

Dense Disparity Maps from Sparse Disparity Measurements

Simon Hawe, Martin Kleinsteuber, and Klaus Diepold

Department of Electrical Engineering and Information Technology
Technische Universität München, Arcisstraße 21, 80290 München, Germany

{simon.hawe, kleinsteuber, kldi}@tum.de

Abstract

In this work we propose a method for estimating disparity maps from very few measurements. Based on the theory of Compressive Sensing, our algorithm accurately reconstructs disparity maps only using about 5% of the entire map. We propose a conjugate subgradient method for the arising optimization problem that is applicable to large scale systems and recovers the disparity map efficiently. Experiments are provided that show the effectiveness of the proposed approach and robust behavior under noisy conditions.

1. Introduction

One of the oldest and also most important research topics in computer vision is the stereo correspondence problem. Given two or more images of the same scene but taken from different viewpoints, the correspondence problem aims at determining which image points in different images are projections of the same physical point in space. Essentially, correspondences are found by comparing image intensities, feature points, or certain shapes extracted from the images at hand. A common simplification for this search is to rectify the stereo images, as after this transformation corresponding points will lie on the same horizontal scanline and therefore the search space is reduced by one dimension. When correspondences for all points of an image have been established, a dense disparity map of the observed scene can be created. The term disparity denotes the difference in coordinates of corresponding points within two images, and is an inverse measurement of distance. Disparity maps are essential for various applications like 3D-television, image based rendering, or robotic navigation. In the following we will shortly review the state-of-the-art concepts used for disparity estimation.

The most basic tool needed for finding corresponding points is a matching cost function that measures image similarity. Most widely used are matching cost functions that compare image intensities by their absolute or squared

differences, but there exist several other techniques [16]. Unfortunately, only considering a simple matching cost is highly ambiguous due to occlusions, or either missing or repetitive textures. To overcome these ambiguities, a variety of dense stereo methods have been proposed throughout the last 40 years, see [25] for an excellent overview and performance comparison. Generally, these approaches can be divided into two groups: local methods that assign a disparity to each position individually, and global methods that minimize an energy function either along a single scanline, or over the entire disparity map.

Local methods aggregate the matching cost of all pixels lying in a support window surrounding a point $\mathbf{q}$, and assign the disparity to $\mathbf{q}$ that results in the lowest aggregated cost. These methods assume constant disparity over the support, which does not hold at disparity discontinuities and leads to boundary fattening artifacts. Explicit occlusion detection [11], multiple windowing [14], or adaptive local support weighting [29] reduce this effect, but cannot avoid it completely.

A further problem appears in regions showing no texture or repetitive textures, due to many indistinguishable possible matches. Global methods tackle these ambiguities by assuming that disparities vary smoothly in general. This is enforced by minimizing an energy function consisting of the matching cost and a discontinuity penalty. In that way, global support is provided for local regions lacking textures. Scanline methods minimize an energy function along each horizontal scanline separately, using dynamic programming [23, 25]. In regions fulfilling the smoothness assumption, this approach results in accurate disparity estimation, but leads to horizontal streaking artifacts around discontinuities. This problem is bypassed by methods that enforce the smoothness constraint in both vertical and horizontal direction. The most prominent stereo algorithms for minimizing the 2D cost function are based on belief propagation [13, 17, 26] and graph cuts [2, 18].

Besides the dense stereo methods reviewed above, algorithms that explicitly produce reliable sparse disparity maps have been proposed in the literature [24]. However,

to the best of our knowledge, no technique has been proposed so far that aims at reconstructing a dense disparity map from these sparse measurements, except simple interpolation which is prone to errors and creates bulky disparity maps, especially at object borders.

In this paper, we present a method that accurately reconstructs dense disparity maps by only using a small set of reliable support points. It is based upon the theory of Compressive Sensing and incorporates prior knowledge about both the general structure of disparity maps and the specific image at hand into the optimization scheme. We propose a conjugate subgradient method that abandons the use of a smoothing parameter and thus leads to unbiased solutions. Our experiments show, that only about five percent of the entire disparities are necessary for our method to produce very accurate results. Hence, the proposed algorithm is suitable for a combined approach in creating and compressing disparity maps.

The paper is organized as follows. In the next section, we shortly recall the basics of Compressive Sensing and introduce the required mathematical notation. In Section 3, we explain the shortcomings of random sampling and how to overcome them. A conjugate subgradient method is presented in Section 4 that efficiently solves the arising non-smooth optimization problem, and in Section 5 we provide numerical simulations that show our algorithm's capability.

2. Compressive Sensing

A common ground to many interesting signals is that they have a sparse representation or are compressible in some transform domain, which means that many transform coefficient are zero or close to zero. For example, the standard representation of an image of a natural scene by its intensities is dense, whereas the wavelet basis admits a sparse representation. The important information is contained in only a few dominant coefficients. To reconstruct the entire image without severe loss of quality, it is sufficient to know these large wavelet coefficients together with their respective positions. If we could directly sample the important coefficients in the sparse domain, we could bypass the inefficient process of first sampling densely, then calculate the sparse representation to afterwards compress. The concept of Compressive Sensing (CS) [4, 8] offers a joint sampling and compression mechanism, which exploits a signal's sparsity to perfectly reconstruct it from a small number of acquired measurements.

Let $\mathbf{s} \in \mathbb{R}^n$ be a column vector that represents a discrete n -dimensional non-sparse real-valued signal. We denote its k -sparse representation by $\mathbf{x} \in \mathbb{R}^n$, where k -sparse means that only $k < n$ entries of $\mathbf{x}$ are nonzero. We write the corresponding linear transformation as

$$\mathbf{s} = \Psi \mathbf{x}, \quad (1)$$

where $\Psi \in \mathbb{R}^{n \times n}$ is an orthonormal basis of $\mathbb{R}^n$, called *representation basis*. Furthermore, let $\Phi \in \mathbb{R}^{m \times n}$ be the *sampling basis* that transforms $\mathbf{s}$ into the vector $\mathbf{y} \in \mathbb{R}^m$ that contains $m < n$ linear measurements

$$\mathbf{y} = \Phi \mathbf{s} = \Phi \Psi \mathbf{x}. \quad (2)$$

We aim to reconstruct $\mathbf{s}$ given only the measurements $\mathbf{y}$ by computing $\mathbf{x}$ from equation (2) and exploiting the fact that $\mathbf{x}$ is sparse. Informally speaking, we are seeking the sparsest vector $\mathbf{x}$ that is compatible with the acquired measurements. Formally, this leads to the following minimization problem

$$\begin{aligned} & \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad \|\mathbf{x}\|_0 \\ & \text{subject to} \quad \mathbf{y} = \Phi \Psi \mathbf{x}, \end{aligned} \quad (3)$$

where $\|\mathbf{x}\|_0$ is the ℓ_0 -pseudo norm of $\mathbf{x}$, i.e. the number of nonzero entries.

Unfortunately, solving (3) is computationally intractable as it is a combinatorial NP-hard problem [20]. Instead, it has been shown in [9] that under some generic assumptions on the matrix $\Phi \Psi$ it is equivalent to replace the ℓ_0 -pseudo norm by the ℓ_1 -norm $\|\mathbf{x}\|_1 := \sum_i |x(i)|$, which leads to the so called Basis Pursuit [7]

$$\begin{aligned} & \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad \|\mathbf{x}\|_1 \\ & \text{subject to} \quad \mathbf{y} = \Phi \Psi \mathbf{x}. \end{aligned} \quad (4)$$

This is a convex optimization problem that can be recast into a linear program, which is solved in polynomial time. The theory of CS says that if the number of measurements m is high enough compared to the sparsity factor k , the solution to equation (4) is exact [5], and consequently the signal is perfectly reconstructed by solving equation (1), using the computed $\mathbf{x}$.

When the signal is sampled in the same basis in which it is sparse, a lot of samples are required for reconstruction (since most of the samples would be zero). Hence, it is intuitively clear that sampling and representation basis have to be *as disjoint as possible*. This is measured by the mutual coherence between Φ and Ψ

$$\mu(\Phi, \Psi) = \sqrt{n} \max_{i,j} |(\Phi \Psi)(i, j)|, \quad (5)$$

where $X(i, j)$ denotes the (i, j) -entry of the matrix X . The smaller the value $\mu(\Phi, \Psi)$ the higher the incoherence and the more favorable is the pair of bases [10]. The relation between the number m of required random samples for perfect reconstruction, the coherence, the sparsity of $\mathbf{x}$, and the dimension n of the signal is provided by the famous formula [3]

$$m \geq C \mu^2(\Phi, \Psi) \|\mathbf{x}\|_0 \log n, \quad (6)$$

where C is some positive constant. Interestingly, randomly drawn sampling matrices provide a low coherence with high probability, [10]. A slightly modified coherence measure is introduced in [12], which is used to find optimized sampling matrices. We emphasize, that this concept is not applicable to our problem at hand, since our sampling space is restricted to the pixel basis.

3. Sampling and Representing Disparity Maps

Let $\mathcal{D} \in \mathbb{R}^{h \times w}$ be a disparity map having $n = hw$ entries and assume that $m \ll n$ disparities are known. In contrast to methods that rely on particular camera set-ups, these disparities can be computed by any method. The accuracy of our recovery algorithm only depends on the sampling *positions*.

In general, disparity maps mainly consist of large homogenous regions of equal disparity with only a few discontinuities at the transitions between those regions. Regarding the wavelet transform, large homogenous regions are represented by only a small number of wavelet coefficients, while the important coefficients cluster around discontinuities. For this, we can assume the wavelet transform of disparity maps to be sparse, and use the wavelet domain as the representation domain.

Let $\mathbf{s} \in \mathbb{R}^n$ be the vectorized unknown disparity map $\mathcal{D}$, and let $\mathbf{y} \in \mathbb{R}^m$ denote the disparity measurements. Moreover, let Ψ denote a Daubechies Wavelet basis, then $\mathbf{s} = \Psi \mathbf{x}$ with $\mathbf{x} \in \mathbb{R}^n$ is the sparse vector of wavelet coefficients. To each measurement y_i there corresponds a standard basis vector $\mathbf{e}_i \in \mathbb{R}^n$ such that

$$y_i = \mathbf{e}_i^\top \mathbf{s} = \mathbf{e}_i^\top \Psi \mathbf{x}, \quad (7)$$

where $(\cdot)^\top$ denotes transpose. Generally, we denote by $\mathbf{v}(i)$ the i -th entry of the vector $\mathbf{v}$. Let $\mathbf{p} \in \mathbb{N}^m$ be a vector containing the indices of the measured disparities, the sampling basis for our problem reads as

$$\Phi = [\mathbf{e}_{\mathbf{p}(1)}, \dots, \mathbf{e}_{\mathbf{p}(m)}]^\top. \quad (8)$$

Unfortunately, the mutual coherence between the canonical basis and the wavelet basis is high. According to equation (6), this requires a high number of random measurements. However, by selecting particular sampling positions, we can achieve accurate reconstruction results using only a small number of measurements, even though the bases do not fulfill the low coherence requirement.

Motivated by the property of the wavelet transform that the relevant coefficients coincide with discontinuities, the particular sampling positions are precisely those disparities lying at the discontinuities. As we do not know the positions of the discontinuities before knowing the disparity map, we use the assumption that disparity discontinuities coincide with image intensity edges. Therefore, we apply the



Figure 1: Sampling pattern covering 5% of the image.

Canny filter [6] to the reference image, and take the positions of the detected edges as the sampling positions. Note, that also other edge detectors are thinkable. Furthermore, to control the minimum sampling density, we divide the image into non-overlapping tiles and select one sampling position inside each tile where no edge has been previously detected, see Figure 1 for an exemplarily sampling pattern. For these selected positions, we calculate the respective disparities, which are the input to the reconstruction algorithm described in the next section.

4. Reconstructing Dense Disparity Maps

Various generic solvers based on second-order methods [15], and algorithms based on first-order methods [1, 21] to tackle problem (4) have been proposed in the literature. Second-order methods are very accurate but require too much computational resources to solve the problem at hand. Due to its large scale, we suggest a first-order method, which roughly follows the ideas in [19], but uses a new sub-gradient method. In order to enhance legibility, we stick to the matrix vector notation. However, regarding the implementation, we want to mention that it is not possible to simply use these matrix vector multiplications, due to the large size of the involved matrices. Fortunately, all matrix operations we are performing can be efficiently implemented without explicitly creating the matrices by using image filtering. A Matlab-code of the algorithm is available at the author's webpage¹.

4.1. Prerequisites

The local variation of the disparity map $\mathcal{D}$ at entry (i, j) is measured by

$$\nabla \mathcal{D}(i, j) := [\mathcal{D}(i, j) - \mathcal{D}(i, j+1), \mathcal{D}(i, j) - \mathcal{D}(i+1, j)]. \quad (9)$$

From this, we define the *total variation (TV) norm* of $\mathcal{D}$ as

$$\|\mathcal{D}\|_{\text{TV}} = \sum_{i=1}^{h-1} \sum_{j=1}^{w-1} \|\nabla \mathcal{D}(i, j)\|_2. \quad (10)$$

¹<http://www.gol.ei.tum.de/index.php?id=25>

The TV-norm is commonly used by various image reconstruction algorithms as a discontinuity preserving smoothness prior, and we also employ it here. It is easily seen that by means of two suitable matrices $\mathcal{G}_x, \mathcal{G}_y \in \mathbb{R}^{n \times n}$ we have

$$\|\mathbf{s}\|_{\text{TV}} := \|\mathcal{D}\|_{\text{TV}} = \sum_{j=1}^n \sqrt{(\mathbf{e}_j^\top \mathcal{G}_x \mathbf{s})^2 + (\mathbf{e}_j^\top \mathcal{G}_y \mathbf{s})^2}. \quad (11)$$

Another extension to problem (4) is that we introduce the square diagonal weighting matrix $\mathcal{W} \in \mathbb{R}^{n \times n}$ for the wavelet coefficients. Recall, that wavelet coefficients can be divided into approximation coefficients representing the low frequency parts, and detail coefficients representing the high frequency parts. Generally, the approximation coefficients are dense and the sparsity is present within the detail coefficients. We explicitly enforce this, by setting all the diagonal entries of $\mathcal{W}$, which coincide with detail coefficients to one, and all the others to zero. This weighting scheme allows to further reduce the number of measurements required for wavelet based Compressive Sensing [27]. Additionally, to increase the algorithm's robustness we relax the data fidelity term for the constraints to account for noisy disparity measurements. Putting all this together, we end up with the optimization problem

$$\begin{aligned} & \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad \|\mathcal{W}\mathbf{x}\|_1 + \gamma \|\Psi\mathbf{x}\|_{\text{TV}} \\ & \text{subject to} \quad \|\mathcal{A}\mathbf{x} - \mathbf{y}\|_2^2 \leq \epsilon. \end{aligned} \quad (12)$$

with $\mathcal{A} = \Phi\Psi$. The parameter $\epsilon \geq 0$ can be interpreted as a bound of the noise energy contained in the measurements and $\gamma \geq 0$ is a weighing parameter. In order to incorporate the constraints into the objective, we follow a common approach and reformulate (12) in unconstrained Lagrangian form

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2} \|\mathcal{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda (\|\mathcal{W}\mathbf{x}\|_1 + \gamma \|\Psi\mathbf{x}\|_{\text{TV}}). \quad (13)$$

The Lagrange multiplier λ weighs between the contribution of the objective and the constraint.

4.2. Conjugate Subgradient

Minimizing the objective of (13) by a first-order method requires to compute its gradient. A common procedure to overcome its non-smoothness is to approximate the absolute value with the smooth differentiable function $|x| \approx \sqrt{x^2 + \nu}$, with $\nu > 0$ being a smoothing factor, [19]. This approach, however, biases the sparse solution, in particular if $\|\mathbf{x}\|_1$ is small. Therefore, we propose a conjugate subgradient method that does not require an approximation of $\|\mathcal{W}\mathbf{x}\|_1$ and consequently yields an unbiased solution. This is in accordance to our experiments. Moreover, we observed that the subgradient method requires less iterations. Hence,

it is computationally cheaper and more accurate compared to using the smoothing term.

In contrast to smooth objectives where a unique gradient exists, the set of subgradients of (13) in general consists of infinitely many elements. We follow the approach proposed in [28] where it is shown that the optimal subgradient for optimization purposes is the one with smallest Euclidian norm.

As $\mathcal{W}$ is diagonal and consequently $\mathcal{W}^\top = \mathcal{W}$, the subdifferential of $\|\mathcal{W}\mathbf{x}\|_1$ is the set $\partial\|\mathcal{W}\mathbf{x}\|_1 \subset \mathbb{R}^n$ with

$$\partial\|\mathcal{W}\mathbf{x}\|_1(i) = \begin{cases} \frac{(\mathcal{W}\mathbf{x})(i)}{|(\mathcal{W}\mathbf{x})(i)|} & \text{if } (\mathcal{W}\mathbf{x})(i) \neq 0 \\ [-1, 1] & \text{otherwise.} \end{cases} \quad (14)$$

An implementation of a subgradient for the prior $\|\Psi\mathbf{x}\|_{\text{TV}}$ is computationally unfeasible due to the computation of the subgradient with smallest Euclidean norm. Therefore, we use the Huber functional

$$h_\nu(x) = \begin{cases} |x| - \frac{\nu}{2} & \text{if } |x| \geq \nu \\ \frac{x^2}{2\nu} & \text{otherwise,} \end{cases} \quad (15)$$

for a smooth approximation of the TV-norm

$$\|\mathbf{s}\|_{\text{TV}} \approx \|\mathbf{s}\|_{\nu, \text{TV}} := \sum_{j=1}^n h_\nu \left(\sqrt{(\mathbf{e}_j^\top \mathcal{G}_x \mathbf{s})^2 + (\mathbf{e}_j^\top \mathcal{G}_y \mathbf{s})^2} \right). \quad (16)$$

Note, that smoothing the prior does not bias the sparse solution in $\|\mathcal{W}\mathbf{x}\|_1$. With the shorthand notation

$$\mathbf{r} := \Psi^\top \sum_{j=1}^n \frac{\mathcal{G}_x^\top \mathbf{e}_j \mathbf{e}_j^\top \mathcal{G}_x \mathbf{s} + \mathcal{G}_y^\top \mathbf{e}_j \mathbf{e}_j^\top \mathcal{G}_y \mathbf{s}}{\sqrt{(\mathbf{e}_j^\top \mathcal{G}_x \mathbf{s})^2 + (\mathbf{e}_j^\top \mathcal{G}_y \mathbf{s})^2}}, \quad (17)$$

the gradient of $\|\Psi\mathbf{x}\|_{\nu, \text{TV}}$ is given by $\nabla\|\Psi\mathbf{x}\|_{\nu, \text{TV}}$ with entries

$$\nabla\|\Psi\mathbf{x}\|_{\nu, \text{TV}}(i) = \begin{cases} \mathbf{r}(i), & \text{if } |\mathbf{r}(i)| \geq \nu \\ \mathbf{r}(i)/\nu & \text{otherwise.} \end{cases} \quad (18)$$

Therefore, the subdifferential of the modified objective

$$f(\mathbf{x}) = \frac{1}{2} \|\mathcal{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda (\|\mathcal{W}\mathbf{x}\|_1 + \gamma \|\Psi\mathbf{x}\|_{\nu, \text{TV}}) \quad (19)$$

is the set

$$\partial f(\mathbf{x}) = \mathcal{A}^\top (\mathcal{A}\mathbf{x} - \mathbf{y}) + \lambda (\partial\|\mathcal{W}\mathbf{x}\|_1 + \gamma \nabla\|\Psi\mathbf{x}\|_{\nu, \text{TV}}). \quad (20)$$

Let us denote

$$\mathbf{b} = \lambda^{-1} \mathcal{A}^\top (\mathcal{A}\mathbf{x} - \mathbf{y}) + \gamma \nabla\|\Psi\mathbf{x}\|_{\nu, \text{TV}}. \quad (21)$$

It is easily verified that the final subgradient with smallest Euclidean norm is given by

$$\mathbf{g}(\mathbf{x}) = \mathcal{A}^\top (\mathcal{A}\mathbf{x} - \mathbf{y}) + \lambda (\nabla\|\mathcal{W}\mathbf{x}\|_1 + \gamma \nabla\|\Psi\mathbf{x}\|_{\nu, \text{TV}}), \quad (22)$$

	Tsukuba	Venus	Teddy	Cones
HS	295	179	248	337
PR	705	500	639	874
FR	607	374	465	701

Table 1: Number of iterations until convergence for different update formulas.

where

$$\nabla \|\mathcal{W}\mathbf{x}\|_1(i) := \begin{cases} \frac{(\mathcal{W}\mathbf{x})(i)}{|(\mathcal{W}\mathbf{x})(i)|} & \text{if } (\mathcal{W}\mathbf{x})(i) \neq 0 \\ -\text{sign}(\mathbf{b}(i)) \min\{|\mathbf{b}(i)|, 1\} & \text{otherwise.} \end{cases} \quad (23)$$

The descent method is initiated with $\mathbf{x}_0 = \mathcal{A}^\top \mathbf{y}$ and iteratively updates

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{h}_i. \quad (24)$$

The scalar $\alpha_i \geq 0$ is the line-search parameter or the step length, and $\mathbf{h}_i$ is the descent direction at the i^{th} iteration. Various line-search techniques for computing α_i exist from which we choose *backtracking line-search* [22] as it is conceptually simple and computationally cheap. Regarding the descent direction, we tested our method with several update formulas for $\mathbf{h}_i$ [22]. Our experiments showed that Hestenes-Stiefel (HS), Polak-Ribière (PR), and Fletcher-Reeves (FR, used in [19]) led to the best reconstruction results. In terms of the convergence speed, HS outperformed all other techniques, see Table 1 for a comparison on our test data. Moreover, we noticed that the rate of convergence of the method with HS is nearly independent from the chosen line-search parameters, whereas all other methods suffer from requiring a cumbersome parameter tuning. These observations led us to implement the HS descent direction update

$$\mathbf{h}_{i+1} = -\mathbf{g}_{i+1} + \frac{\mathbf{g}_{i+1}^\top (\mathbf{g}_{i+1} - \mathbf{g}_i)}{\mathbf{h}_i^\top (\mathbf{g}_{i+1} - \mathbf{g}_i)} \mathbf{h}_i, \quad (25)$$

where $\mathbf{g}_i := \mathbf{g}(\mathbf{x}_i)$ and initial value $\mathbf{h}_0 = -\mathbf{g}_0$.

Equations (24) and (25) are iterated either until convergence is achieved, or a maximum number of iterations has been reached. As a stopping criterion we choose the norm of $\mathbf{g}_i$. Usually, conjugate gradient methods use a reset after n iterations, i.e. $\mathbf{h}_i := -\mathbf{g}_i$ if $i \bmod n = 0$. Since our algorithm converges in relatively few steps compared to the dimension n , we do not stress the issue of resetting. The output of this algorithm are the reconstructed wavelet coefficients $\mathbf{x}$ from which we finally compute the dense disparity map using equation (1).

5. Results

In this section we evaluate the accuracy of our reconstruction algorithm on the standard Middlebury dataset

[25], which provides stereo images together with the respective ground truth disparity maps. As usual, the disparity is given in units between 0 and 255. We measure the reconstruction quality in two ways:

- the amount of defective pixels, i.e. the percentage of pixels whose reconstructed value differs by more than one from its true value
- the mean absolute disparity error over the entire disparity map.

All experiments were performed with the same parameter setting. We used Daubechies four tap wavelets (db2-wavelets) due to their low computational cost and their qualitatively high results. In contrast, the usage of Haar wavelets led to pronged edges in the reconstructed disparity map. We chose the wavelet decomposition level one, since higher decomposition levels did not improve the reconstruction quality. The scaling factors from equation (19) were experimentally determined to $\lambda = 0.01$, $\gamma = 10$, and $\nu = 0.01$.

We compare reconstruction by interpolation via Delaunay triangulation (RDT) with:

1. our Compressive Sensing based reconstruction algorithm using Canny Edge (CSC) detector as explained in Section 3
2. our Compressive Sensing based reconstruction algorithm using random samplings (CSR).

For RDT we use the same sampling positions as for CSC. The two quality criteria are plotted against the number of measurements m in Figure 2. The results were gathered from reconstructing the standard images Tsukuba, Venus, Teddy, and Cones [25]. The plots in each row correspond to the reconstruction of the respective disparity map shown in the first column.

To show the effectiveness of our method, we first discuss the reconstruction accuracy *using ideal disparity measurements* extracted from the ground truth disparity maps, cf. column (b) and (d) of Figure 2. It can be seen that all three algorithms perform reasonably well if the number of measurements is high, and that CSC always performs best. The advantage of CSC over the other methods becomes significantly high when few measurements are used.

There is another very important advantage of the CSC method. While RDT leads to fringy reconstruction results in the neighborhood of discontinuities, CSC does not yield this highly unwanted effect and produces very accurate disparity maps, in particular in the neighborhood of edges, cf. Figure 3.

To evaluate the influence of *corrupted measurements* on the reconstruction results, we use the same sampling positions as for the ideal case and add uniformly distributed

noise drawn from the interval $[-15, 15]$ to randomly chosen 25% of the data. Columns (c) and (e) of Figure 2 illustrate the performance of the tested methods. While CSC and CSR are very robust to corrupted measurements, the performance of RDT dramatically drops if corrupted measurements are present. From column (c) it can be noticed that the reconstruction error slightly increases once a certain number of measurements has been reached. The reason for this is that the total number of corrupted measurements increases, which consequently leads to a higher percentage of badly reconstructed disparities.

Finally, we evaluate our method using *real estimated disparities*. To compute the disparities, we used the adaptive support window approach combined with a left right consistency check [29]. The last column of Figure 3 shows the reconstructed disparity maps with *computed* disparities. Our results are compared with the dense approach from [29] and RDT using the Middlebury criteria, cf. Table 2. The results of our algorithm are comparable to the dense reference method, while RDT is significantly worse. Recall, that we did not implement a specific sparse algorithm. By improving the quality of the disparity measurements for example by incorporating accurate feature matchers like SIFT, the accuracy of the measurements will increase and consequently, the reconstruction quality will also improve.

6. Conclusion and Outlook

In this paper we present an algorithm for dense disparity map reconstruction using sparse reliable disparity measurements. Based on the theory of Compressive Sensing, we perform the reconstruction by exploiting the sparsity of disparity maps in the Wavelet domain and a particular sampling strategy. To efficiently solve the arising large scale optimization problem, we introduce a conjugate subgradient method. Our numerical results show that we achieve accurate disparity map reconstruction by using only 5% of the disparity information. Moreover, they illustrate the algorithm's robustness to corrupted measurements.

Yet, we have not optimized the computation time of our algorithm. Our Matlab implementation reconstructs a 640×480 disparity map in approximately one minute on a standard desktop computer. As the computationally most demanding part is the wavelet transform which has to be performed at each iteration, a massive speedup can be achieved when parallelizing the computations within each iteration.

Subject matter of future research is to investigate if the reconstruction can be improved by further incorporating weights based on confidence measurements of the computed disparities into the data fidelity term. Furthermore, we currently develop a disparity estimator explicitly designed for sparse estimation which also delivers confidence measurements. As we do not pose any requirements on how

the disparities are computed, applications like single image 2D-3D conversion where disparity cues are sparsely spread over an image could benefit from our approach.

Acknowledgements

This work has been partially supported by the Cluster of Excellence *CoTeSys* - Cognition for Technical Systems, funded by the Deutsche Forschungsgemeinschaft (DFG).

References

- [1] S. Becker, J. Bobin, and E. Candès. Nesta: a fast and accurate first-order method for sparse recovery. *SIAM Journal on Imaging Sciences*, 4(1):1–39, 2009. 3
- [2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001. 1
- [3] E. J. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, 2007. 2
- [4] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006. 2
- [5] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006. 2
- [6] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986. 3
- [7] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic Decomposition by Basis Pursuit. *SIAM Journal on Scientific Computing*, 1999. 2
- [8] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. 2
- [9] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. *Proceedings of the National Academy of Sciences of the United States of America*, 100(5):2197–2202, 2003. 2
- [10] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001. 2, 3
- [11] G. Egnal and R. P. Wildes. Detecting binocular half-occlusions: empirical comparisons of five approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1127–1133, 2002. 1
- [12] M. Elad. Optimized projections for compressed sensing. *IEEE Transactions on Signal Processing*, 55(12):5695–5702, 2007. 3
- [13] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70(1):41–54, 2006. 1
- [14] A. Fusiello, V. Roberto, and E. Trucco. Efficient stereo with multiple windowing. In *Proceedings of the 10th IEEE Con-*

Algorithm	Rank Avg.	Tsukuba			Venus			Teddy			Cones		
		nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
[29]	42.00	1.38	1.85	6.90	0.71	1.19	6.13	7.88	13.30	18.60	3.97	9.79	8.26
CSC	43.20	1.64	2.18	8.84	0.31	0.68	4.32	7.22	12.60	19.40	4.33	10.40	11.60
RDT	84.60	4.57	5.52	24.1	1.18	1.70	16.0	10.9	16.50	29.70	9.27	15.60	24.40

Table 2: Excerpt of the Middlebury evaluation table, comparing our algorithm (CSC) with [29] and RDT.

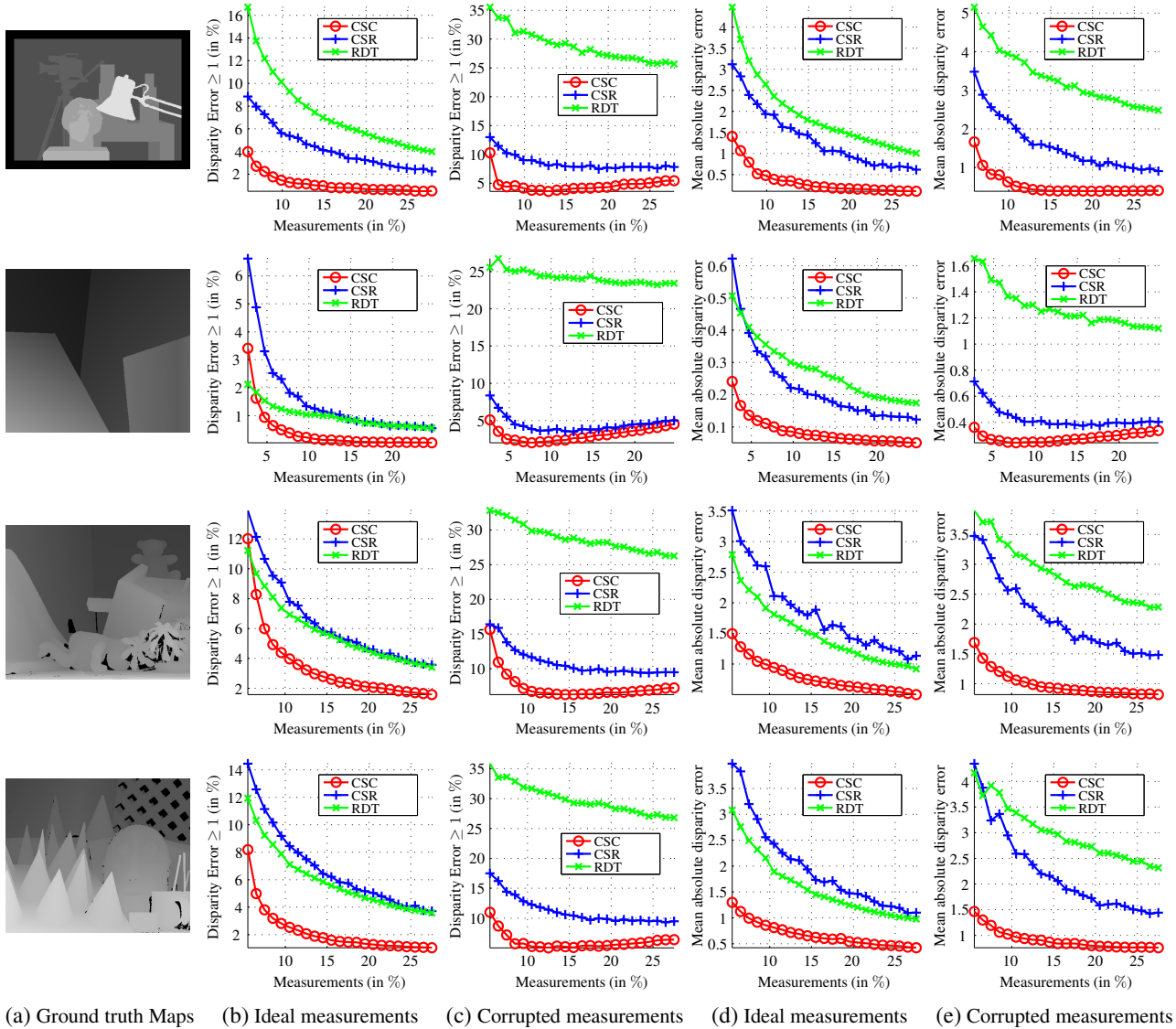


Figure 2: Comparison of reconstruction by interpolation via Delaunay triangulation (RDT) with our CS based reconstruction algorithm using Canny Edge detector (CSC) and our CS based reconstruction algorithm using random sampling (CSR). In column (a): ground truth images; (b) and (c): the percentage of pixels whose reconstructed value differs by more than one from its true value, (b) with ground truth samplings and (c) from corrupted measurements; (d) and (e): the mean absolute disparity error over the entire disparity map, (d) with ground truth samplings and (e) from corrupted measurements.

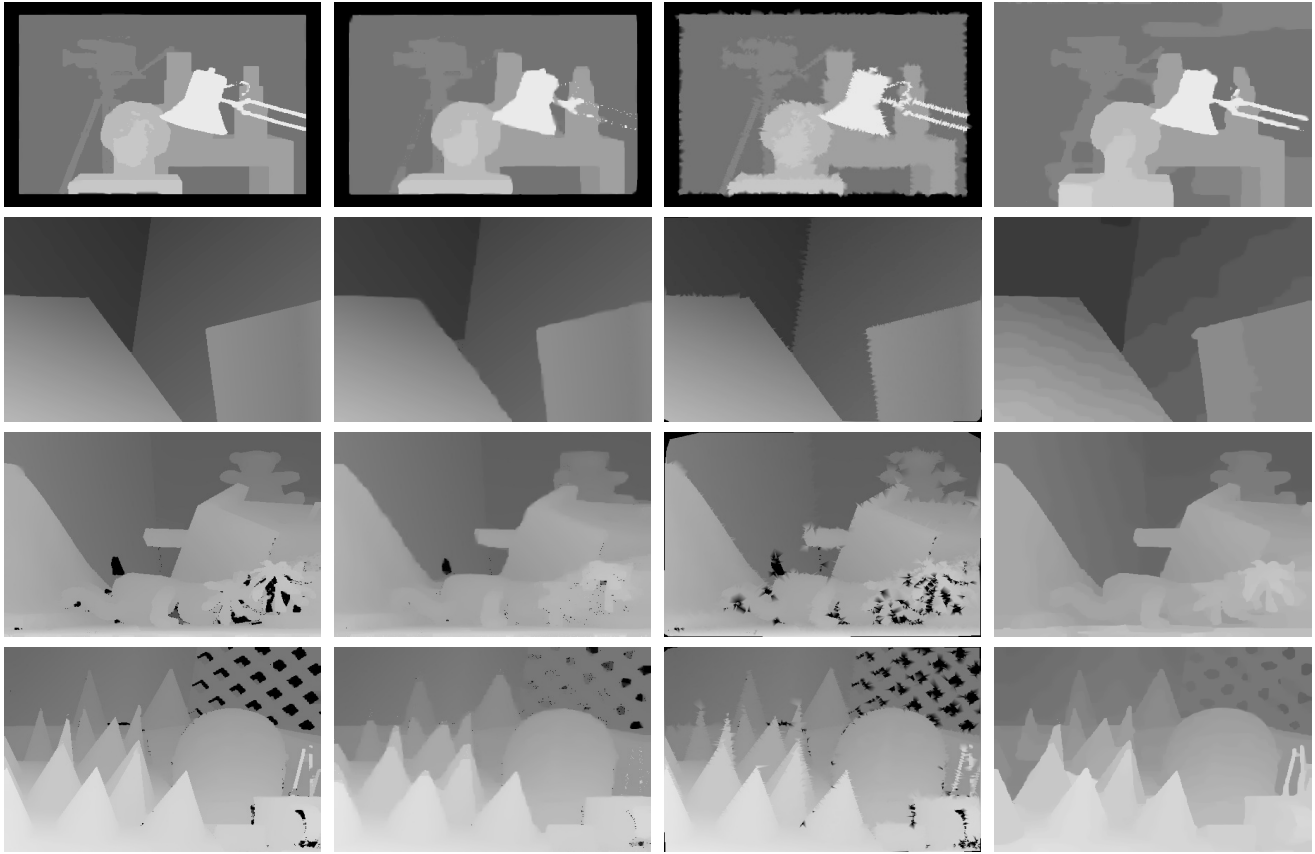


Figure 3: Reconstruction of the four test disparity maps from 5% measurements: CSC (1st column), CSR (2nd column), RDT (3rd column), reconstruction from computed disparities (4th column).

- [16] H. Hirschmüller and D. Scharstein. Evaluation of cost functions for stereo matching. In *Proceedings of the 20th IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007. 1
- [17] A. Klaus, M. Sormann, and K. Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In *18th International Conference on Pattern Recognition*, pages 15–18, 2006. 1
- [18] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *Proceedings of the 7th European Conference on Computer Vision*, pages 82–96, 2002. 1
- [19] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007. 3, 4, 5
- [20] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, 24(2):227–234, 1995. 2
- [21] Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–152, 2005. 3
- [22] J. Nocedal and S. J. Wright. *Numerical Optimization*, 2nd Ed. Springer, New York, 2006. 5
- [23] Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):139–154, 1985. 1
- [24] M. Sarkis and K. Diepold. Sparse stereo matching using belief propagation. In *IEEE International Conference on Image Processing (ICIP)*, pages 1780–1783, 2008. 1
- [25] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, 2002. 1, 5
- [26] J. Sun, N.-N. Zheng, and H.-Y. Shum. Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):787–800, 2003. 1
- [27] N. Vaswani and W. Lu. Modified-CS: modifying compressive sensing for problems with partially known support. *IEEE Transactions on Signal Processing*, 58(9):4595–4607, 2010. 4
- [28] P. Wolfe. A method of conjugate subgradients for minimizing nondifferentiable functions. In *Nondifferentiable Optimization*, volume 3 of *Mathematical Programming Studies*, pages 145–173. 1975. 4
- [29] K.-J. Yoon and I. S. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):650–656, 2006. 1, 6, 7