by warping into a standard form
2. Image Transformation – transforms each pixel in grayscale image to an appropriate normalized form based on average local intensity
3. Correlation Measure – each pixel of the reference image with pixels along the epipolar line of the shifted image
4. Extrema Extraction – the extreme value of correlation at each pixel is determined yielding a disparity image
5. Post-processing – to fill in holes and remove noise in the map.

3 Compressive Sampling

Compressive Sampling or Compressed Sensing (CS) [5, 6] addresses the problem of solving an under-determined system of linear equations where the solution is known to be sparse. Therefore the problem to be solved is,

\[ y_{m \times 1} = A_{m \times n} x_{n \times 1}, \quad m < n \] (1)

This is an under-determined system and in general does not have a unique solution. But theoretical studies in CS show that, if the solution is sparse, it is necessarily unique.

Since the solution is unique, one is interested in finding it. Since the sparse solution is unique, finding the sparsest solution of (1) will be the desired solution; since there cannot be more than one sparse solution, the sparsest solution is the one we want to obtain. The sparsest solution can be found by solving the following optimization problem,

\[ \min \| x \|_0 \quad \text{subject to } y = Ax \] (2)

Here the \( \ell_0 \)-norm counts the number of non-zeros in \( x \), and hence minimizing the number of non-zeros yields the sparsest solution. Unfortunately minimizing the \( \ell_0 \)-norm is an NP (Non-deterministic Polynomially) hard problem. Thus solving (2) is not practical. CS proves that, when the NP hard objective function ( \( \ell_0 \)-norm) is replaced by its closest convex surrogate (\( \ell_1 \)-norm), minimizing the surrogate (subject to the constraints) yields the correct result. Thus, CS says that the results obtained from (2) is the same as solving the following (3) when the solution is known to be sparse.

\[ \min \| x \|_1 \quad \text{subject to } y = Ax \] (3)

The \( \ell_1 \)-norm is the sum of absolute values in \( x \). This (3) is a convex problem which can be solved via linear programming. In recent times dedicated fast and efficient \( \ell_1 \)-norm minimization solvers such as Spectral Projected Gradient L1 [7] are available; these solvers are 2 orders of magnitude faster than standard linear programs.

4 Proposed Method

In this paper we first compute the disparity maps via a correlation based method. But we do not compute the dense disparity map. Disparity maps are computed along randomly chosen epipolar lines. In this work, we assume that the epipolar lines are horizontal. In Fig. 1, we show in white the lines where the disparity maps will be computed, the disparity map will not be computed along the black lines. Fig. 1 shows the sampled epipolar lines (Fig. 1b) and the corresponding disparity map (Fig. 1b) obtained via correlation based algorithm [9]. Thus Fig. 1b is a sampled (masked) version of the full disparity map (Fig. 1c).

Mathematically, the operation can be expressed as,

\[ b = Rd \] (4)

Here \( b \) is the sampled disparity map (Fig. 1c), \( R \) is the sampling mask (Fig. 1b) and \( d \) is the full disparity map (Fig. 1a). Since the ultimate goal is to find a dense disparity map, we have to estimate \( d \) from (4). This is a classic linear inverse problem. Unfortunately the problem is under-determined owing to the sampling. Thus, there is no unique solution.
In the last section, we have learnt that CS can solve an under-determined system of equations when the solution is known to be sparse. The disparity map is not sparse. However it has a sparse representation in Fourier frequency domain. The disparity map is smooth in most places except where there is a sudden change in depth; these are the discontinuities in the disparity map. Thus, if we take a Fourier transform of the disparity map, most of the energy will be concentrated in the low frequency components. Thus the Fourier transform of the disparity map will be sparse. This is experimentally verified in Fig. 2, where we plot the sorted absolute values of the Fourier coefficients of the disparity map in Fig. 1a. We see that the sorted Fourier coefficients have a very fast decay; meaning that the Fourier transform of the disparity maps is very sparse.

The Fourier transform is orthogonal, therefore it is possible to express it in the following analysis-synthesis form,

\begin{align*}
\text{analysis} & : \alpha = Fd \\
\text{synthesis} & : d = F^T \alpha
\end{align*}

If \(d\) is the signal (in our case the disparity map), then the forward transform (analysis) \(\alpha\) is sparse. The inverse transform is just a transpose (synthesis). Incorporating the synthesis equation into (4) we get,

\[ b = RF^T \alpha \]

The problem now is to solve \(\alpha\). Since the solution is known to be sparse, we can invoke CS to solve it. We solve for the sparse Fourier transform coefficients by solving \(\ell_1\)-norm minimization (3) using [7]. Once the sparse Fourier coefficients are obtained, the disparity map can be computed from the synthesis equation (5).

### 4.1 Hole-Filling

Correlation based methods cannot compute some portions of the disparity maps. These portions are the holes which we have already mentioned before. In our proposed framework, the holes can also be treated as unsampled locations. Till now we were considering a mask \(R\) which corresponded to sampled epipolar lines. The mask contains ones in sampled locations and zeroes at unsampled positions. As we just mentioned, even within the sampled epipolar lines, there may be positions for which disparity map could not be computed from correlation based methods. The positions of these holes can also be incorporated in the mask, i.e. those positions in \(R\) which have been sampled, but do not have disparity values are treated as unsampled locations and put to zeroes. Thus \(R\) not only takes into account the unsampled epipolar lines, but also considers points in the sampled lines where disparity values could not be computed.

## 5 Experimental Evaluation

The proposed method is tested on some standard datasets from the Middlebury evaluation site [10]. The groundtruth disparity maps are the ones supplied with the dataset. For our method, we use two sampling ratios of 50% and 75% sampling. The points in the left image for which no match is found are not interpolated and are represented as black pixels towards the left portion of the disparity map. The disparity maps along with the ground truth are shown in Fig. 2.

For visual evaluation we show the reference images, the groundtruth disparity maps and the reconstructed disparity maps in Fig. 2. We see that our proposed methods show some reconstructed artifacts along the unsampled lines. For quantitative evaluation, the results are shown in Table 1. The Root Mean Square Error (RMSE) between groundtruth is calculated. The RMSE as an evaluation metric has proposed in [10].

We see that even though visual evaluation shows horizontal reconstruction artifacts, the reconstruction accuracy from our proposed method with 50% and 75% sampling is quiet close to disparity map from full data in terms of RMSE. For 50% sampling, the percentage degradation in terms of RMSE is 1.03 and for 75% sampling, the degradation in RMSE is only 0.53%. Such degradations in RMSE are negligible in terms as far as quantitative evalu-
Figure 2. 1\textsuperscript{st} Column – Cones; 2\textsuperscript{nd} Column – Teddy; 3\textsuperscript{rd} Column – Tsukuba. Top to Bottom Reference Image, Supplied Disparity map, Disparity map computed by [9], Disparity map computed by proposed method with 75\% sampling; Disparity map computed by proposed method with 50\% sampling.
<table>
<thead>
<tr>
<th>Name of Dataset</th>
<th>Baseline algorithm [9]</th>
<th>Proposed-50% sampling</th>
<th>Proposed-75% sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cones</td>
<td>15.87</td>
<td>15.61</td>
<td>15.77</td>
</tr>
<tr>
<td>Teddy</td>
<td>15.888</td>
<td>15.61</td>
<td>15.77</td>
</tr>
<tr>
<td>Tsukuba</td>
<td>13.31</td>
<td>13.25</td>
<td>13.28</td>
</tr>
</tbody>
</table>

The reconstruction artifacts are easily discernible, yet when we look at the quantitative results the errors are negligible. This phenomenon can be explained by the hole-filling capability of our proposed method. Owing to subsampling the reconstructed disparity map is of poorer quality, but our method fills in the holes that are left behind by correlation based methods. Since these holes are filled by our method, the reconstruction error (RMSE) with respect to the groundtruth is reduced. Thus one on one hand, subsampling introduces error, whereas on the other the hole-filling reduces some error. The overall effect is that the reconstruction is balanced, and we perceive negligible degradation in reconstruction error.

6 Conclusion

In this work we propose a new technique for disparity map computation. This work is a proof-of-concept. We show that it is possible to reconstruct disparity maps from disparity values computed along a subset of all possible epipolar lines. Even with 50% sampling, our method shows marginal degradation in disparity map quality.

There are several possible improvements to this work. The first improvement can be effected by changing the sampling pattern. In this work we have considered cases where the sampling is along epipolar lines. Some of the epipolar lines are fully sampled while others are completely omitted. Compressed Sensing do not advocate such sampling; for best CS reconstruction the sampling locations should be random. In future, we will compute randomly sampled sparse disparity maps; in theory one will get better reconstruction results from such random sampling.

In this work, we have not considered errors in the computed (via correlation based method) disparity maps. Generally the errors are sparse but appear as shot noise. It is possible to remove such sparse shot noise through reformulating the CS problem. This would improve the depth map computation even further.

References


[10] vision.middlebury.edu/stereo/