

Active Range Imaging via Random Gating

Grigorios Tsagkatakis^a, Arnaud Woiselle^b, George Tzagkarakis^c,
Marc Bousquet^b, Jean-Luc Starck^c, Panagiotis Tsakalides^a

^a ICS - Foundation for Research & Technology - Hellas (FORTH), Crete, Greece

^b Sagem Défense Sécurité, 95100 Argenteuil, France

^c CEA/DSM, Service d'Astrophysique, Centre de Saclay, F-91191 Gif-Sur-Yvette cedex, France

ABSTRACT

Range Imaging (RI) has sparked an enthusiastic interest recently due to the numerous applications that can benefit from the presence 3D data. One of the most successful techniques for RI employs Time-of-Flight (ToF) cameras which emit and subsequently record laser pulses in order to estimate the distance between the camera and an object. A limitation of this class of RI is the requirement for a large number of frames that have to be captured in order to generate high resolution depth maps. In this work, we propose a novel approach for ToF based RI that utilizes the recently proposed framework of Compressed Sensing to dramatically reduce the number of necessary frames. Our technique employs a random gating function along with state-of-the-art minimization techniques in order to estimate the location of a returning laser pulse and infer the distance. To validate the theoretical motivation, software simulations were carried out. Our simulated results have shown that reconstruction of a depth map is possible from as low as 10% of the frames that traditional ToF cameras require with minimum reconstruction error while 20% sampling rates can achieve almost perfect reconstruction in low resolution regimes. Our experimental results have also shown that the proposed method is robust to various types of noise and applicable to realistic signal models.

Keywords: Active Range Imaging, Compressed Sensing, Gated Time-of-Flight Cameras

1. INTRODUCTION

The objective in Range Imaging (RI) is to generate a 2D image called depth map, where pixel intensity corresponds to distance between the camera and the object being imaged. RI has been employed in numerous applications including remote sensing, gaming, security, search and rescue and medical applications.¹ RI systems can be broadly classified as active or passive, depending on the presence or absence of a user-controlled illumination source. Passive RI systems are designed such that 3D information is acquired from a single or multiple images using properties of the scene or the imaging setup, while active RI systems enjoy a boarder range of applications, e.g., night time imaging, due to the active illumination source.

The most representative example of passive RI technique is stereo imaging while novel approaches include depth-from-defocus and coded apertures. Stereo and in general multi-view imaging employs two or more cameras in order to take advantage of the disparity between the focal planes of these cameras and the 3D map is extracted by means of triangulation. The objective in this scenario is to utilize the correspondences between locations from the two (or more) images in order to estimate the fundamental matrix. The fundamental matrix can then be used to triangulate the 3D location of each point in the common field of view from both/multiple cameras.¹⁰ Limitations of this class of methods include the low quality of the depth map, caused by erroneous correspondence matching and the need for multiple cameras with known disparity for increasing the field-of-view, which is limited in the two-camera setups. In addition, the correspondence problem is non-trivial and requires high processing

Further author information: (Send correspondence to G.Ts.) E-mails: G.Ts.: greg@ics.forth.gr, A.W.: arnaud.woiselle@cea.fr, G.Tz.: georgios.tzagkarakis@cea.fr, M.B.: marc.bousquet@sagem.com, J-L.S.: jstarck@cea.fr, P.T.: tsakalid@ics.forth.gr

This work was funded by the IAPP CS-ORION (PIAP-GA-2009-251605) grant within 7th Framework Program of the European Community.

power which may limit the application of such a system in real-time scenarios. On the other hand, the resolution of the depth map usually matches the resolution of the camera sensor and this type of system does not need an active illumination source.

Depth-from-Defocus (DfD) refers to a class of passive depth estimation methods that try to estimate the depth map of a scene using only cues from a monocular image. They are based on the intuition that the amount of blur (defocus) is related to the relative distance of the object to the camera, i.e., points that lie near the focal plane will be in focus while points at greater distances will appear more blurred. These methods are easy to apply since they do not require specific hardware or setups. Nevertheless, DfD methods usually require a large number of images to be captured in order to extract the depth and assume that the scene remains static during the multiple exposures.³ Coded Apertures (CA) is a recently proposed technique that employs sophisticated apertures in order to acquire more information about the scene. By replacing the traditional circular aperture with a coded one and applying appropriate processing on the captured image, one may obtain a higher temporal or spatial resolution of the scene, increased depth-of-field, higher dynamic range etc.⁸ Zhou et al.³ proposed the application of CA for obtaining depth-from-defocus. They proposed an optimization method that generated a pair of coded apertures corresponding to a pair of well defined Point Spread Functions (PSFs). Using the known PSFs, a depth map of the scene is evaluated by performing deconvolution with the filter corresponding to the known PSF. In general, the benefits of coded apertures include the generation of RGB and depth (RBG-D) from a single shot, without significant loss in image resolution and with minimal hardware intervention. However, the depth estimation is very coarse and the process requires software deblurring, which is a non trivial task.²

In contrast to passive RI, active RI systems such as time-of-flight (ToF) and Structured Light (SL) cameras make use of a light source in addition to the camera sensor to generate the depth map. In SL methods, a light projection system, either a coherent source such as a Laser or an incoherent one such as a LCD projector, is responsible for illuminating the scene (typically in the Near IR part of the spectrum) while a camera sensor captures the reflected light.⁵⁻⁷ Depth information can be extracted by investigating the distortion patterns of the projective light field. A key design consideration of an SL based system is the projection pattern which may consist of points, lines, grids, color etc. SL approaches can be broadly divided in sequential and full frame, according to the number of frames that are required to form the depth map. In all cases, the extraction of the depth map follows a similar mentality as that of stereo imaging, albeit with known (or easily estimated) point correspondences. Compared to stereo, SL avoids issues with correspondences, while compared to DfD, SL offers higher accuracy. However, limitations of SL arise from specular reflections, object motion and the need for careful calibration.

Time-of-Flight (ToF) approaches utilize a ToF sensor to measure depth by exploiting the time light takes to travel a distance in order to infer the depth map of a scene. Two classes of ToF methods have been developed, Gated and Continuous-Wave (CW) modulation.⁴ In CW-ToF, a Near IR projector, usually LED based, projects a sinusoidal modulated light while a CMOS or a CCD sensor measures the reflected light. To estimate depth, each sensor element must sample the reflected light four times at equal intervals for every period.⁹ Using these measurements, one can estimate the phase, the offset and the amplitude of the reflected light, and extract a depth estimate. In Gated RI, the shutter of the camera sensor is tightly controlled in order to precisely manipulate the amount of reflected light that is captured at each frame. By limiting the exposure time, only laser pulses that have been reflected from an object at a specific distance are captured. To capture a high range of depth level, multiple frames have to be captured. Figure 1 shows a visual example of the principle of gated ToF RI where the depth map is presented along with the color-to-depth scale.

Specific source of noise that affect ToF cameras were explored by Foix et al.⁴ The authors identified five types of systematic errors occurring in this class of depth estimation, which include depth distortions caused by imperfections in the modulation of the emitted light, integration-time errors where different integration times cause different depth estimation, pixel related errors caused by the sensors such as latencies due to capacitors charge, amplitude related errors due to low or overexposed reflected amplitudes and temperature related errors caused by the operating temperature of the sensor. CW-ToF sensors are also affected by low SNR in low illumination areas, multiple light reception from the sensor, light scattering, especially from objects close to the scene and motion blurring. Similar to SL, ToF sensors require a careful calibration while current ToF sensors are still characterized by low depth map resolution.

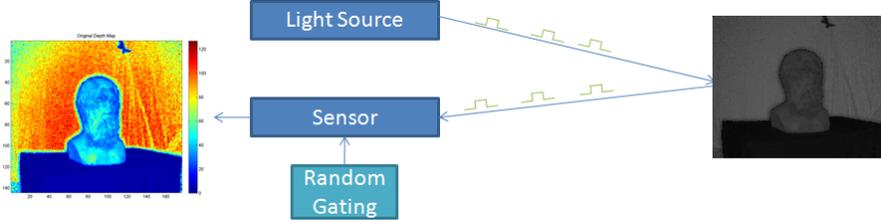


Figure 1. Overview of gated Range Imaging.

In general, SL and ToF are among the highest performing range imaging systems, an observation consistent with the fact that active imaging methods are superior in performance, although specific hardware is required (illumination source). Comparing SL and ToF, SL requires a more careful calibration while depth can only be extracted at illuminated locations. On the other hand, ToF requires high performance sensors, with very fast and accurately controlled electronic components and usually produces a lower resolution depth map, compared to SL.

Despite these benefits, generation of a depth map using classical ToF-RI gating techniques requires the acquisition of a large number of frames, which is proportional to the required depth resolution. In this work, we propose a system based on the premises of compressed sensing (CS),² which is able to achieve excellent depth map reconstruction from a significantly reduced number of frames. The rest of the paper is organized as follows: Section 2, we present an overview of Compressed Sensing while Section 3 discusses the application of Compressed Sensing on gated ToF RI. Experimental results are presented in Section 4 and the paper concludes with Section 5.

2. COMPRESSED SENSING

Compressed sensing (CS) is a novel approach in signal representation and sampling that was introduced by Donoho² and by Candes and Tao.³ The main concept of CS is that a signal can be recovered from a small number of random measurements, that can be far below the Nyquist-Shannon limit. The key assumptions are that the signal is *sparsely* represented in an dictionary, usually an orthogonal basis, and that enough *random* measurements are taken. Formally, let $X \in R^n$ be an input signal that can be represented as a linear combination of basis vectors, as follows: $X = \mathbf{D}S$, where \mathbf{D} is the dictionary matrix and S is the vector of the linear combination coefficients. The signal X is called k -sparse if $\|S\|_0 < k$, where the zero norm counts the non-zero elements. The main premise of CS is that the signal can be fully recovered from a low-dimensional representation $Y = \Psi X = \Psi \mathbf{D}S \in R^M$, where $m \ll b$.

In order to guarantee the stable recovery of the original signal, the $m \times n$ measurement matrix (also called sensing matrix) Ψ should satisfy the so-called Restricted Isometry Property (RIP). Formally, a sensing matrix $\Psi \in \mathbb{R}^{m \times n}$ satisfies the RIP with isometry constant $0 \leq \delta < 1$ if for all k -sparse signals X it holds that

$$(1 - \delta)\|X\|_2^2 \leq \|\Psi X\|_2^2 \leq (1 + \delta)\|X\|_2^2 \quad (1)$$

Designing such a measurement matrix has proven to be a challenging task. However, it has been shown that matrices whose elements are drawn from an appropriate distribution satisfy the RIP with high probability.¹⁴ Examples of such distributions include normalized zero-mean, bounded variance, Gaussian and Bernoulli distributions. Under these conditions, recovery of the sparse signal is possible by solving the following l_0 minimization

$$\min \|S\|_0 \quad \text{subject to} \quad Y = \Psi \mathbf{D}S \quad (2)$$

Even though solving the l_0 problem will produce the correct solution, the l_0 minimization is an NP-hard problem and therefore impractical for moderate sized problems. The main breakthrough of the CS theory is that under the sparsity constraint and the incoherence of the measurement matrix, the solution, i.e. reconstructing the

original signal x and the coefficient vector S from Y , can be found by solving the much easier l_1 optimization problem:

$$\min \|S\|_1 \text{ subject to } Y = \Psi \mathbf{D} S \quad (3)$$

In order to acquire stable solutions, the number of random measurements should be greater than $O(k \log(n/k))$. The above optimization is called Basis Pursuit² and can be solved in polynomial time.

Although strictly sparse signals may be difficult to find in nature, approximately sparse signals encompass a large family of signals that follow a power-law when they are expressed in terms of a basis. In these cases, the largest k coefficients are enough to achieve sufficiently small distortion. Natural images fall under the compressible signals category and image compression algorithms such as JPEG and JPEG2000 rely heavily on this assumption by keeping only a few DCT or wavelet coefficients and discarding the rest. For compressible signals, the goal is not the exact reconstruction of the signal, but the reconstruction of a close approximation of the original signal. In this case, Eq. 3 becomes

$$\min \|S\|_1 \text{ s.t. } \|Y - \Psi \mathbf{D} S\|_2 < \epsilon \quad (4)$$

where ϵ is a bound on the residual error of the approximation which is related to the amount of noise in the data. Besides the convex optimization approach in Eq. 4, greedy methods such as Orthogonal Matching Pursuit (OMP)¹³ have also been proposed. OMP greedily tries to identify the elements of the dictionary that contain most of the signal energy. At each iteration step of the OMP, the element of the dictionary corresponding to the higher absolute inner product with the residual is selected. This element is added to the collection and a new approximation of the signal is evaluated by projecting the input signal to the linear span of the collection. The new residual is found by subtracting the approximated signal from the input signal. The process continues until the residual is within tolerance limits.

3. COMPRESSED SENSING RANGE IMAGING

In order to apply the CS framework on RI, the two main requirements of CS, i.e., sparsity of signal and incoherence of the sampling matrix have to be satisfied. Our proposed CS based RI technique satisfies both requirements. In our model, we assume that the laser pulses that are emitted from the ToF camera at specific time instances, are reflected by an object and subsequently captured by the camera. The laser pulses are emitted such that during the interval between the minimum and the maximum distance, only a single pulse will be captured. From a signal processing perspective, this model implies that the signal in question will be approximated by a Dirac function at the specific time instance the laser pulse returns to the camera, while the rest of the signal between the minimum and the maximum time will be zero.

$$s_i = \begin{cases} 1 & \text{if } i=i_r \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where i_r is the location of the returning pulse. As a result, the captured signal will be extremely sparse and therefore appropriate for CS type sampling. By the fact the speed of light is constant, this time instance can be transformed to the distance between the camera and the object. This process is extended to multiple pulses within each frame in an effort to increase the signal-to-noise ratio of the camera.

The second requirement of CS, that of random linear measurements, can be implemented by selecting a random gating function. The random gating function guarantees the incoherence of the measurement matrix and the stable reconstruction of the signal. Due to the physical characteristics of the system, we employ a Bernoulli type sampling where the gate is either open or close corresponding to a sampling matrix

$$\Phi_{ij} = \begin{cases} 0 & \text{with probability } 1/2 \\ 1 & \text{with probability } 1/2 \end{cases} \quad (6)$$

Mathematically, the problem can be formulated using the CS framework as $\min \|S\|_1 \text{ s.t. } \|Y - \Phi S\| < \epsilon$.

Figure 2 offers a visual explanation of the proposed scheme’s design. In this figure, the left and right sub-figures illustrate the behaviour of periodic and random gating when imaging objects at different depths. The top rows indicate the returning laser pulse, after it has been reflected on an object. The second and fourth rows present the gating function where we observe that periodic sampling leaves the gate open for one time interval in each case while the random gating opens and closes the gate multiple times within the integration time. The third and fifth rows illustrate the captured signals. We observe that in the left sub-figure both the periodic and the random gating functions record a value that can be used for inferring the location of the returning laser pulse. However, in the right sub-figure, we observe that the misalignment of the returning pulse and the gating results in an all-zero signal for the periodic sampling while the random sampling records a valuable signal. As a result, the recorded values by the random gating can be used to infer the location of the pulses and estimate the distance to the object while in traditional gating, more frames have to be acquired.

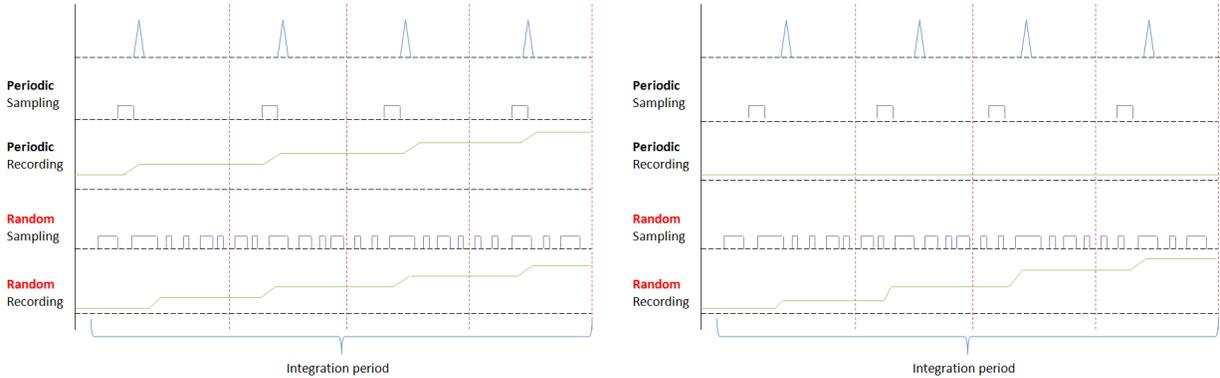


Figure 2. Illustrative example of the differences between traditional and random gating.

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed CS based RI system, we performed a series of experiments on a simulated Gated ToF camera, using data from the Max-Planck institute.¹⁵ In order to provide realistic results, we considered various aspects of the simulated ToF camera. First, we considered the attenuation of laser pulse due to the atmosphere. The attenuation is modelled as $exp(-2 * (support/attenuationFactor))$ where support is the length of the interval between the minimum and the maximum distance in meters and attenuationFactor is a user defined factor that was set to 1000 in our experiments. In addition to attenuation, we also considered signal divergence due to distance. Divergence is modelled as $1/(support^2)$. Attenuation and divergence were considered as multiplicative forms of noise.

In addition to the systematic source of noise, we also considered two forms of random noise. The first class of noise was modelled as white Gaussian and is attributed to the random effects of the atmosphere that affect the laser pulses during propagation. In our experiments the standard deviation of this form of noise was set to 3×10^{-7} . The second type of noise was also modelled as white Gaussian and was introduced during the final sampling of the signal in order to account for sensor related noise such as read-out noise. During our experiments we considered various values for the noise power. To reconstruct the depth maps, we employed the OMP algorithm with 10 iterations for each reconstruction. We note that in our experimental results the dictionary was set as the identity matrix since the signals in question are sparse by nature and thus no explicit dictionary was necessary.

A critical parameter of the simulation is the number of acquired frames. By dividing the acquired frames with the total number of frames that a classical system requires, we obtain the corresponding sampling rate. The total number of necessary frames are dictated by the requested resolution. In our experimental results we considered resolutions of 128 and 256 depth bins. Figure 3 presents the original and reconstructed depth map for sampling rates 10%, 20% and 30% with resolution 128 depth bins. Figure 4 presents a similar set of results for 256 depth level resolution at sampling rates 2%, 5% and 10%. Based on these images, we can make the following

observations. First, we observe that the system is able to achieve excellent quality even at very low sampling rates. Furthermore, we observe that there is a sharp transition between highly distorted and mildly distorted reconstruction, a phenomenon termed phase transition that has been extensively discussed in the CS literature.¹⁶ Last, we observe that the noise that affects the reconstruction has salt-and-paper type characteristics which suggests a specific class of denoising algorithms.

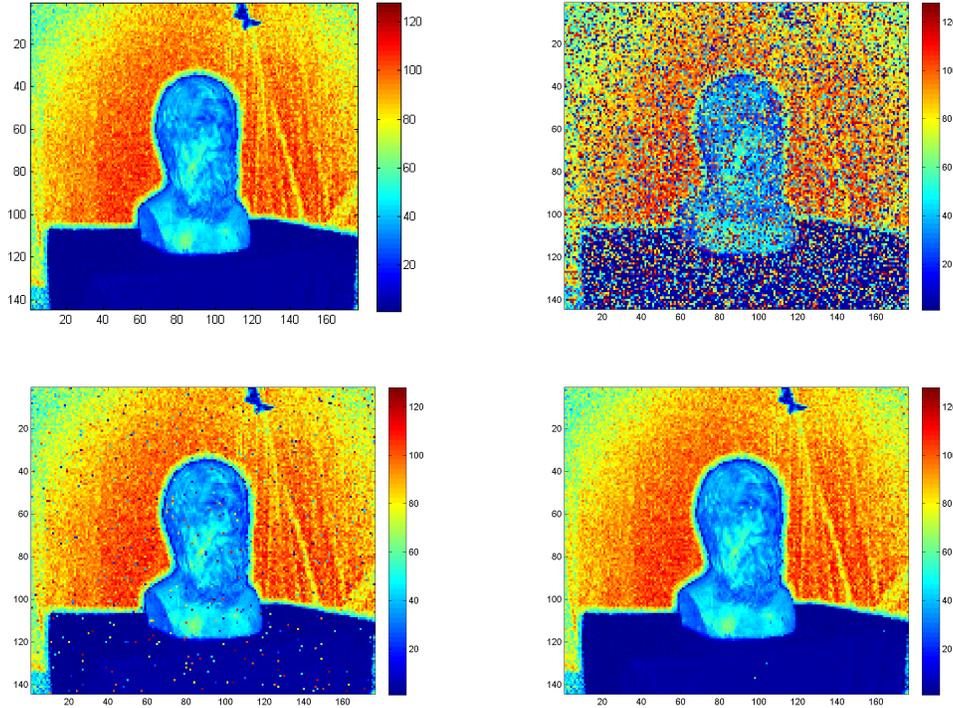


Figure 3. Original (top left) with resolution 128 and reconstructed depth maps from 12 (top right), 24 (bottom left) and 36 (bottom right) frames. The corresponding RMSE are 36.61 for sampling rate 10%, 8.33 for sampling rate 20% and 0.54 for sampling rate 30%.

In addition to the images, Figure 5 presents the reconstruction error (RMSE) as a function of sampling rate for 128 (left) and the 256 (right) depth resolution levels. In these plots we observe that the reconstruction error quickly approaches zero at sampling rates as low as 20% for the low resolution and 9% for the high resolution. We observe that the reconstruction is error free from as low as 8.6 corresponding to sampling rate 2%. The superior performance of CS-RI in higher depth resolutions is attributed to the increased sparsity of the recorded signal since increasing the levels of quantization while maintaining the number of pulses increases the degree of signal sparsity.

To further explore the behaviour of the proposed CS based RI technique, we considered reconstructing noisy versions of the signals (returning laser pulses). Noise in this scenario is modelled as white Gaussian with varying Signal-to-Noise ratios. This type of noise corresponds mainly to read-out noise of the sensor but could also be considered as an approximation to generic sources of disturbance. Table 1 presents the reconstruction error as a function of sampling rate at 256 bins resolution for various SNRs. We observe that the proposed method is very robust to noise and is able to achieve excellent quality even under significant amount of noise.

5. DISCUSSION

In this paper we propose a novel application of Compressed Sensing for the acquisition of range images by gated Time-of-Flight cameras. The proposed system is able to reconstruct depth maps with minimum reduction

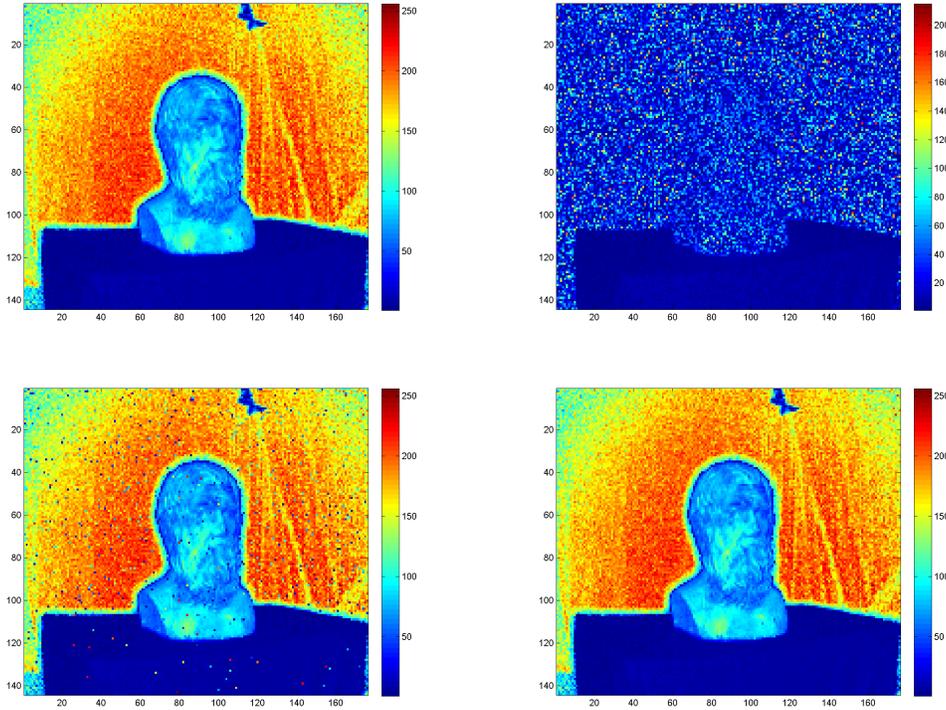


Figure 4. Original (top left) with resolution 256 and reconstructed depth maps from 6 (top right), 13 (bottom left) and 26 (bottom right) frames. The corresponding RMSEs are 118.5 for sampling rate 2%, 8.6 for sampling rate 5% and 0.0 for sampling rate 10%.

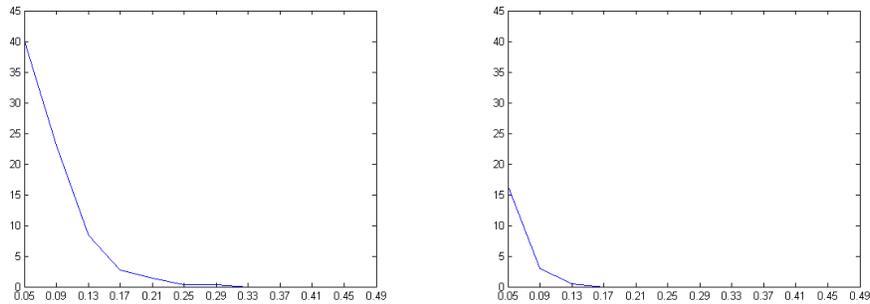


Figure 5. Reconstruction error as a function of sampling rate at depth resolution 128 (left) and 256 (right).

in quality from significantly fewer measurements, when compared to a traditional system, by exploiting the CS framework. Formally, the depth map is constructed by acquiring a small number of frames, where each frame accumulates a large number of returning laser pulses. To achieve the incoherence property required by the CS theory,⁵ the proposed system employs a random gating function where the shutter opens and closes at random intervals during each frame. This random gating function serves as the random projections process in the framework of CS. To recover the depth map, the method assumes that the received signals, that is, the returning laser pulses, are sparse by nature and employ the identity matrix as a dictionary. Experimental results on simulated data have proved the validity of our claims and suggest that CS-RI is a viable technique for gated active range imaging. Future work includes the design of more sophisticated dictionaries and the application of the proposed scheme on real hardware.

Table 1. Reconstruction error for different SNR

<i>SamplingRate</i>	<i>5db</i>	<i>10db</i>	<i>15db</i>	<i>20db</i>
9%	32.2	16.3	9.3	8.3
18%	18.5	3.7	0.4	0.5
28%	11.8	1.8	0.0	0.0
37%	7.8	0.6	0.0	0.0

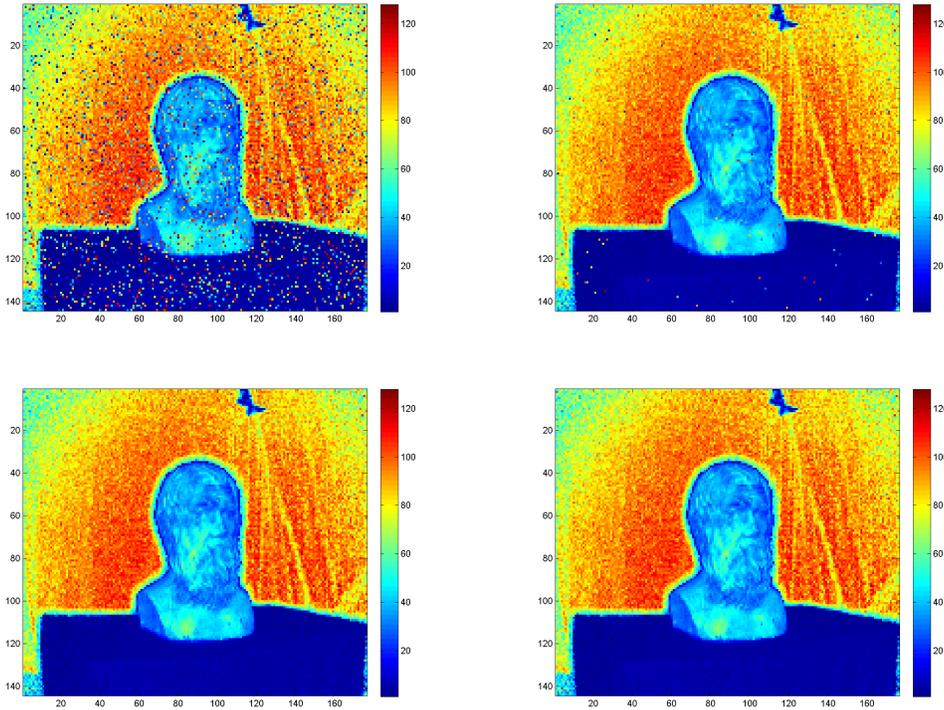


Figure 6. Depth map reconstruction from noisy data. The figure show the reconstructed depth map when the acquired signal is affected by noise. Four scenarios are considered with SNR equal to 5 (top left), 10 (top right), 15 (bottom left) and 20db (bottom right) at sampling rate 18%. The corresponding RMSEs are 18.5, 3.7, 0.4 and 0.0 respectively.

REFERENCES

- [1] Sansoni, G., Trebeschi, M. and Docchio, F., "State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation", *Sensors*, vol. 9(1), pp. 568-601, (2009).
- [2] Levin, A. and Fergus, R. and Durand, F. and Freeman, W.T., "Image and Depth from a Conventional Camera with a Coded Aperture", *ACM Transactions on Graphics (TOG)*, vol. 26(3), (2007).
- [3] Zhou, C., Lin, S. and Nayar, S.K. "Coded Aperture Pairs for Depth from Defocus and Defocus Deblurring", *International journal of computer vision*, vol. 93(1), pp. 53-72, (2011).
- [4] Foix, S., Alenya, G. and Torras, C., "Lock-in Time-of-Flight (ToF) Cameras: a survey", *IEEE journal of sensors*, vol.11(9), pp. 1917-1926, (2011).
- [5] Fofi, D., Sliwa, T. and Voisin, Y., "A comparative survey on invisible structured light", *Proceedings of SPIE*, vol. 5303, pp. 90, (2004).
- [6] Ribo, M. and Brandner, M., "State of the Art on Vision-Based Structured Light Systems for 3D Measurements", *International Workshop on Robotic Sensors: Robotic and Sensor Environments*, pp. 2-6, (2005).
- [7] Geng, J., "Structured-light 3D surface imaging: a tutorial", *Advances in Optics and Photonics*, Vol. 3, Issue 2, pp. 128-160, (2011).

- [8] Raskar, R., "Less Is More: Coded Computational Photography", *Proceedings of the 8th Asian conference on Computer vision-Volume Part I*, pp. 1-12, 2007).
- [9] Hussmann, S. and Ringbeck, T. and Hagebeuker, B., "A performance review of 3D TOF vision systems in comparison to stereo vision systems", *InTech Stereo Vision*, pp. 103-120, (2008).
- [10] Hartley, R. and Zisserman, A., "Multiple view geometry in computer vision", *Cambridge University Press*, UK, (2003).
- [11] Donoho, D.L., "Compressed sensing", *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289-1306, (2006).
- [12] Candes, E.J. and Tao T., "Near-optimal signal recovery from random projections: Universal encoding strategies?", *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406-5425, (2006).
- [13] Tropp, J.A. and Gilbert, A.C., "Signal recovery from random measurements via orthogonal matching pursuit", *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655-4666, (2007).
- [14] Baraniuk, R., Davenport, M., DeVore, R. and Wakin, M., "A simple proof of the restricted isometry property for random matrices", *Constructive Approximation*, vol. 28, no. 3, pp. 253-263, (2008).
- [15] Cui, Y., Schuon, S., Chan, D., Thrun, S. and Theobalt, C., "3D Shape Scanning with a Time-of-Flight Camera", *Proceedings of Computer Vision and Pattern Recognition conference*, (2010), <http://www.mpi-inf.mpg.de/theobalt/tof/>
- [16] Maleki, A. and Donoho, D.L., "Optimally tuned iterative reconstruction algorithms for compressed sensing", *IEEE Journal of Selected Topics in Signal Processing*, vol. 4(2), pp. 330-341, (2010).