2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)

SPARSE RECONSTRUCTION FOR DISPARITY MAPS USING COMBINED WAVELET AND CONTOURLET TRANSFORMS

Lee-kang Liu and Truong D. Nguyen*

University of California, San Diego
Electrical and Computer Engineering
http://videoprocessing.ucsd.edu

ABSTRACT

Disparity estimation is a key component in 3D image processing, yet dense estimation is a computationally intensive task. In this paper, we propose to estimate the dense disparities from a small set of spatial measurements. Observing that disparity maps mainly contain contours and smooth regions, we formulate the problem as a sparse reconstruction problem using a combined wavelet and contourlet bases. We show that the combined transform yields better reconstruction results than existing methods.

Index Terms— Sparse reconstruction, dense disparity estimation, wavelet, contourlet, combined transform, conjugate subgradient

1. INTRODUCTION

As 3D technology is getting mature, depth data has been widely used for many applications, e.g., 3D model construction [10], online fitting rooms [4], etc. To acquire depth information, a variety of devices have been developed (e.g., Kinect, time-of-flight camera, LiDAR). In addition to these *direct* acquisition methods, an *indirect* method is to estimate disparities from a pair of stereo images using computational methods. Although direct acquisition methods are faster, these devices are more expensive than stereo cameras. However, indirect depth estimation is sensitive to illumination, noise, stereo camera alignments, and other camera factors. Although some feature points are relative robust, they are usually sparse. For instance, image corners and blob shape features are used for localization applications [8], [11]. According to the characteristic of distinct local information, these feature points are comparatively robust during matching process. Since the number of feature points is typically small, it is an interesting research problem of how to estimate a dense disparity map from sparse samples.

Recently, Elad et al. proposed a well-known method, K-SVD, that reconstructs an image from sparse samples, utilizing a learned overcomplete dictionary [1], [15]. K-SVD has been extended for color image denoising, face image recovery and image inpainting [7], [12]. However, unlike natural images, disparity maps, which encode depth information, have simple structures and possess piecewise smooth regions. Taking advantage of these properties, several works studied the problem of reconstructing dense disparity maps from sparse measurements [2], [3], [6]. Hawe et al. proposed a model that reconstructs dense disparity maps by considering sparsity in wavelet transform domain [9]. Although the authors claim that disparity maps can be reconstructed from sparse samples, the sampling points are limited to be on edge locations. As the sampling locations are not restricted, the resulting disparity map has staircase artifact along object boundaries. Besides edges and object boundaries, disparity map also has abundant smooth regions. Our proposed method uses both wavelet and contourlet to preserve structures within disparity maps. In this paper, we have two contributions.

- We propose a method for reconstructing dense disparity maps from sparse samples by utilizing combined wavelet and contourlet transforms.
- Experimentally, we justify that wavelet has better representation for local set of points, and contourlet has better representation for image contours.

The paper is organized as follows. After reviewing existing methods for disparity reconstruction in Section 2, we present the proposed method in Section 3. Experimental results are shown in Section 4. Finally, we summarize the paper and address our future work in Section 5.

2. SPARSE REPRESENTATION AND RECONSTRUCTION OF DISPARITIES

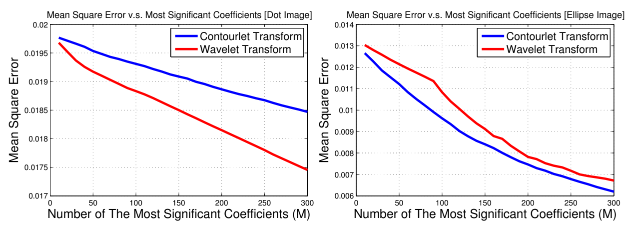
2.1. Representations of Disparities

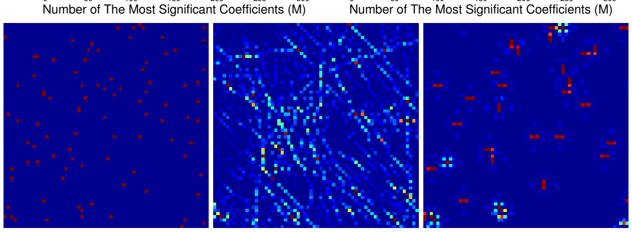
Wavelet and contourlet transforms are widely used in image processing applications since images can be represented by a few coefficients in both transform domains. However, the difference between wavelet and contourlet is that wavelet transform has square compact support, whereas contourlet has di-

*Corresponding author: L. Liu, email: `l7liu@ucsd.edu`. This work is partially supported by NSF grant CCF-1065305

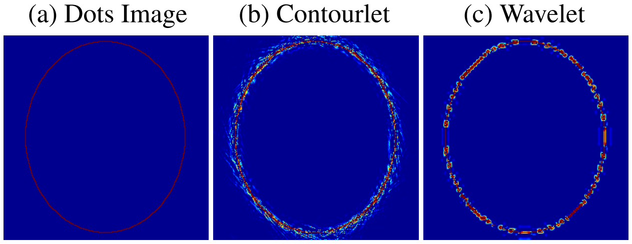
978-1-4799-2893-4/14/$31.00 ©2014 IEEE

rectional rectangular compact support. By applying directional filter banks, contourlet is suitable for analysing image contours [5], [14]. Fig. 1 summarizes a study of how efficient wavelet and contourlet perform for localized set of singular points (dots) and contours. The columns show the image (first column), its reconstruction using a few large coefficients for contourlets (second column) and for wavelets (third column). As observed, the wavelet transform has better reconstruction for localized area, whereas the contourlet is especially suited for contours. Fig. 2 compares the reconstruction quality for disparity map and color image of the Aloe image. As observed, both wavelets and contourlets perform well for disparity maps.





(a) Dots Image (b) Contourlet (c) Wavelet



(d) Ellipse Image (e) Contourlet (f) Wavelet

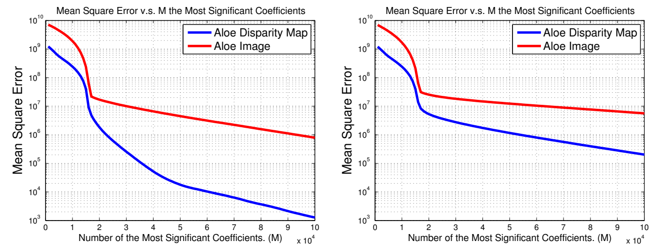
Fig. 1: Mean square error v.s M the most significant coefficients. Erroneous line structures appear in (b) and result in higher MSE than (c). Set of singular points results in non-efficient representation for image contours (f), hence wavelet have higher MSE than contourlet for ellipse image.

2.2. Sparse Reconstruction of Disparities

From compressed sensing theory, sparsity in transform domain corresponds to the number of samples in spatial domain for image reconstruction [3]. Given transform bases, $\mathbf{M} \in \mathbb{R}^{n\times n}$, the minimum number of samples, m, in spatial domain is.

$$m \geq \mathbf{C} \cdot \mu^2(\mathbf{M}) \cdot \|\mathbf{x}\|_0 \cdot \log{(n)} \tag{1}$$

where the ℓ_0 norm indicates the non-zero elements in transform coefficients, $\mathbf{x} \in \mathbb{R}^{n\times 1}$. The variable $\mu(\mathbf{M})$ represents



(a) Wavelet (b) Contourlet

Fig. 2: Mean square error v.s. M the most significant coefficients. Both wavelet and contourlet has better efficiency in representing disparity maps.

the mutual coherence of $\mathbf{M}$, and it is defined as, $\mu(\mathbf{M}) = \sqrt{n} \cdot \max_{i,j}|\mathbf{M}(i,j)|$.

Before setting up the problem formulation, we first describe the notation used. Variables $\mathbf{B}_1 \in \mathbb{R}^{n\times n}$ and $\mathbf{B}_2 \in \mathbb{R}^{n\times n}$ are wavelet and contourlet bases. $\mathbf{W}_1 \in [0,1]^{n\times n}$ and $\mathbf{W}_2 \in [0,1]^{n\times n}$ are two diagonal matrices with zeros at the locations of approximation coefficients and ones at the locations of detail coefficients. $\mathbf{S} \in [0,1]^{m\times n}$ is a sampling matrix. Variables $\mathbf{b} \in \mathbb{R}^{m\times 1}$ and $\mathbf{u} \in \mathbb{R}^{n\times 1}$ denote the observations (sparse samples) and disparity map, respectively.

Hawe et al. proposed a model for dense disparity map reconstruction by utilizing wavelet transform [9]. Since disparity maps have piecewise smooth regions, for preserving the discontinuity, the authors introduced *total variation* as a prior. Moreover, since the approximation of total variation results in bias in low frequency components and the sparsity exists in detail coefficients, the authors applied the weighting matrix $\mathbf{W}_1$ for discarding approximation coefficients, and the problem is described as follows.

$$\begin{aligned} \underset{\mathbf{u}}{\text{minimize}} \quad & \|\mathbf{W}_1\mathbf{B}_1\mathbf{u}\|_1 + \beta\|\mathbf{u}\|_{TV} \\ \text{subject to} \quad & \mathbf{b} = \mathbf{S}\mathbf{u} \end{aligned} \tag{2}$$

Due to the reasons that disparity maps have sparsity in both wavelet and contourlet transform domains, wavelet has better representation for localized features, and contourlet has better representation for contours, we propose to use both wavelet and contourlet transforms for dense disparity map reconstruction.

$$\begin{aligned} \underset{\mathbf{u}}{\text{minimize}} \quad & \|\mathbf{W}_1\mathbf{B}_1\mathbf{u}\|_1 + \|\mathbf{W}_2\mathbf{B}_2\mathbf{u}\|_1 + \beta\|\mathbf{u}\|_{TV} \\ \text{subject to} \quad & \mathbf{b} = \mathbf{S}\mathbf{u} \end{aligned} \tag{3}$$

According to the Lagragian duality and the equivalent form, $\mathbf{u} = \mathbf{B}_1^{-1}\mathbf{x}$, the problem can be reformulated as an unconstrained minimization problem.

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{b} - \mathbf{S}\mathbf{B}_1^{-1}\mathbf{x}\|_2^2 + \lambda\|\mathbf{W}_1\mathbf{x}\|_1 \\ & + \gamma\|\mathbf{W}_2\mathbf{B}_2\mathbf{B}_1^{-1}\mathbf{x}\|_1 + \beta\|\mathbf{B}_1^{-1}\mathbf{x}\|_{TV} \end{aligned} \tag{4}$$

3. PROPOSED ALGORITHM

For solving the unconstrained minimization problem, similar approach is discussed in [9]; however, the main difference is that we consider both wavelet and contourlet transforms. Referring to Algorithm 1, finding gradient of each terms of (4) is the first step, and the gradient of $\|\mathbf{W}_2\mathbf{B}_2\mathbf{B}_1^{-1}\mathbf{x}\|_1$ at location k is,

$$\partial_{\mathbf{x}}\|\mathbf{W}_2\mathbf{B}_2\mathbf{B}_1^{-1}\mathbf{x}\|_1(k) = \left\{\mathbf{B}_1\mathbf{B}_2^{-1}\text{sign}\left[\mathbf{W}_2\mathbf{B}_2\mathbf{B}_1^{-1}\mathbf{x}\right]\right\}(k) \tag{5}$$

where the definition of the sign function is,

$$\text{sign}\,(v) = \begin{cases} 1 & \text{if} \quad v > 0 \\ 0 & \text{if} \quad v = 0 \\ -1 & \text{if} \quad v < 0 \end{cases} \tag{6}$$

Since the anisotropic total variation is equivalent to ℓ_1 norm, by introducing difference operator $\mathbf{D} = [\mathbf{D}_x, \mathbf{D}_y]^T$, we can rewrite $\|\cdot\|_{TV}$ as,

$$\begin{aligned}\|\mathbf{s}\|_{TV} &= \|\mathbf{D}\mathbf{s}\|_1 \\ &= \sum_{x,y}\left[\sqrt{|\mathbf{s}_{x+1,y} - \mathbf{s}_{x,y}|^2} + \sqrt{|\mathbf{s}_{x,y+1} - \mathbf{s}_{x,y}|^2}\right]\end{aligned} \tag{7}$$

Therefore, the gradient of $\|\mathbf{B}_1^{-1}\mathbf{x}\|_{TV}$ is,

$$\partial_{\mathbf{x}}\|\mathbf{B}_1^{-1}\mathbf{x}\|_{TV}(k) = \partial_{\mathbf{x}}H_\delta\left(\sqrt{\left(\mathbf{e}_k\mathbf{D}_x\mathbf{B}_1^{-1}\mathbf{x}\right)^2}\right) + \partial_{\mathbf{x}}H_\delta\left(\sqrt{\left(\mathbf{e}_k\mathbf{D}_y\mathbf{B}_1^{-1}\mathbf{x}\right)^2}\right) \tag{8}$$

where $\mathbf{e}_k$ is a vector with 1 at location k, and 0 otherwise.

$$\partial_{\mathbf{x}}H_\delta\left(\sqrt{\mathbf{s}_x^2}\right) = \begin{cases} \mathbf{B}_1\mathbf{D}_x^T\mathbf{e}_k^T\text{sign}(\mathbf{s}_x) & \text{if} \quad |\mathbf{s}_x| \geq \delta \\ \dfrac{\mathbf{B}_1\mathbf{D}_x^T\mathbf{e}_k^T\mathbf{s}_x}{\delta} & \text{otherwise} \end{cases} \tag{9}$$

and

$$\partial_{\mathbf{x}}H_\delta\left(\sqrt{\mathbf{s}_y^2}\right) = \begin{cases} \mathbf{B}_1\mathbf{D}_y^T\mathbf{e}_k^T\text{sign}(\mathbf{s}_y) & \text{if} \quad |\mathbf{s}_y| \geq \delta \\ \dfrac{\mathbf{B}_1\mathbf{D}_y^T\mathbf{e}_k^T\mathbf{s}_y}{\delta} & \text{otherwise} \end{cases} \tag{10}$$

where $\mathbf{s}_x = \mathbf{e}_k\mathbf{D}_x\mathbf{B}_1^{-1}\mathbf{x}$ and $\mathbf{s}_y = \mathbf{e}_k\mathbf{D}_y\mathbf{B}_1^{-1}\mathbf{x}$. Finally, the conjugate subgradient of $\|\mathbf{W}_1\mathbf{x}_1\|_1$ is as follows

$$\bigtriangledown\|\mathbf{W}_1\mathbf{x}\|_1 = \sum_k \begin{cases} \text{sign}(\left[\mathbf{W}_1\mathbf{x}\right](k)) & \text{if} \quad |\mathbf{W}_1\mathbf{x}|(k) \neq 0 \\ -\,\text{sign}(\mathbf{q}(k)) \cdot \min\{\text{abs}(\mathbf{q}(k)), 1\} & \text{otherwise} \end{cases} \tag{11}$$

where the variable $\mathbf{q}$ is,

$$\begin{aligned}\mathbf{q} = \frac{1}{\lambda}\{&-\mathbf{B}_1^{-T}\mathbf{S}^T\left(\mathbf{b} - \mathbf{S}\mathbf{B}_1^{-1}\mathbf{x}\right) + \gamma \bigtriangledown \|\mathbf{W}_2\mathbf{B}_2\mathbf{B}_1^{-1}\mathbf{x}\|_1 \\ &+ \beta \bigtriangledown \|\mathbf{B}_1^{-1}\mathbf{x}\|_{TV}\}\end{aligned} \tag{12}$$

Algorithm 1 Disparity Reconstruction Algorithm

procedure DENSERECONSTRUCTIONWTCT($\mathbf{b}, \mathbf{S}$)
  initialization
  $\mathbf{x}_0 = \mathbf{B}_1^{-T}\mathbf{S}^T\mathbf{b}$, $\mathbf{h}_0 = -\mathbf{d}_0$, $\alpha_{-1} = 1$, $i = 0$
  while *not converge* **do**
    $\mathbf{d}_{i+1}$ = Subgradient of (3) using (5), (8) and (11).
    α_i = BacktrackingLineSearch$(\mathbf{x}_i, \alpha_{i-1})$
    $\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i\mathbf{h}_i$
    $\mathbf{h}_{i+1} = -\mathbf{d}_{i+1} + \frac{\mathbf{d}_{i+1}^T(\mathbf{d}_{i+1} - \mathbf{d}_i)}{\mathbf{h}_i^T(\mathbf{d}_{i+1} - \mathbf{d}_i)}$
  end while
end procedure

As the gradients are calculated, this unconstrained minimization problem can be solved by directional gradient descent. Referring to Algorithm 1, the step size, α_i, is estimated by backtracking line-search algorithm [13], and we use Hestenes-Stiefel (HS) method for updating the directional gradient, $\mathbf{h}_i$. Finally, the overall algorithm for dense disparity reconstruction is shown in Algorithm 1.

4. EXPERIMENTS AND RESULTS

4.1. Experimental Setup

In our experiment, we test disparity maps from Middlebury dataset (vision.middlebury.edu/stereo/). The ground truth disparities and reconstructed results are shown in Fig. 3. The size of each disparity maps is 512×512, and the range of disparity values is [0-255]. Regarding the parameters, we use *"db2"* and *level*=2 in wavelet transform, and choose *nLevel* = [5, 6] for contourlet transform. In addition, the regularization parameters are $\lambda = 0.01$, $\gamma = 0.01$ and $\beta = 0.5$. We examine reconstruction performance by randomly selecting 5%, 10%, 15%, 20% and 25% sampling points. Two comparisons are presented in our experiment. Besides the model proposed in [9], we also consider the case that only uses contourlet transform. The PSNR and mean absolute error (MAE) are presented to evaluate the performance of the algorithms.

Given an estimated image $\hat{\mathbf{I}}$, ground truth image $\mathbf{I}$ and total number of pixels, N, the definition of MAE is:

$$\text{Mean Absolute Error} = \frac{1}{N}\sum_{i,j}|\mathbf{I}(i,j) - \hat{\mathbf{I}}(i,j)| \tag{13}$$

4.2. Discussions

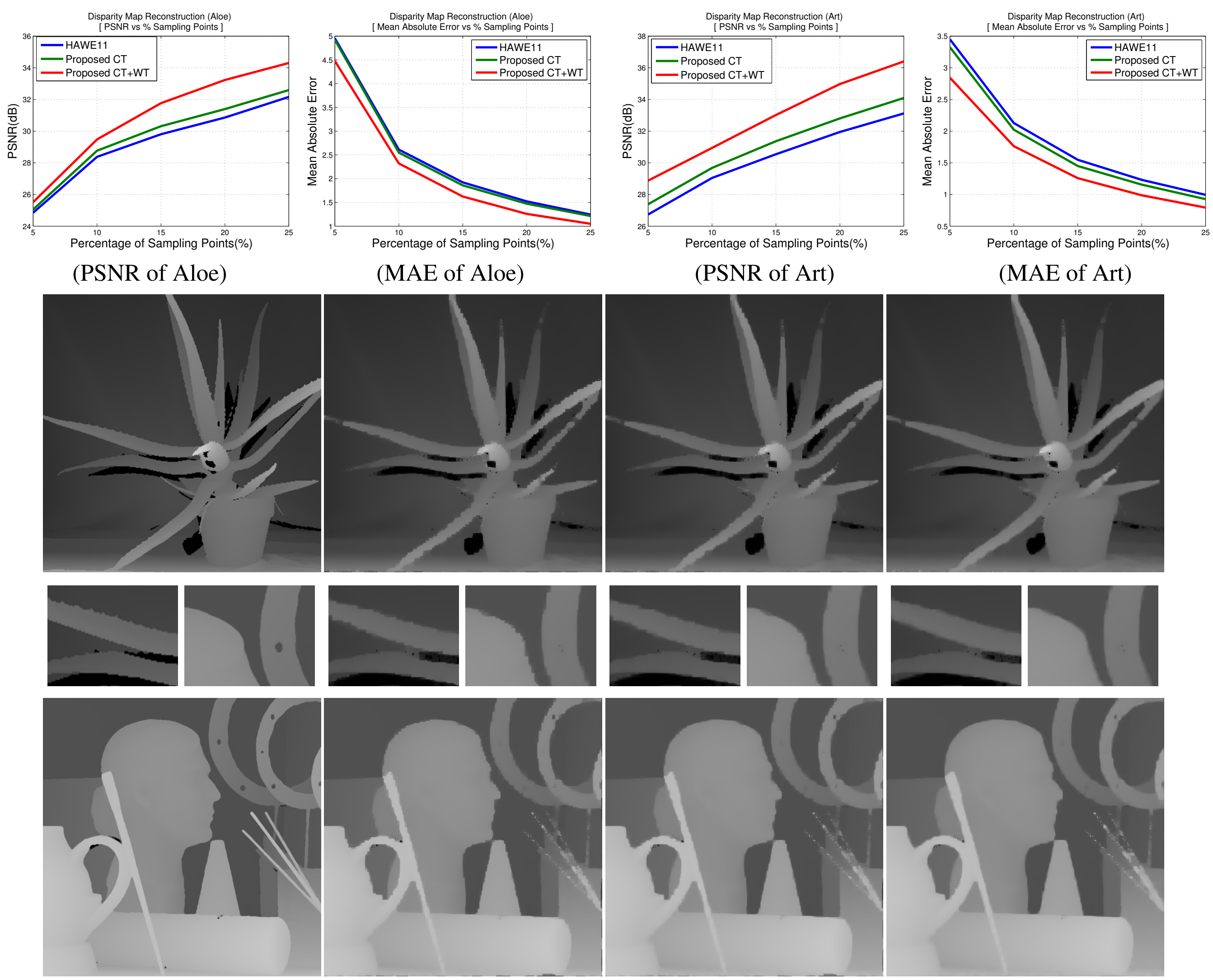
As shown in the first row of Fig. 3[1], the proposed CT+WT method has the highest PSNR. Since the proposed CT+WT method utilizes both contourlet and wavelet transforms, local features and contours are reconstructed with high quality. Additionally, the proposed CT method has higher PSNR than

[1]For more experimental results please refer to our research website: http://videoprocessing.ucsd.edu/~LeeKang/research.html

HAWE'11 because the major structures of disparity maps are contours. Referring to MAE curves, the proposed CT+WT method outperforms the proposed CT and HAWE'11. As disparity maps correspond to depth information, the lower MAE infers better dense depth estimation performance. Thus, the proposed CT+WT method not only reconstructs disparity structures but also has less depth errors. While visually comparing object boundaries, the proposed CT+WT and proposed CT have smooth boundaries, whereas the staircase artifact along object boundaries exists in HAWE'11 method (Row 3 of Fig. 3). In summary, the experiment shows that using combined wavelet and contourlet transforms has better reconstruction performance than utilizing either wavelet or contourlet transform.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a sparse reconstruction algorithm that utilizes both wavelet and contourlet transforms, and demonstrate that wavelet has better representation for localized features and contourlet has better representation for object edges. The experimental results show that the proposed CT+WT method outperforms the proposed CT and HAWE'11 methods since contours and singular points are two components that exist in disparity maps.



(a) Ground Truth (b) HAWE'11[9] (c) Proposed CT (d) Proposed CT + WT

Fig. 3: Reconstructed disparity maps from 10% random samples. (Row 1) Evaluations, (Row 2) Aloe, (Row 3) Zoom in of Aloe and Art, (Row 4) Art. Referring to Row 3, results from proposed CT + WT have smooth object boundaries (Art) and better representations for zigzag leaves (Aloe).

6. REFERENCES

[1] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311 –4322, Nov. 2006.

[2] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.*, 51(1):34–81, February 2009.

[3] E. J. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, November 2006.

[4] F. Cordier, Hyewon S., and N. Magnenat-Thalmann. Made-to-measure technologies for an online clothing store. *Computer Graphics and Applications, IEEE*, 23(1):38–48, 2003.

[5] M.N. Do and M. Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *Image Processing, IEEE Transactions on*, 14(12):2091–2106, 2005.

[6] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001.

[7] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *Image Processing, IEEE Transactions on*, 15(12):3736–3745, 2006.

[8] R.S. Feris, J. Gemmell, K. Toyama, and V. Kruger. Hierarchical wavelet networks for facial feature localization. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 118–123, 2002.

[9] S. Hawe, M. Kleinsteuber, and K. Diepold. Dense disparity maps from sparse disparity measurements. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2126–2133, 2011.

[10] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and Andrew Fitzgibbon. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *In Proc. UIST*, pages 559–568, 2011.

[11] Y. Ke and R. Sukthankar. Pca-sift: a more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–506–II–513 Vol.2, 2004.

[12] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *Image Processing, IEEE Transactions on*, 17(1):53–69, 2008.

[13] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, 2nd edition, 2006.

[14] D.D.-Y. Po and M.N. Do. Directional multiscale modeling of images using the contourlet transform. *Image Processing, IEEE Transactions on*, 15(6):1610–1620, 2006.

[15] R. Rubinstein, T. Peleg, and M. Elad. Analysis k-svd: A dictionary-learning algorithm for the analysis sparse model. *Signal Processing, IEEE Transactions on*, 61(3):661–677, 2013.