

# MULTI IMAGE SUPER RESOLUTION USING COMPRESSED SENSING

Torsten Edeler<sup>1</sup>, Kevin Ohliger<sup>1</sup>, Stephan Hussmann<sup>1</sup>, and Alfred Mertins<sup>2</sup>

<sup>1</sup> Westcoast University of Applied Sciences (FHW), Heide, 25746 Germany

<sup>2</sup> Institute for Signal Processing, University of Lübeck, 23538 Germany

## ABSTRACT

In this paper we present a new compressed sensing model and reconstruction method for multi-detector signal acquisition. We extend the concept of the famous single-pixel camera to a multi-detector device with the benefit of reducing measurement time, while still providing resolution enhancement and deblurring. We provide a scalable model which allows the trade off between system complexity (number of detectors) and time (number of measurements).

We test our model on simulated sparse and compressible data and show convergence using the proposed reconstruction method. We also show that our model allows for significant reduction of necessary measurements.

A real-live setup for data acquisition according to the new model is presented and we show successful reconstruction of the acquired data. With this setup it is possible to acquire super-resolution images with a low resolution camera. The measurements can also be corrupted by a considerable amount of blurring and noise.

**Index Terms**— Single pixel camera, Compressed sensing, Wavelet, Super resolution, Deconvolution, Deblurring

## I. INTRODUCTION

The Shannon sampling theorem gives a limit for the resolution of sampling systems due to its statement that a signal's information can be perfectly reconstructed if it is sampled uniformly at a sampling rate which is at least twice the maximum signal frequency. In recent years, a theory named Compressive Sensing (CS) emerged which overcomes the limitations that Shannon's theorem imposes on sampling systems. CS is motivated by the fact that most natural signals are sparse or at least approximately sparse in a certain basis like, for example, a wavelet or Fourier basis. CS exploits the signal compressibility during the sampling process by measuring a few informative signal parts directly and therefore makes it possible to reduce the sampling rate drastically.

Assume a signal vector  $\mathbf{x} \in \mathbb{R}^N$  has a K-Sparse representation  $\mathbf{x} = \Psi\mathbf{c}$  in the basis  $[\psi_n]_1^N = \Psi$ , which is named representation matrix in this paper. The rows of the matrix  $\Psi$  have an (bi-)orthogonal relationship with the columns of the matrix  $\tilde{\Psi}$ , which is named sparsifying matrix ( $\Psi\tilde{\Psi} = \mathbf{I}$ ). K-sparse means that  $\|\mathbf{c}\|_0 \leq K$ , where  $\|\cdot\|_0 : \mathbb{R}^N \rightarrow \mathbb{R}$  counts non-zero entries and  $0 < K \leq N$ .

Sampling  $M$  projections of  $\mathbf{x}$  leads to the model

$$\mathbf{y} = [(\phi_m, \mathbf{x})]_M^1 + \mathbf{v} = \Phi\mathbf{x} + \mathbf{v} = \Phi\Psi\mathbf{c} + \mathbf{v}, \quad (1)$$

where  $\Phi$  is the *system sensing matrix* as it represents the behavior of the measurement system. Together with the *representation matrix* they form the *sampling matrix*  $\Phi\Psi$  for the vector  $\mathbf{c}$  which is the K-sparse representation of  $\mathbf{x}$ . Observation noise is taken into account by  $\mathbf{v}$ .

Compressive sensing theory tells us that, in certain situations, there exists a unique solution to the inverse problem of finding  $\mathbf{c}$  given  $M < N$  measurements. While it has been proven that

finding  $\mathbf{c}$  by brute force methods is NP-Hard, the problem can be reformulated as a convex program

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} \|\mathbf{c}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \Phi\Psi\mathbf{c}\|_2 \leq \epsilon, \quad (2)$$

where  $\|\mathbf{c}\|_1 = \sum_{i=1}^N |c_i|$  is the  $\ell_1$ -norm and  $\epsilon$  bounds the amount of noise in the observation data. Providing  $\|\mathbf{c}\|_0 \leq K$  and a noiseless sampling process ( $\epsilon = 0$ ), Candes et al. [1] showed that exact recovery of  $\mathbf{c}$  is possible if the *Restricted Isometry Property* (RIP) [2] holds for the sampling matrix.

The Restricted Isometry Property and more specifically the *Restricted Isometry Constant* (RIC)  $\delta_K$  as matrix quantity are very important in CS context. Many algorithms' error boundings rely on the RIC. The RIP imposes the following condition on the sampling matrix:

$$(1 - \delta_K)\|\mathbf{c}\|_2^2 \leq \|\Phi\Psi\mathbf{c}\|_2^2 \leq (1 + \delta_K)\|\mathbf{c}\|_2^2 \quad (3)$$

The RIC  $\delta_K$  is the smallest constant for which the inequalities hold, given all possible K-sparse signals  $\mathbf{c}$ . The RIP is said to *hold* if  $\delta_{2K} < 1$ , which implies unique reconstruction [2]. If  $\delta_K \ll 1$  then the sampling matrix nearly maintains the  $\ell_2$  distance of different vectors  $\mathbf{c}$ , which implies good reconstruction (inverting) behavior. Unfortunately determining a RIC of a given sampling matrix is an NP-Hard problem [3] and therefore it is not practical to choose  $\Phi$  to be a structured matrix. But if the elements of  $\Phi$  are independently drawn from a random distribution like Gaussian or Bernoulli, we obtain (given an (almost) orthonormal representation matrix  $\Psi$ ) a RIC of  $\delta_K \ll 1$ . Prerequisite is a sufficiently high number measurements of  $M = CK \log(N/K)$ , where  $K$  is the expected sparsity level,  $N$  is the number of signal components, and  $C$  is an arbitrary but small constant which may change from appearance to appearance.

In real-life situations, signals are (i) noisy and (ii) not sparse, but compressible [4], which means that there exists a good K-Sparse approximation. For such signals,  $\epsilon$  in (2) depends on the observation noise level and the desired sparsity. The smaller the RIC for the sensing system, the smaller is the reconstruction error for those signals.

In this paper we extend the traditional compressive sensing model to a multi-detector model, which covers the simultaneous measurements of different but similar mixtures of a single source signal. The detectors are placed on an arbitrary grid on the source signal and each detector integrates the source signals only locally. The motivation for this new model is derived from applications where arrays of detectors are available but due to either size or cost, resolution is very low. One example are time-of-flight cameras where each pixel is a signal correlator and therefore needs a significant amount of area on the silicon for the electronic circuitry. In addition to this motivation, which has been used to justify traditional compressed sensing system like the single-pixel camera [5] before, our new model allows the reduction of required measurements (or measurement time) by using more detectors. This has significant importance because it allows the trade off between number of detectors (system costs or complexity) against measurement time. However, introducing more than one detector

is not a trivial extension to the single-pixel camera, where each element can be treated on its own. The detector signals might be correlated caused by crosstalk, which has to be considered during reconstruction.

Using multiple detectors in compressive sensing context is not new and has been used in [6] to reconstruct a single source image by assuming sparse representation of single image tiles using over complete dictionaries. In [7] the authors used several audio receivers to reconstruct a source signal. They assumed the measurements to be a shifted and filtered version of the original signal. In [8] the authors present a general framework for related but not identical signals supplied by several detectors. They introduce a structured sparsity model and assume the detector signals to be represented by such a model.

Even though our model also allows several detectors capturing only local, but possibly correlated parts of a source signal we do not assume a sparse representation of the individual detector signals but a sparse representation of the whole source signal. To test our new model we built up a test setup for compressive image acquisition similar to that in [5] but with a multi detector receiver. We also present simulations to show convergence and the trade off between system complexity and number of measurements.

In the next section we will introduce our new multi detector compressive sensing model, in Section III we show an approach of signal reconstruction using this model. Section IV presents experimental and simulation results.

## II. SAMPLING MODEL

We assume a sampling system where we have control over a weighting element, which is able to multiply the original source signal by a predefined pattern. In our experimental setup this weighting element is a digital mirror device (DMD) with thousands of controllable mirrors. Further we assume that the weighted original signal is captured by one or more detectors either capturing the whole or parts of the weighted signal. With no loss of generality we assume the signals to have two dimensions and call them images from now on. For the sake of better readability we represent the images in vector form by lexicographic ordering.

Our model covers the acquisition of  $l^2$  different (but possibly similar) linear mixtures of an original image:

$$\mathbf{y}_k = \hat{\mathbf{P}}\mathbf{H}\mathbf{F}_k\mathbf{x} + \mathbf{v}_k \quad \forall k \in \{1 \dots M\}, \quad (4)$$

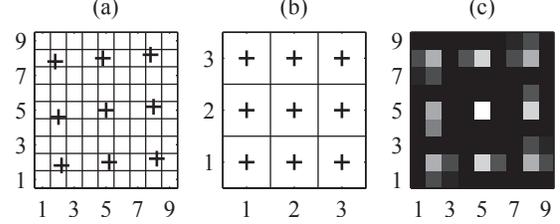
where  $\mathbf{x} \in \mathbb{R}^{[N \times 1]}$  is the high-resolution image and  $\mathbf{y}_k \in \mathbb{R}^{[l \times 1]}$  is the  $k$ -th measurement of  $l^2$  mixtures. Matrix  $\mathbf{F}_k$  is diagonal and includes the weights for each element in  $\mathbf{x}$ . The matrix  $\hat{\mathbf{P}}\mathbf{H}$  models the sampling process, where this process is splitted into an FIR convolution operation  $\hat{\mathbf{H}} \in \mathbb{R}^{[N \times N]}$ , modeling the influence of each element in  $\mathbf{F}_k\mathbf{x}$  on each detector and a Dirac sampling operator  $\hat{\mathbf{P}} \in \mathbb{R}^{[l^2 \times N]}$ . Together they can model detectors integrating over different region of interests.  $\mathbf{v}_k$  represents additive zero mean, white Gaussian noise with  $E[\mathbf{v}_k\mathbf{v}_k^T] = \sigma^2\mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix. The sampling operator  $\mathbf{P}$  models the detector sampling grid on the high resolution input image. If applied to an image  $\mathbf{x}$ , the result is a low-resolution image  $\mathbf{y} = \hat{\mathbf{P}}\mathbf{x}$ , where the dimension of  $\mathbf{y}$  and the sampling locations on  $\mathbf{x}$  are defined by the actual setup. The same operator can also be used for upsampling by applying its transposed. Then  $\mathbf{z} = \hat{\mathbf{P}}^T\mathbf{y}$  is a high-resolution representation of  $\mathbf{y}$ , which is also defined by the sampling grid. The sampling values are bilinearly interpolated, if necessary, and locations not covered by the sampling grid are left zero, by the upsampling operator  $\hat{\mathbf{P}}^T$ . Fig.1 illustrates this behavior.

To incorporate all  $M$  measurements into a single model, we extend (4) to

$$\mathbf{y} = \mathbf{P}\mathbf{H}\mathbf{F}\mathbf{x} + \mathbf{v}, \quad (5)$$

where  $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_M^T]^T$  holds all measurements,  $\mathbf{P} = \mathbf{I}^M \otimes \hat{\mathbf{P}}$ ,  $\mathbf{H} = \mathbf{I}^M \otimes \hat{\mathbf{H}}$ , and  $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_M]^T$ . Vector  $\mathbf{v}$  represents Gaussian noise with  $E[\mathbf{v}\mathbf{v}^T] = \sigma^2\mathbf{I}^{M \cdot l^2}$ . The notation  $\mathbf{I}^r$  represents the identity matrix with dimension  $[r \times r]$ . The operator  $\otimes$  denotes the Kronecker matrix product.

The model (5) also covers the model (1) if  $\hat{\mathbf{P}}\hat{\mathbf{H}} = \mathbf{1}^{1 \times N}$ , that is, there is only a single detector integrating over the whole weighted images  $\mathbf{F}_k\mathbf{x}$  with  $k = 1 \dots M$ , where the notation  $\mathbf{1}^{r \times t}$  represents a matrix filled with ones with dimension  $[r \times t]$ .



**Fig. 1.** Illustration of sampling operator  $\hat{\mathbf{P}}$ . Values are bilinearly interpolated, if sampling grid lays on non-integer locations. (a): Original image  $\mathbf{x}$  is sampled by the operator at 9 locations (down-sampling). (b): Result of downsampling operation  $\hat{\mathbf{P}}\mathbf{x}$ . (c): Result of upsampling operation  $\hat{\mathbf{P}}^T\hat{\mathbf{P}}\mathbf{x}$ .

## III. RECONSTRUCTION

Reconstruction of the original signal  $\mathbf{x}$  of length  $N$ , given the measurements  $\mathbf{y}$  as described by model (5) with  $l^2$  detectors and  $M$  measurements is an ill-posed problem for two reasons. First, if the total number of measurements  $M_{tot} = l^2 \cdot M < N$ , (5) forms an underdetermined system of equations. Second, the correlation of the detector signals modeled by the blurring operator  $\hat{\mathbf{H}}$  requires a deconvolution of the measured signals which is also known to be ill-posed.

To recover the original signal  $\mathbf{x}$  from measurements  $\mathbf{y}$ , we minimize objective function

$$\begin{aligned} f(\hat{\mathbf{x}}) &= f_{\text{data}}(\hat{\mathbf{x}}) + \lambda f_{\text{prior}}(\hat{\mathbf{x}}) \\ &= \|\mathbf{y} - \mathbf{P}\mathbf{H}\hat{\mathbf{F}}\hat{\mathbf{x}}\|_2^2 + \lambda \|\hat{\Psi}\hat{\mathbf{x}}\|_1, \end{aligned} \quad (6)$$

where  $f_{\text{data}}(\hat{\mathbf{x}})$  is the data fidelity term which is the  $\ell_2$ -norm of the error between measurements and the estimate  $\hat{\mathbf{x}}$ . Due to the ill-posed nature of the problem we introduce the  $f_{\text{prior}}(\hat{\mathbf{x}})$  which is the  $\ell_1$ -norm of the sparse representation of  $\hat{\mathbf{x}}$ . The regularization constant  $\lambda$  weights the prior against data fidelity. A good representation space, defined by sparsifying matrix  $\hat{\Psi}$ , concentrates the signal's energy to a few components and therefore allows a good sparse approximation. For example smooth signals are good represented in Fourier domain, piecewise constant signals have a good representation in wavelet domain.

Note that (6) leads to the same solution as (2), if  $\mathbf{P}\mathbf{H} = \mathbf{I}^M \otimes \mathbf{1}^{1 \times N}$ , in which case (5) is equivalent to (1).

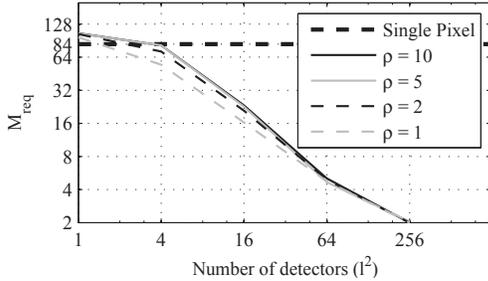
## IV. EXPERIMENTS

In our experiments we show the efficiency of our proposed model in terms of number of detectors and number of measurements trade off. We show that using more detectors reduces the amount of necessary measurements significantly. Simulations on sparse and approximately sparse (compressible) images show the influence of model parameters and a real world test setup which acquires data according to the model (5) is presented as well.

#### IV-A. Simulation for exactly sparse images

In this section we present simulation result for signals which are exactly sparse in a representation domain. We determine required measurements  $M_{req}$  for perfect reconstruction of the original image, which was constructed from the lena-image by scaling it to  $n \times n = 32 \times 32$  pixel and only taking the dominating  $K = 10$  coefficients in 5-3 wavelet domain. The diagonal elements of  $\mathbf{F}_k$  are chosen from a Bernoulli distribution with the elements  $\{-1, 1\}$ . The impulse response  $h$  of the blurring operator  $\hat{\mathbf{H}}$  was set to a 2D Gaussian blurring kernel with variance  $\rho n/l$  and size  $(2n/l + 1)^2$ , where we chose different values for  $\rho$  in the experiment. The experiments were repeated 30 times. Mean number of required measurements  $\overline{M}_{req}$  are plotted in Fig. 2. For comparison purpose, we also plotted the minimal measurements needed for the traditional model (1) as dashed line.

The figure shows that the required number of measurements for perfect reconstruction decreases with increasing number of detectors, thus showing the trade off between system complexity and measurement time. It also shows a minor dependency of  $\overline{M}_{req}$  from the amount of blurring, where the  $\overline{M}_{req}$  tends to be lower with less blurring.



**Fig. 2.** Number of measurements  $M_{req}$  necessary for perfect reconstruction using  $l^2$  detector elements. The experiments involved a ground truth image of size  $n \times n = 32 \times 32$  pixel, with a  $(K = 10)$ -sparse representation in 5-3 wavelet domain. The measurements were not corrupted by noise but with blurring defined by a Gaussian blurring kernel with variance  $\rho n/l$  and size  $(2n/l + 1)^2$ .

#### IV-B. Reconstruction of compressible signals

Natural signals are almost never exactly sparse in any space. But they can be approximated well by sparse signals, because few coefficients in representation space contain most of their energy. This class of images is called compressible. Indeed many lossy compression algorithms like MP3, JPG, JPG2000, etc. rely on the property of good sparse signal approximation. Therefore to be able to sample and reconstruct natural signals, it is essential for compressive sensing application to also work for compressible signals. In this experiment we test our model (5) and the reconstruction approach of Section III on compressible images.

In this experiment we use the lena-image of size  $l \times l = 64 \times 64$  and take  $M = 20$  measurements with  $l \times l = 16 \times 16$  detectors according to model (5). The blurring operator  $\hat{\mathbf{H}}$  is defined with a Gaussian kernel of size  $[11 \times 11]$  and a variance of two.  $\mathbf{F}_k$  is a diagonal matrix with its values independently drawn from a Bernoulli distribution with elements  $\{-1, 1\}$ . The ground truth image ( $\mathbf{x}$ ) of the experiment and the effect of blurring and downsampling ( $\hat{\mathbf{P}}\hat{\mathbf{H}}\mathbf{x}$ ) is shown in Fig. 3(a) and (b), respectively. Note that even though we have  $16 \times 16$  detectors, Fig. 3(b) is bilinearly upsampled to destination resolution of  $64 \times 64$  pixel.

To show the influence of measurement noise on the reconstructed image, we added different levels of noise to the measurements

and applied the reconstruction method of Section III. The regularizer uses the 5-3 wavelet space for signal representation. For the regularization parameter  $\lambda$ , we choose the value with the best reconstruction quality at a measurement-signal-to-noise-ratio (SNR) of 20dB and use the corresponding regularization parameter  $\lambda_{20}$  as reference for the other noise levels. They are calculated by  $\lambda_S = \lambda_{20} 10^{2-S/10}$ , where  $S$  is the noise level, as SNR, of the corrupted measurements.

In Fig. 4 the quality of the reconstructed image, measured as PSNR against the known ground truth, is shown. Interestingly the reconstruction method needs longer to converge when less noise is added to the measurements, even though the quality of the reconstruction is better. It also shows that the reconstructed image is considerably degraded if noise level rises. As a reference we plotted the quality of a  $(K = 800)$ -sparse approximation of the ground truth image as horizontal gray dashed line in the plot. The red dashed lines are the reconstruction results for a single pixel camera using  $M = 20$  and  $M = 2048$  measurements with an SNR of 60dB. It is obvious that our model with  $M = 20$  measurements provides substantially better results. Fig. 3 shows the reconstructed images using measurements with an SNR of 60dB (Fig. 3(c)) and 20dB (Fig. 3(d)).

#### IV-C. Reconstruction of real life signals

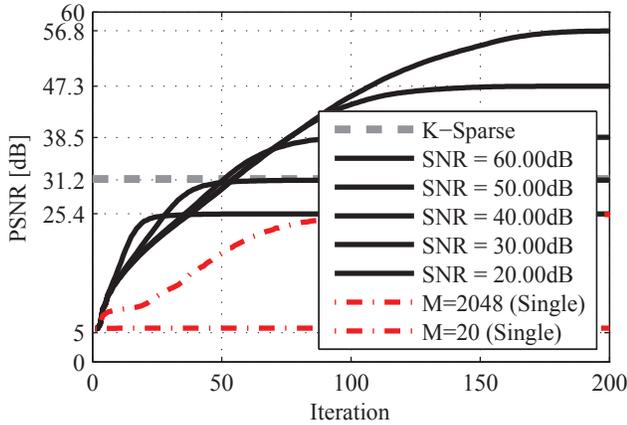
To test our model and reconstruction on real data we built up a setup with a DMD (digital mirror device) as optical multiplication element (modeled by matrix  $\mathbf{F}_k$ ), a detector array (actually a matrix camera) and two lenses (see Fig. 5). The DMD is a micro chip device with many thousands of micro mirrors. They are placed upon SRAM-cells and are individually switchable to one of two directions. The optical multiplication is performed by either switching towards the scene (multiply by 1) or towards a black background (multiply by 0).

The diagonal elements of the multiplication matrix  $\mathbf{F}_k$  are again independently drawn from a Bernoulli distribution with elements  $\{-1, 1\}$ . Since the DMD only allows multiplication with 0 and 1 the sampling process of a single measurement  $\mathbf{y}_k$  from model (4) is splitted into two measurements and we use the difference of both measurements.

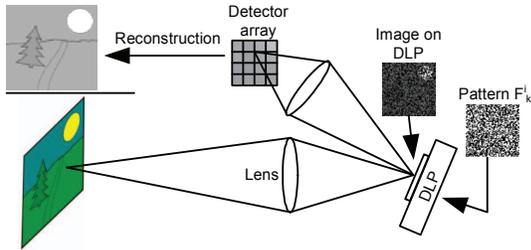
The reconstruction results of the letter "R" can be seen in Fig. 6. The sub figure (a) shows the original image from the camera bilinearly interpolated. Sub figure (b) shows a single measurements  $\mathbf{y}_k$  according to model (4). In sub figure (c) the reconstruction result is shown. The result is upsampled by a factor of 6.75 in each dimension and  $M = 64$  measurements were taken. The reconstruction obeys a high-resolution image of the scene, which would not even be achievable with a higher resolution camera, because of the blurring effect the camera optic introduces. But of course, the reconstruction can not be better than the projection onto the DMD and resolution can not be increased beyond that of the DMD.



**Fig. 3.** Reconstruction using multi-pixel compressive sensing. (a) Original Image with  $64 \times 64$  pixels. (b) Bilinearly upsampled output of  $16 \times 16$  detectors. (c) Reconstruction using  $M = 20$  measurements with an SNR of 60dB. (d) Reconstruction using  $M = 20$  corrupted measurements with an SNR of 20dB.



**Fig. 4.** Convergence of quality (measured as PSNR against ground truth) plotted over the iteration steps of the solver for different measurement noise levels (specified as SNR) for the proposed model. Higher SNR corresponds to higher PSNR result. The dashed horizontal line is the PSNR of a K-Sparse approximation of Fig. 3(a) with  $K = 800$  coefficients in 5-3 wavelet domain. The two red dashed lines are the reconstruction results for the single-pixel model for comparison purpose at  $M = 20$  and  $M = 2048$  measurements with SNR=60dB.



**Fig. 5.** Schematic of data acquisition system used for our experiments. A scene is projected onto a DMD (digital mirror device), which performs an optical multiplication of the projected scene with a pattern  $F_k$  ( $i \in \{1, 2\}$ ). The resulting image on the DMD is projected through a second lens on an array of light detectors. Several measurements of this detector array together with several patterns  $F_k$  are used to reconstruct a high-resolution image.

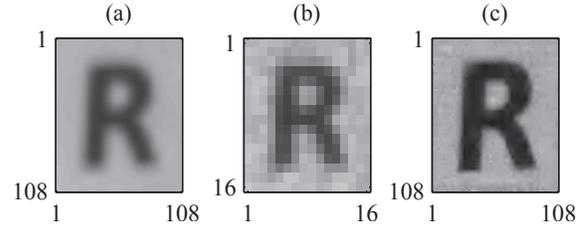
## V. CONCLUSIONS

In this paper we presented a new compressive sensing model using multiple signal detectors. The measurement signals represent different, but possibly similar, observations of a high-resolution scene. We presented a reconstruction method to successfully reconstructed a high-resolution scene from multi-detector measurements. We also showed the trade off between number of detectors and number of necessary measurements ranging from a single detector to full resolution of the output image. Beside the super resolution, our method also performs an implicit deconvolution, with very good reconstruction quality regardless of blurring intensity.

We showed that our method works on simulated sparse and compressible data and we presented a setup for real data acquisition and successfully reconstructed high-resolution images with measurements from a real low-resolution detector array.

Even though the proposed reconstruction method using convex programming converges, the convergence speed is very low. Many algorithms have been developed for the traditional and extended

compressed sensing models to improve convergence speed like Orthogonal Matching Pursuit (OMP), Compressive Sampling Matching Pursuit (CoSaMP), and iterative hard thresholding (ITH) (See [9] and references therein). We have to investigate the application of those algorithms to our new model. Beside that, theoretical guaranties for the proposed model and reconstruction methods have to be given.



**Fig. 6.** Reconstruction results of real data. (a): Best image observable by the detector array bilinear interpolated to destination resolution. (b) Single measurement of the scene optically multiplied by a random pattern  $F_k^1$ . (c) Reconstruction of scene using  $M = 64$  measurements. The up scaling factor is 6.75.

## VI. REFERENCES

- [1] Emmanuel Candès and Justin Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969, 2007.
- [2] E.J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [3] B.K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM journal on computing*, vol. 24, pp. 227, 1995.
- [4] T. Blumensath and M.E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.
- [5] M.F. Duarte, M.A. Davenport, D. Takhar, J.N. Laska, T. Sun, K.F. Kelly, and R.G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.
- [6] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," *CVPR*, 2008.
- [7] A. Griffin and P. Tsakalides, "Compressed sensing of audio signals using multiple sensors," *EUSIPCO08*, 2008.
- [8] D. Baron, M.B. Wakin, M.F. Duarte, S. Sarvotham, and R.G. Baraniuk, "Distributed compressed sensing," *preprint*, 2005.
- [9] J.A. Tropp and S.J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.