Compressive sensing for reconstruction of 3D point clouds in smart systems

Ivo Stančić
Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture
University of Split
Split, Croatia
istancic@fesb.hr

Miloš Brajović, Irena Orović
Faculty of Electrical Engineering
University of Montenegro
Podgorica, Montenegro
milosb@ac.me, irenao@ac.me

Josip Musić
Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture
Split, Croatia
jmusic@fesb.hr

***Abstract*—Performing an accurate 3D surface scan of everyday objects is sometimes difficult to achieve. Using the 3D scanner as a main sensor in a fast-moving mobile robot emphasizes this issue even further. When small robots with limited payload are considered, professional Lidar systems are not likely to be embeded due to their weight, dimensions and/or high cost. Introduction of simple structured-light scanners made possible fast scanning, effective robot detection and evasion of obstacles. Nevertheless, some obstacles may still be difficult to detect and recognize, primarily due to limitations of scanner's performance which results in low number of reconstructed surface points. In this paper a compressed sensing technique, primarily used for reconstruction of 2D images, is utilized to enhance quality of 3D scan, by increasing number of reconstructed 3D points to the scanners theoretical maximum. Obtained results demonstrated feasibility of the approach in terms of mean square error.**

I. Introduction

Basic requirement for any mobile robot is the accurate knowledge of its surroundings, in order to localize itself in space, avoid obstacles in its path, and demostrate smart behaviour in general. This leads to introduction of robot perception systems, whose main function is to detect surrounding objects and avoid possible collisions. Detection and recognition of potential obstacles, especially in dynamic surroundings is a challenging task, and simple detection of obstacles (without actual recognition) in robots vicinity is sometimes insufficient. In a search for a solution, researchers and engineers have introduced several sensor systems, which are based on optoelectronics. They utilize simple geometry principles [1], [2], with emphasis on multi-camera systems and various 3D scanners. The basic principle is similar to the function of human vision [1]. Downside of this approach is introduction of complex calculations, which are often difficult to execute in real-time with robot's on-board computer.

In contrast to camera systems, laser scanners offer simpler calculation of objects location in space, where laser point (line) is moved by precise electromechanical system (rotating mirror in most systems). Since the number of points returned is relatively low, the detection based solely on laser data may be unreliable [3]. Industry-grade laser scanners (SICK) [4] are in some cases more expensive than the robot itself, or too bulky to be efficiently utilized.

Structured light scanners were introduced as a low cost alternative to Lidar scanners [5], [6], since they use simple light pattern projector and off the shelf camera. Development of Microsoft Kinect system, which is in fact more advanced structured light sensor, inspired researchers to use it as a robot navigation device. Correa et. al. [7] have successfully implemented Microsoft Kinect as a sensor for navigation of autonomous surveillance mobile robot indoors.

Both structured light scanner, and laser scanner are easier to implement than multi-camera system, nevertheless they inherit some of the common imperfection of optoelectronic system. Challenging surface patterns like camouflaged objects, dark object with low reflective ratio, part of objects in shade and object lighted by other light source may be difficult to recognize / reconstruct, due to low number of 3D points in resulting point cloud. Each scan can retrieve fixed (maximum) number of reconstructed points (one projection of light structure, or full LIDAR swipe). Techniques used in compressed sensing, which are primarily used for reparations of damaged / undersampled 2D images, may be used for increasing quality of a 3D scan, and supplement missing 3D points in point cloud. Similar approach was already successfully used for refinement of depth-maps obtained by Microsoft Kinect sensor [8].

Sparse signal processing and the compressed sensing (CS) attracted a significant research interest during the last decade [9]-[23]. CS deals with the reconstruction of randomly undersampled signals with the assumption that these are sparse in a known transformation domain. The reduced set of measurements often represents a consequence of a sampling strategy, in order to reduce the data size requirements and the number of acquisitions, preserving the same quality of the information as if these values are available [14]-[19]. On the other hand, known denoising techniques from the robust theory, such as the L-statistics, are used to eliminate signal samples corrupted by high noise [13], [20], [21]. Due to the random nature of the noise, the corrupted values assume also random positions and if the spasity condition is satisfied, CS reconstruction algorithms can be applied in the reconstruction of the eliminated values.

The theoretical foundation of CS lies in fact that the missing samples can be reconstructed by solving an undetermined sys-

tem of linear equations with the additional sparsity constraint [9]-[12], [14]-[19]. Hence, an adequate measure of sparsity is exploited in the reconstruction procedures. A natural way to measure the sparsity is the so-called ℓ_0-norm, that is, the number of non-zero signal coefficients in the observed sparse transform domain.

It is crucial to emphasize that direct variations of unavailable samples values measuring ℓ_0-norm at the same time is an NP (non-deterministic polynomial-time) hard problem. Linear programming techniques and gradient-based algorithms are applied in the reconstruction by relaxing the sparsity constraint involving the ℓ_1-norm. Many studies [9]-[12] have confirmed that in the domain of interest this relaxation procedure is adequate in the CS context. Several reconstruction procedures are based on this relaxation [14]-[19]: well-known convex optimization algorithms such as primal-dual interior point methods, gradient-based method (Orthogonal Matching Pursuit (OMP), Gradient Pursuit and CoSaMP).

In this paper, for the purpose of the observed 2D reconstruction problem, the gradient-based reconstruction algorithm presented in [13] is applied. The adaptations of the algorithm for the case of 2D signals and transforms are presented in [22].

II. Theoretical Background

A. 3D scanner system

Method introduced and evaluated in this paper is presented as a possible improvement of in-house build structured light scanner [5]. Scanner is used primarily for mobile robot navigation, offering accuracy in 3rd dimension of 1.52 mm and RMSE below 1 cm. It is built from off-the shelf components, which includes DLP (Digital Light Processing) projector, digital camera and a computer. Design is based on simple stereovision system [6], where DLP projector acts as an active component and camera is the passive one. Also, use of DLP projector enables dynamic modifications of projected patterns, which could be dynamically altered to cope with current robot's environment. Principle of operation is similar to the laser scanner (and some structured light scanner), where each projected point is analyzed, and scanned surface is reconstructed point by point. If a single light ray is projected from projector (noted as B on Fig.1) it passes through projector frame (noted as point B') and hits the target in point C. The reflected light ray is captured by camera through its plane (noted as point A'). The pixel on camera plane (A') and the pixel on projector plane (B') correspond to an angle between camera and object (α) and angle between projector and object (β). If the exact position of camera and projector in the world coordinate frame is known, the problem of reconstructing exact position of point C is reduced to triangulation problem.

The position of point C in the reference coordinate frame is derived using equations (1) and (2), where A' and B' are its coordinates in the camera and projector planes, respectively; while P_c and P_p are the camera and projector matrices; is a triangulation function; and H is the linear transformation that transforms *A'= HB'* [1].

$$C = \tau\left(A', B', P_c, P_p\right) \tag{1}$$

$$\tau = H^{-1}\left(A', B', P_c H^{-1}, P_p H^{-1}\right) \tag{2}$$

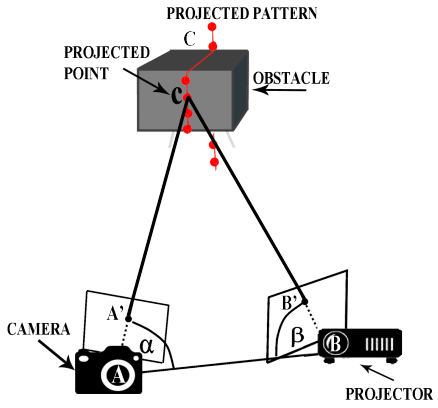

Figure 1. Triangulation of a surface point

If projector and camera are ideally positioned in horizontal direction, projected points forming a horizontal line should also form a horizontal line in camera image. In similar situation, where vertical line is projected, and there is object present in front of scanner system, line is curved on the camera side (Fig.2). This basic principle is also used in proposed scanner system.

In simulation, by placing a plane at infinity distance from scanner, projected pattern (matrix containing 41 x 41 points) captured by camera is a template image for objects at infinity. By placing any object between plane at infinity and scanner system, similar image of projected matrix is captured but with projected points horizontally shifted towards the projector side, where disparity (horizontal shift) is in direct function of distance between scanner and object (Fig.2).

Horizontal displacement of each projected point with index (*i,j*) is stored as a value of matrix on position (*i,j*). Complex 3D reconstruction problem is in this part reduced to 2D problem, which can be represented with the 2D image with rows/ columns that directly correspond to row / columns of the projected matrix, and intensity that corresponds to intensity at given 2D location. In real scenario obtained 2D matrix may have some missing elements.

B. Scanner simulator

Simulator implemented in this paper is mimicking real-life scanner [5]. Virtual scanner is used rather than real scanner as it allows fully controllable scene, which is hard to achieve even in the laboratory conditions. The simulated virtual scanner is created in Blender environment, and consist of camera and light ray projector. Objects of known characteristics and dimensions are placed in front of scanner system in the virtual scene. Result of Blender simulation is video stream recorded from simulated scanner's camera, which has similar

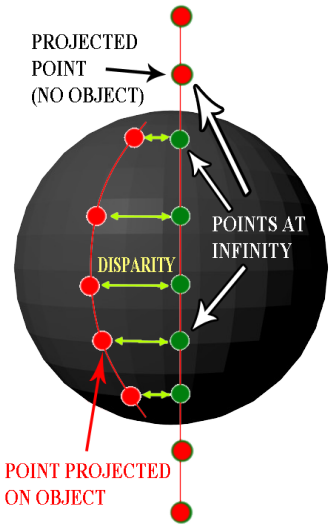

Figure 2. Disparity as seen from camera, as a result of an object placed in front of the scanner system

characteristics as a real camera. Projected pattern consist of 41 x 41 points in matrix format, where each point is projected at a single frame, thus eliminating possible error due to point missclassification.

Simulated scanner is limited to 41 x 41 projected point, due to fact that projected light ray object (cone) in Blender can only accept rounded numbers as angle of cone tip. Smallest possible angle of 1 degree was chosen, in order to achieve maximum density of projected rays, without creating a new non-standard Blender object. Higher density would be possible in real-life projector, which is limited only by projector resolution.

Algorithm that calculates 3D locations of projected points is using previously obtained video stream, which is principally the same algorithm utilized in real-life 3D scanner [5]. Whole algorithm is fully implemented in Matlab environment, all data analysis and results visualizations are also done using Matlab tools.

III. The Compressed Sensing and the Disparity Matrix

A. Modelling the disparity matrix

Let us observe the disparity matrix $f(n,m)$ of size $M \times M$. The 2D DCT of the considered disparity matrix has the following form (DCT II) :

$$C(k_1,k_2) = \sum_{n=0}^{M-1}\sum_{m=0}^{M-1} a_{k_1}a_{k_2} f(n,m) b_{n,m}(k_1,k_2). \quad (3)$$

The corresponding inverse transform is defined by:

$$f(n,m) = \sum_{k_1=0}^{M-1}\sum_{k_2=0}^{M-1} a_{k_1}a_{k_2} C(k_1,k_2) b_{m,n}(k_1,k_2), \quad (4)$$

with

$$b_{n,m}(k_1,k_2) = \cos\left(\frac{2\pi(2n+1)}{2M}k_1\right)\cos\left(\frac{2\pi(2m+1)}{2M}k_2\right), \quad (5)$$

representing 2D DCT basis functions, and $a_{k_1} = a_{k_2} = \sqrt{1/M}$ for $k_1 = 0$ and $k_2 = 0$ respectively and $a_{k_1} = a_{k_2} = \sqrt{2/M}$ for $k_1 \neq 0$, i.e. $k_2 \neq 0$ respectively.

Since it can be observed as a digital image, we assume that the disparity matrix is a K-sparse signal in the 2D DCT domain where $K \ll M^2$, that is

$$f(n,m) = \sum_{i=1}^{K} A_i b_{n,m}(k_{1i},k_{2i}), \quad (6)$$

where A_i denotes the amplitude of the i-th signal component. Disparity matrix has non-zero coefficients at positions (k_{1i},k_{2i}) $i = 1,...,K$ in the 2D DCT domain. This means that only K 2D DCT coefficients of the disparity matrix have significant non-zero values, while other values are equal to zero or negligible.

We further assume that only M_A elements of the disparity matrix are available at positions $(n,m) \in \mathbf{M_A}$ (i.e. $(M-M_A)$ values are missing at random positions). If the signal (i.e. the disparity matrix) satisfies that $K \ll M^2$, according to CS theory, missing samples can be exactly reconstructed if certain conditions are met [9].

B. 2D DCT gradient-based reconstruction algorithm

In the gradient reconstruction algorithm all values at missing samples positions are set to zero. In further iterations these values are considered as minimization variables, and they are varied with a small, appropriately chosen step $\pm\Delta$. For every observed missing value position concentrations of the both 2D DCTs are evaluated, in order to determine the gradient direction, which is defined as the difference of the concentration measures. The missing disparity matrix values are then updated simultaneously in a steepest descent manner. A good starting value of the step can be obtained as:

$$\Delta = \max|f(n,m)|, \ (n,m) \in \mathbf{M_A}. \quad (7)$$

Before the algorithm starts, the signal consisted of available signal samples and with zeros at missing samples positions is formed, according to :

$$y(m,n) = \begin{cases} 0, & \text{for } (n,m) \in \mathbf{M}\backslash\mathbf{M_A} \\ f(n,m), & \text{for } (n,m) \in \mathbf{M_A}, \end{cases}$$

where $\mathbf{M}$ denotes the full set of signal positions.

For each iteration k, until the desired precision is obtained, the following steps are repeated:

Step 1: For each missing value at the position $(n,m) \in \mathbf{M}\backslash\mathbf{M_A}$, form two auxiliary matrices according to:

$$y_1^{(k)}(n,m) = \begin{cases} y_1^{(k)}(n,m) + \Delta, & \text{for } (n,m) \in \mathbf{M}\backslash\mathbf{M_A} \\ y_1^{(k)}(n,m), & \text{for } (n,m) \in \mathbf{M_A} \end{cases},$$

$$y_2^{(k)}(n,m) = \begin{cases} y_2^{(k)}(n,m) - \Delta, & \text{for } (n,m) \in \mathbf{M} \backslash \mathbf{M_A} \\ y_2^{(k)}(n,m), & \text{for } (n,m) \in \mathbf{M_A} \end{cases},$$

Step 2: Calculate the finite differences of the signal transform concentration measures [24]

$$g(n,m) = \frac{1}{2\Delta}\left[\mathcal{M}^+ - \mathcal{M}^-\right] \tag{8}$$

where

$$\mathcal{M}^+ = \frac{1}{M^2}\sum\nolimits_{k_1}\sum\nolimits_{k_2}\left|C^+(k_1,k_2)\right|$$
$$\mathcal{M}^- = \frac{1}{M^2}\sum\nolimits_{k_2}\sum\nolimits_{k_2}\left|C^-(k_1,k_2)\right|$$

represent concentration measures. Note that $C^+(k_1,k_2)$ and $C^-(k_1,k_2)$ denote calculated 2D DCTs of two previously defined signals $y_1^{(k)}(n,m)$ and $y_2^{(k)}(n,m)$ respectively.

Step 3: Form the gradient matrix $\mathbf{G}^{(k)}$ of the same size as the disparity matrix $f(n,m)$ with elements defined as follows:

$$G^{(k)}(n,m) = \begin{cases} g(n,m), & \text{for } (n,m) \in \mathbf{N} \backslash \mathbf{M_A} \\ 0, & \text{for } (n,m) \in \mathbf{M_A} \end{cases}$$

with $g(n,m)$ calculated in the Step 2.

Step 4: Correct the values of $y(n,m)$ using the gradient matrix $\mathbf{G}^{(k)}$ with the steepest descent approach:

$$y^{(k+1)}(n,m) = y^{(k)}(n,m) - 2\Delta G^{(k)}(n,m).$$

We multiply the gradient matrix with the factor 2Δ to eliminate the dependence on Δ that appears in (8). By decreasing Δ when the algorithm convergence slows down a high level of precision can be achieved. The proper decrease of the step can be achieved when the oscillatory nature of the adjustments is detected [13].

IV. Simulation Results and Discussion

In order to illustrate the presented theoretical concepts, several disparity matrices are observed.

Example 1: We consider the compressed Sensing scenario, where $M - M_A = 400$ randomly positioned disparity matrix values are missing. Moreover, it is important to emphasize that certain disparity matrix elements have infinite values, in corresponding to the background at infinity, and thus we considered them also as unavailable, besides the observed 400 unavailable samples. The number of these elements is 253, 225, 421, 204, 148, 232 respectively, These values represent an additional challenge for the reconstruction algorithm, especially since they are grouped. The corresponding disparity matrices with missing elements are shown in Fig.3, where missing values are denoted with dark blue color. Light blue denotes infinity while yellow to red shades denote from furthers to closest object or its parts. We apply the gradient reconstruction algorithm and the obtained results are shown in Fig. 4 (a)-(f). Please note that scanned artificial objects were: a) Blender monkey standard object, b) sphere, c) multiple objects (spheres and cubes), d) small sphere with a background, e) single human, and f) humans.

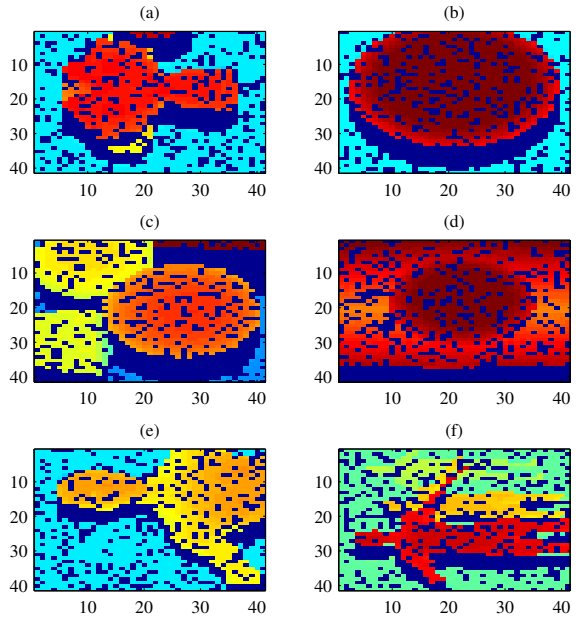

Figure 3. Disparity matrices with $M - M_A = 400$ missing values at random positions

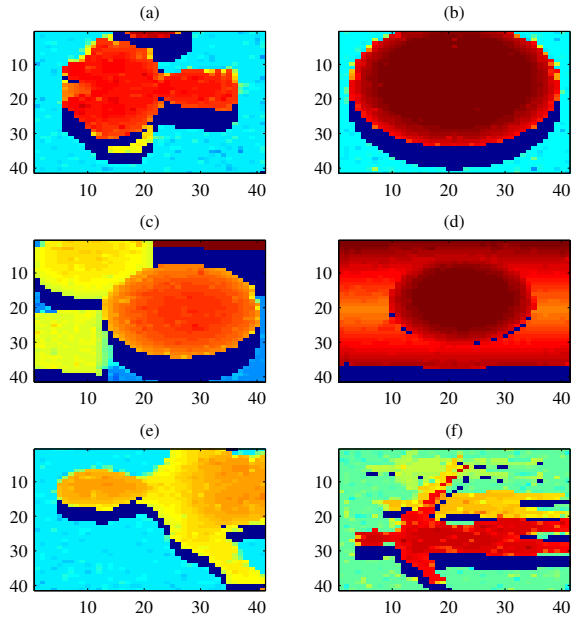

Figure 4. Reconstructed disparity matrices

Example 2: In order to additionally validate the results, we calculate MSE between the original and undersampled disparity matrices, as well as the MSE between the original and reconstructed disparity matrices. We consider scenarios from 10% to 80% of missing disparity matrix values, with the step of 10%. MSEs are obtained based on averaging the squared reconstruction errors calculated for 100 independent random realizations of missing values positions in disparity matrices, for each considered percent of missing values. Results are

shown in Fig.5 proving the significant MSE improvement after the reconstruction is done.

With 80% of disparity missing, sphere object has the MSE of 39.9 dB, while humans object has MSE of 44.9 dB. CS algorithm demonstrated significant improvement, lowering MSE for both object to 15.7 dB and 15.8 dB respectively, which is significantly better than undersampling of original matrices by 10%. From Fig.5 it can be seen that difference between undersampled and reconstructed MSE value approximately is the same across whole missing values range. Also, as expected MSE value increases as the number of missing values increases.

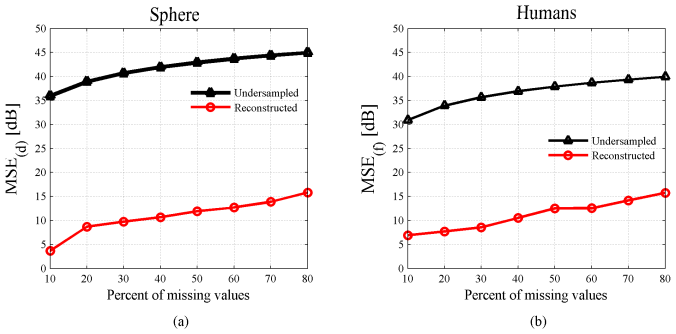

Figure 5. MSE analysis

V. Conclusion

The objective of the paper was to improve quality of a 3D scanner data, by introducing the compressed sensing based signal reconstruction. this in turn enables more reliable behavior of different smart systems, which can potentially use 3D scanners output. In real-life scenarios, projected pattern are not completely detected by camera, and thus maximum number of surface points cannot be achieved. In order to provide better reconstruction of scanned surface, complex 3D reconstruction problem is partly reduced to 2D problem (disparity matrix), and with the use of compressed sensing algorithms, missing data are filled. Hence, almost maximum theoretical number of points can be obtained. In simulated environment, proposed method performs very well. Presented results show that even when 80% percent of disparity matrix is lost, compressed sensing reconstruction algorithm provided high quality of reconstructed surface and successfully filled missing parts of scanned objects. Despite time-consuming and complexity of CS algorithm, small resolution of disparity matrix allow future implementation of the algorithm in mobile robots scanner system. In future research, presented CS algorithms would be included in more complex structured-light scanners with dynamically adaptable density of projected points, as well as implemented in parallel processing fashion for increased computational efficiency.

Acknowledgment

This research is supported by the Croatian-Montenegrian bilateral project "Compressed sensing and super-resolution in surveillance systems based on optical sensors and UAVs".

References

[1] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision, Second edition*, Cambridge University Press, 2004.
[2] G. Shapiro, G.C. Stockman, *Computer vision*, Prentice-Hall, 2001.
[3] E. P. Fotiadis, M. Garzn, A. Barrientos "Human Detection from a Mobile Robot Using Fusion of Laser and Vision Information."*Sensors* , Vol 13(9), pp 11603-11635, 2013.
[4] Y. Cang , J. Borenstein, (2002, May). "Characterization of a 2-D Laser Scanner for Mobile Robot Obstacle Negotiation "*International Conference on Robotics and Automation*, IEEE. 2002.
[5] I. Stancic, J. Music, M. Cecic, "A Novel Low-Cost Adaptive Scanner Concept for Mobile Robots "*Ingeniera e Investigacin*, Vol 34, pp 37-43, 2014.
[6] C. Rocchini, P. Cignoni, C. Montani, P. Pingi, R. Scopigno "A low cost 3D scanner based on structured light "*Computer Graphics Forum*, Vol 20(3), pp 299-308, 2001.
[7] D. S. O. Correra, D. F. Scotti, M. G. Prado, D. F. Sales, "Mobile Robots Navigation in Indoor Environments Using Kinect Sensor. "*Brazilian Conference on Critical Embedded Systems*, IEEE, 2012.
[8] Q. Zhang, S. Li, W. Guo, P. Wang, J. Huang "Refinement of Kinect Sensors Depth Maps Based on GMM and CS Theory," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, Vol 8(5), pp 87-92, 2015.
[9] E. Candès, J. Romberg, T. Tao: "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,"*IEEE Transactions on Information Theory*, Vol 52(2), pp 489-509, 2006.
[10] E. J. Candès, M. B. Wakin, "An Introduction To Compressive Sampling", *IEEE Signal Processing Magazine*, Vol 25(2), pp 21-30, 2008.
[11] D. Donoho: "Compressed sensing," *IEEE Trans. on Information Theory*, 2006, Vol 52(4), pp 1289-1306, 2006.
[12] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, Vol 24(4), pp 118-121, 2007.
[13] LJ. Stanković, M. Daković, and S. Vujović, "Adaptive Variable Step Algorithm for Missing Samples Recovery in Sparse Signals," *IET Signal Processing*, Vol 8(3), pp 246-256, 2014.
[14] M. Elad, *Sparse and Redudant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.
[15] H. Rauhut, "Stability Results for Random Sampling of Sparse Trigonometric Polynomials," *IEEE Transactions on Information theory*, vol 54(12), pp 5661-5670, 2008.
[16] C. Studer, P. Kuppinger, G. Pope, H. Bolcskei, "Recovery of sparsely corrupted signals," *IEEE Transactions on Information Theory*, Vol 58(5), pp 3115-3130, 2012.
[17] M. Davenport, M. Duarte, Y. Eldar, G. Kutyniok, "Introduction to compressed sensing "*Chapter in Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.
[18] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, Vol 20(3), pp 301321, 2009.
[19] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, " *IEEE Journal of Selected Topics in Signal Processing*, Vol 1(4), pp 586597, 2007.
[20] M. Daković, LJ. Stanković, and I. Orović, "Adaptive Gradient Based Algorithm for Complex Sparse Signal Reconstruction, "*22nd Telecommunications Forum TELFOR*, Belgrade, 2014.
[21] S. Stanković, I. Orović, and E. Sejdić, *Multimedia signals and systems*, Springer-Verlag, 2012.
[22] I. Stanković, I. Orović, and S. Stanković, "Image Reconstruction from a Reduced Set of Pixels using a Simplified Gradient Algorithm,"22nd Telecommunications Forum TELFOR, 2014.
[23] E. Sejdić, A. Cam, L.F. Chaparro, C.M. Steele and T. Chau, "Compressive sampling of swallowing accelerometry signals using TF dictionaries based on modulated discrete prolate spheroidal sequences," *EURASIP Journal on Advances in Signal Processing*, 2012:101 doi:10.1186/1687–6180–2012–101.
[24] LJ. Stanković, "A measure of some time–frequency distributions concentration," *Signal Processing*, Vol 81, pp 621-631, 2001.