

## Research Article

# Gradient Compressive Sensing for Image Data Reduction in UAV Based Search and Rescue in the Wild

Josip Musić,<sup>1</sup> Irena Orović,<sup>2</sup> Tea Marasović,<sup>1</sup> Vladan Papić,<sup>1</sup> and Srdjan Stanković<sup>2</sup>

<sup>1</sup>Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Ruđera Boškovića 32, 21000 Split, Croatia

<sup>2</sup>Faculty of Electrical Engineering, University of Montenegro, Džordža Vašingtona bb, 81000 Podgorica, Montenegro

Correspondence should be addressed to Josip Musić; [jmusic@fesb.hr](mailto:jmusic@fesb.hr)

Received 1 April 2016; Revised 30 September 2016; Accepted 11 October 2016

Academic Editor: Agathoklis Giaralis

Copyright © 2016 Josip Musić et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Search and rescue operations usually require significant resources, personnel, equipment, and time. In order to optimize the resources and expenses and to increase the efficiency of operations, the use of unmanned aerial vehicles (UAVs) and aerial photography is considered for fast reconnaissance of large and unreachable terrains. The images are then transmitted to control center for automatic processing and pattern recognition. Furthermore, due to the limited transmission capacities and significant battery consumption for recording high resolution images, in this paper we consider the use of smart acquisition strategy with decreased amount of image pixels following the compressive sensing paradigm. The images are completely reconstructed in the control center prior to the application of image processing for suspicious objects detection. The efficiency of this combined approach depends on the amount of acquired data and also on the complexity of the scenery observed. The proposed approach is tested on various high resolution aerial images, while the achieved results are analyzed using different quality metrics and validation tests. Additionally, a user study is performed on the original images to provide the baseline object detection performance.

## 1. Introduction

In modern society, people often engage in different outdoor activities for fun or recreation. However, they sometimes overestimate their abilities or even get lost and need help and assistance. In such situations even if the lost person has a mobile phone, it is difficult to provide exact location or the signal strength is too low to be useful. In order to provide medical or other kinds of assistance, the emergency services need to locate a person. The number of such situations increases especially during summer months when people are more active both on the sea and in the mountains [1]. On average, there were 1,862 calls to emergency services in UK in 2014. Review of search and rescue operations in the US over time period of 15 years was examined in [2]. In the considered time period there were about 65,000 search and rescue incidents involving around 78,000 persons in need of assistance. Similar trends can be found in somewhat less developed countries with large number of tourists [3]. The exponential growth in total number of mountain rescue

operations is observed for the period between 1991 and 2005. For the period of years 2002–2005, there were 1,217 operations in total: 12.16% were search operations and 27.12% were rescue operations. Out of all search operations, 81.75% of persons were found (out of which 10.71% involved fatalities). Among all search actions, in 17.52% of cases, search dogs were employed and in 10.22% of cases the military helicopters were used. This large number of operations and limited time frame requires significant resources in terms of money, equipment, and manpower. Hence, the search and rescue operations are very time and resource demanding. Usually, a number of ground personnel (including police officers, firefighters, ambulances, and even locals) are involved as well as military or civilian helicopters with their crews. These resources are in general expensive and the majority of them are used for the search part of the operation. Search operation also needs to be completed as fast as possible to ensure positive outcome. Thus, the efforts are mainly focused towards optimizing search procedure in order to provide reliable information to the search team where to look first/next. In this way, it is

possible to significantly increase the probability of positive outcome and to ensure lower costs. One of the solutions is to include aerial photography (via UAVs) and associated pattern recognition and image processing algorithms [4]. Namely, UAVs enable fast reconnaissance of large and unreachable areas, while taking photos in regular time intervals. UAVs in such scenarios are intended to complement the search procedure and not to completely replace search on the ground [5].

The use of UAVs for supporting search and rescue operations was also suggested in [6], where the authors used different victim detection algorithms and search strategies in simulated environment. It is also suggested that four features should be taken into account when designing the appropriate search system: (1) quality of image data, (2) level of information exchange between UAVs and/or UAVs and ground station, (3) UAV autonomy time, and (4) environmental hazards.

In this paper we are concerned with image quality and level of information exchange, requirements which are in general in opposition: better quality sensors consume larger network bandwidth and vice versa. Transmission resources (bandwidth, signal availability, and quality) are not readily available in the wild, where search and rescue operations usually take place. Thus, finding a good compression algorithm to reduce amount of data [7, 8] or using smart cameras with reduced number of pixels could be of great importance. For instance, using reduced number of pixels allows consuming less energy from UAVs battery and thus achieving greater autonomy/range. In [7], three levels of filters with different knowledge (redundancy, task, and priority) were used. Depending on the filter (or combination of filters) applied, the reduction of transmission bandwidth is achieved in the range from 24.07% to 67.83%. The authors however noted that in general MPEG reduction level was not achieved. Additional common difficulty is the presence of noise in the areas where signal coverage is limited [4]. This noise can manifest in different manners such as salt-and-pepper type of noise or even as whole blocks of missing pixels. Thus, finding effective ways to deal with such situations is of importance, since the noise can reduce (or completely cover) victim's footprint in the image. Therefore, it is of great importance to explore possibilities of reducing the amount of captured and transmitted data, while maintaining the level of successful target objects detection using specially designed algorithms, as it was initially introduced in [9]. Such algorithms should also exhibit certain level of robustness in the presence of noise. In that sense, we propose using a popular compressive sensing (CS) approach in order to deal with randomly undersampled terrain photos, obtained as a result of smart sensing strategy or removal of impulse noise. The CS reconstruction algorithms may deal with images having large amount of missing or corrupted pixels [10–16]. The missing pixels can be subject of the reduced sampling strategy when a certain amount of pixels at random positions are omitted from observations or may appear as a consequence of discarding pixels affected by certain errors or noise as discussed in the sequel [14]. In a general CS scenario, it is possible to observe two noisy effects called the observation noise and the sampling noise [17, 18]. The observation noise appears in the phenomena before sensing process, while the sampling noise appears as an error

on the measurements after the sensing process. As proved in [17], both types of noise would cause distortion in the final image, whereas the observation noise in CS case is less detrimental than the sensing noise. In the considered imaging system for aerial photography, the noise can be occasionally present in the form of salt-and-pepper noise (or spike noise). It is mainly caused by the analog-to-digital converter errors that could be considered as observation noise appearing prior to CS process [18]. It can be also caused by the errors in transmission, which indeed correspond to the observation noise. In both cases, our assumption is that we can discard the pixels affected by the salt-and-pepper noise and proceed with the image reconstruction upon receipt in the control center using small amount of available nonnoisy pixels. Hence, either the images are randomly undersampled in the absence of noise, or the random noisy pixels are discarded causing the same setup as in the former case. Moreover, if some of the measurements are subject to the transmission errors, these can be also discarded upon the receipt in the control center. Additionally, in some applications facing other heavy-tailed noise models, robust sampling operators based on the weighted myriad estimators can be considered [18].

Therefore, we consider the following scenario: (a) randomly undersampled images (with missing pixels caused by the sampling strategy or by discarding noisy pixels) are obtained and sent over the network; (b) upon reception in the control center, the images are completely restored/reconstructed and (c) are used as the input of image processing algorithm for target objects detection. Only the images of interest are flagged by the object detection algorithm (as the images with suspicious objects) and further examined by a human operator. Hence, the proposed system consists of two segments, both running in the control center: image reconstruction and object detection. It is important to emphasize that the output of the first segment influences the performance of the second segment depending on the number of missing pixels (i.e., image degradation level). In the paper, we examine the performance of both segments under different amounts of missing pixel with the aim to draw conclusions on their respective performances and possible improvements.

## 2. Materials and Methods

An overview of the entire proposed approach is presented in Figure 1. In the next sections, the particular algorithms from the figure will be presented and explained in more detail. However, it should be noted that the entire approach is not yet fully integrated into UAV and adopted for real-time application. This is a topic of an ongoing research and it will be implemented in the future.

*2.1. Compressive Sensing and Image Reconstruction.* Compressive sensing appeared as a new sensing paradigm, which allows acquiring significantly smaller number of samples compared to the standard approach based on the Shannon-Nyquist sampling theorem [10–13]. Particularly, compressive sensing assumes that the exact signal can be reconstructed from a very small set of measurements, under the condition

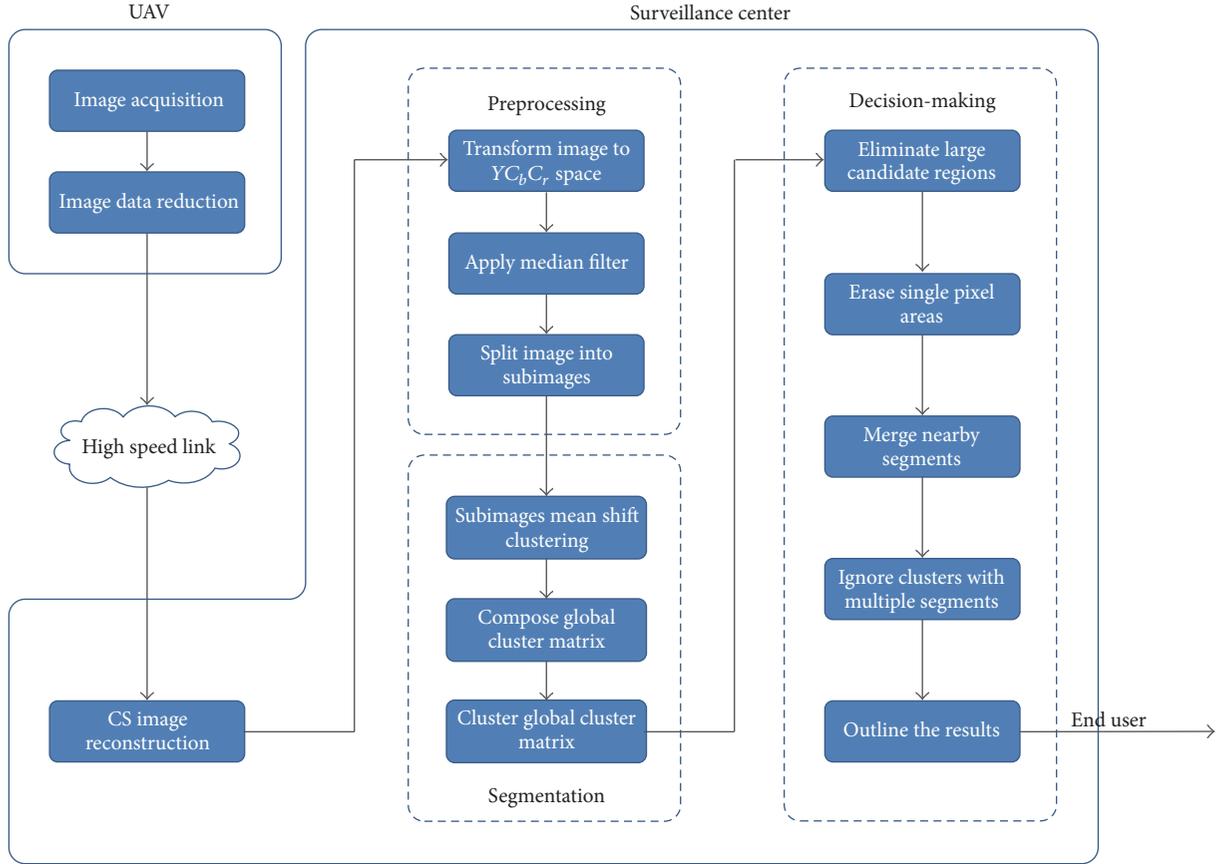


FIGURE 1: General overview of the proposed procedure.

that the measurements are incoherent and the signal has sparse representation in certain transform domain.

Consider a signal in  $R^N$  that can be represented using the orthonormal basis vectors  $\{\Psi_i\}_{i=1}^N$ . The desired signal  $\mathbf{x}$  can then be observed in terms of the transform domain basis functions [10, 11]

$$\mathbf{x} = \sum_{i=1}^N X_i \Psi_i \quad (1)$$

or equivalently in the matrix form

$$\mathbf{x} = \Psi \mathbf{X}, \quad (2)$$

where the transform domain matrix is  $\Psi := [\psi_1 | \psi_2 | \dots | \psi_N]$ , while  $\mathbf{X}$  represents the transform domain vector. In the case of sparse signals, the vector  $\mathbf{X}$  contains only  $K \ll N$  significant components, while others are zeros or negligible. Then we might say that  $\mathbf{X}$  is a  $K$ -sparse representation of  $\mathbf{x}$ . Instead of measuring the entire signal  $\mathbf{x}$ , we can measure only a small set of  $M < N$  random samples of  $\mathbf{x}$  using a linear measurement matrix  $\Phi := [\phi_1 | \phi_2 | \dots | \phi_N]$ . The vector of measurements  $\mathbf{y}$  can now be written as follows:

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \mathbf{X} = \Theta \mathbf{X}, \quad (3)$$

where  $\Theta = \Phi \Psi$  of size  $M \times N$ .

The main challenge behind compressive sensing is to design a measurement matrix  $\Phi$ , which can provide exact and unique reconstruction of  $K$ -sparse signal where  $M \geq K$ . Here, we consider the random measurement matrix  $\Phi$  of size  $M \times N$  ( $M \ll N$ ) that contains only one value equal to 1 at the random position in each row (while other values are 0). Particularly, in the  $i$ th row, the value 1 is on the position of the  $i$ th random measurement. Consequently, the resulting compressive sensing matrix  $\Theta$  in (3) is usually called random partial transform matrix [19, 20]. It contains partial basis functions from  $\Psi$  with values at the random positions of measurements. Note that, in the case of images, the two-dimensional measurements are simply rearranged into one-dimensional vector  $\mathbf{y}$ , while  $\Psi$  should correspond to a two-dimensional transformation.

Since  $M < N$ , the solution is ill-posed: the number of equations is smaller than the number of unknowns. If we could determine the support of  $K$  components within  $\mathbf{X}$ , then the problem would be determined by the  $M \geq K$  system of linear equations ( $M$  equations with  $K$  unknowns). The stable recovery is assured if the compressive sensing matrix  $\Theta$  satisfies the restricted isometry property (RIP). However, in real situations we almost never know the signal support in  $\mathbf{X}$ , and the signal reconstruction needs to be performed using

certain minimization approaches. The most natural choice assumes the minimization of  $\ell_0$  norm, which is used to search for the sparsest vector  $\mathbf{X}$  in the transform domain:

$$\begin{aligned} \mathbf{X} &= \operatorname{argmin} \|\mathbf{X}\|_0 \\ \text{subject to } & \Theta \mathbf{X} = \mathbf{y}. \end{aligned} \quad (4)$$

However, this approach is computationally unstable and represents NP (nondeterministic polynomial-time) hard problem. Therefore, the  $\ell_0$  norm minimization is replaced by the  $\ell_1$  norm minimization, leading to the convex optimization problem that can be solved using linear programming:

$$\begin{aligned} \mathbf{X} &= \operatorname{argmin} \|\mathbf{X}\|_1 \\ \text{subject to } & \Theta \mathbf{X} = \mathbf{y}. \end{aligned} \quad (5)$$

In the sequel, we consider a gradient algorithm for efficient reconstruction of images, which belongs to the group of convex optimization methods [15]. Note that the gradient based methods generally do not require the signals to be strictly sparse in certain transform domain and in that sense provide significant benefits and relaxations for real-world applications. Particularly, it is known that the images are not sparse in any of the transform domains but can be considered as compressible in the two-dimensional Discrete Cosine Transform (2D DCT) domain. Hence, the 2D DCT is employed to estimate the direction and value of the gradient used to update the values of missing data towards the exact solution.

*2.1.1. Gradient Based Signal Reconstruction Approach.* Previous minimization problem can be solved using the gradient based approach. The efficient implementation of this approach can be done in a block by block basis, where the block sizes are  $N \times M$  (the square blocks  $N = M = 16$  are used in the experiments). The available image measurements within the block are defined by the pixels indices

$$(i, j) \in \Omega \quad \text{where } \Omega \subset \{(1, 1), (1, 2), \dots, (N, M)\}, \quad (6)$$

while

$$\operatorname{card} \{\Omega\} = M_a N_a \ll NM. \quad (7)$$

The original (full) image block is denoted as  $f(n, m)$ . The image measurements are hence defined by

$$f(i, j), \quad \text{for } (n, m) = (i, j) \in \Omega. \quad (8)$$

Let us now observe the initial image matrix  $\mathbf{z}$  that contains the available measurements and zero values at the positions of unavailable pixels. The elements of  $\mathbf{z}$  can be defined as

$$z(n, m) = f(n, m) \delta(n - i) \delta(m - j), \quad (9)$$

where  $\delta(n)$  is a unit delta function. The gradient method uses the basic concept of steepest descent method. It treats only the missing pixels such that their initial zero values are varied in both directions for a certain constant  $\Delta$ . Then, the  $\ell_1$  norm is applied in the transform domain to measure the level of sparsity and to calculate the gradient, which is also used to update the values of pixels through the iterations. In that sense, we might observe the matrix  $\mathbf{Z}$  comprising  $N$  vectors  $\mathbf{z}$  formed by the elements  $z(n, m)$ :

$$\mathbf{Z} = [\mathbf{z} \ \mathbf{z} \ \cdots \ \mathbf{z}]. \quad (10)$$

In order to model the process of missing pixels variations for  $\pm\Delta$ , we can define the two matrices:

$$\begin{aligned} \mathbf{Z}_k^+ &= [\mathbf{z}_{1k}^+ \ \mathbf{z}_{2k}^+ \ \cdots \ \mathbf{z}_{Nk}^+] = \mathbf{Z}_k + \Delta, \\ \mathbf{Z}_k^- &= [\mathbf{z}_{1k}^- \ \mathbf{z}_{2k}^- \ \cdots \ \mathbf{z}_{Nk}^-] = \mathbf{Z}_k - \Delta, \end{aligned} \quad (11)$$

where  $k$  denotes the number of iterations. The initial value of  $\Delta$  can be set as  $\Delta = \max(|\mathbf{z}|)$ . The previous matrices can be written in an expanded form

$$\begin{aligned} \mathbf{Z}_k^+ &= \mathbf{Z}_k \pm \Delta \\ &= [\mathbf{z}_k \ \mathbf{z}_k \ \cdots \ \mathbf{z}_k] \pm \Delta \operatorname{diag} [\delta(n - i) \delta(m - j)], \end{aligned} \quad (12)$$

where

$$\mathbf{Z}_k = \begin{bmatrix} \mathbf{z}_k(1, 1) & & \mathbf{z}_k(1, 1) \\ \mathbf{z}_k(1, 2) & & \mathbf{z}_k(1, 2) \\ \vdots & \cdots & \vdots \\ \mathbf{z}_k(N, M) & & \mathbf{z}_k(N, M) \end{bmatrix}, \quad (13)$$

while

$$\operatorname{diag} [\delta(n - i) \delta(m - k)] = \begin{bmatrix} \delta(1 - i_1) \delta(1 - j_1) & 0 & \cdots & 0 \\ 0 & \delta(1 - i_1) \delta(2 - j_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta(N - n_N) \delta(M - m_M) \end{bmatrix}. \quad (14)$$

Based on the matrices  $\mathbf{Z}_k^+$  and  $\mathbf{Z}_k^-$ , the gradient vector  $\mathbf{G}$  is calculated as

$$\mathbf{G}^k = \frac{1}{2\Delta} \left[ \|\text{DCT}_{2D}\{\mathbf{Z}_k^+\|_1^{\text{col}} - \|\text{DCT}_{2D}\{\mathbf{Z}_k^-\|_1^{\text{col}} \right], \quad (15)$$

where  $\text{DCT}_{2D}\{\cdot\}$  is 2D DCT being calculated over columns of  $\mathbf{Z}_k^+$  and  $\mathbf{Z}_k^-$ , while  $\|\cdot\|_1^{\text{col}}$  denotes the  $\ell_1$  norm calculated over columns. In the final step, the pixels values are updated as follows:

$$\mathbf{z}_{k+1} = \mathbf{z}_k + \frac{\mathbf{G}}{(2\Delta)}. \quad (16)$$

The gradient is generally proportional to the error between the exact image block  $f$  and its estimate  $z$ . Therefore, the missing values will converge to the true signal values. In order to obtain a high level of the reconstruction precision, the step  $\Delta$  is decreased when the algorithm convergence slows down. Namely, when the pixels values are very close to the exact values, the gradient will oscillate around the exact value and the step size should be reduced; for example,  $\Delta = \Delta/3$ . The stopping criterion can be set using the desired reconstruction accuracy  $\varepsilon$  expressed in dB as follows:

$$\text{MSE} = 10 \log_{10} \frac{\sum |z_p - z_{p-1}|^2}{\sum |z_{p-1}|^2} < \varepsilon, \quad (17)$$

where Mean Square Error (MSE) is calculated as a difference between the reconstructed signals before and after reducing step  $\Delta$ . Here, we use the precision  $\varepsilon = -100$  dB. The same procedure is repeated for each image block, resulting in the reconstructed image  $y$ . Computational complexity of the reconstruction algorithm is analyzed in detail in light of possible implementation of the algorithm in systems (like FPGA) with limited computational resources. The 2D DCT of size  $M \times M$  is observed, with  $M$  being the power of 2. Particularly,  $M = 16$  is used here. Hence, for each observed image block the total number of additions is  $(M - M_A)\{2[(3M/2)(\log_2(M) - 1) + 2]^2 + 4\}$ , where  $M_A$  denotes the available samples, while the total number of multiplications is  $(M - M_A)\{2[M\log_2(M) - 3M/2 + 4]^2 + 7\}$ . Note that the most demanding operation is DCT calculation, which can be done using fast algorithm with  $(3M/2)(\log_2(M) - 1) + 2$  additions and  $M\log_2(M) - 3M/2 + 4$  multiplications. These numbers of operations are squared for the considered 2D signal case.

The performance of CS reconstruction algorithm can be seen in Figure 2 for three different numbers of compressed measurements (compared to the original image dimensionality): 80% of measurements, 50% of measurements, and 20% of measurements. Consequently, we may define the corresponding image degradation levels as 20% degradation, 50% degradation, and 80% degradation.

**2.2. Suspicious Object Detection Algorithm.** Figure 1 includes the general overview of the proposed image processing algorithm. The block diagram implicitly suggests a three-stage operation: the first stage being the preprocessing stage is represented by the top left part of the diagram and the second

stage being the segmentation stage is represented by the lower left part of the diagram, whereas the third stage being the decision-making stage is represented by the right part of the block diagram. It should be noted that the algorithm has been deployed for Croatian Mountain Rescue Service for several months as field assistance tool.

**2.2.1. Image Preprocessing.** The preprocessing stage comprises two parts. At the start, images are converted from the original  $RGB$  to  $YC_bC_r$  color format. Traditional  $RGB$  color format is not convenient for computer vision applications, due to the high correlation between its color components. Next, blue-difference ( $C_b$ ) and red-difference ( $C_r$ ) color components are denoised by applying a  $3 \times 3$  median filter. The image is then divided into nonoverlapping subimages, which are subsequently fed to the segmentation module for further processing. Number of subimages was set to 64 since the number 8 divides both image height and width without the residue and ensures nonoverlapping.

**2.2.2. Segmentation.** The segmentation stage comprises two steps. Each subimage is segmented using the mean shift clustering algorithm [21]. Mean shift is extremely versatile, nonparametric iterative procedure for feature space analysis. When used for color image segmentation, the image data is mapped into the feature space, resulting in a cluster pattern. The clusters correspond to the significant features in the image, namely, dominant colors. Using mean shift algorithm, these clusters can be located and dominant colors can therefore be extracted from the image to be used for segmentation.

The clusters are located by applying a search window in the feature space, which shifts towards the cluster center. The magnitude and the direction of the shift in feature space are based on the difference between the window center and the local mean value inside the window. For  $n$  data points  $x_i$ ,  $i = 1, 2, \dots, n$ , in the  $d$ -dimensional space  $R^d$ , the shift is defined as

$$m_h(x) = \frac{\sum_{i=1}^n x_i g(\|(x - x_i)/h\|^2)}{\sum_{i=1}^n g(\|(x - x_i)/h\|^2)} - x, \quad (18)$$

where  $g(x)$  is the kernel,  $h$  is a bandwidth parameter which is set to the value 4.5 (determined experimentally using different terrain-image datasets),  $x$  is the center of the kernel (window), and  $x_i$  is the element inside kernel. When the magnitude of the shift becomes small, according to the given threshold, the center of the search window is declared as cluster center and the algorithm is said to have converged for one cluster. This procedure is repeated until all significant clusters have been identified.

Once all the subimages have been clustered, a global cluster matrix  $\mathbf{K}$  is formed by merging resulting cluster matrices obtained during subimage segmentation. This new matrix  $\mathbf{K}$  is clustered again using cluster centers (i.e., mean  $C_b$  and  $C_r$  of each cluster) from the previous step, instead of subimage pixel values, as input points. This two-step approach is introduced in order to speed up segmentation,

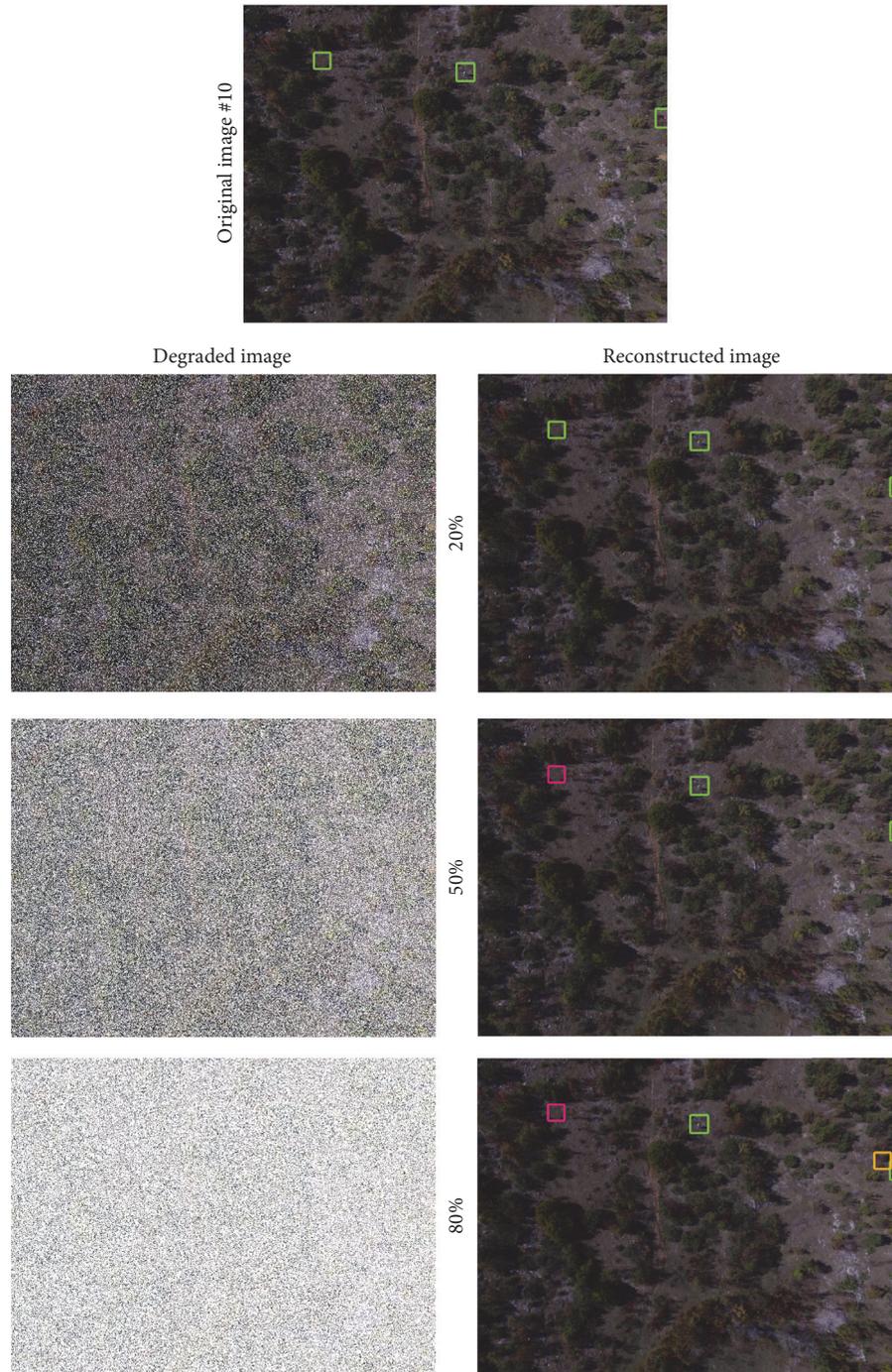


FIGURE 2: Example of compressive sensing image reconstruction algorithm's performance on one of the images (#10) from used dataset. Detection results are also indicated: green squares represent correct detections (in comparison to original image), red squares represent FNs, and orange squares represent FPs.

while keeping almost the same performance. It assures that the number of points for data clustering stays reasonably low in both steps. The number of pixels in subimages is, naturally,  $T$  times smaller than the number of pixels in the original image and the number of cluster centers used in the second step is even smaller than the number of pixels in the first step.

The output of the segmentation stage is a set of candidate regions, each representing a cluster of similarly colored pixels.

The bulk of computational complexity of segmentation step is due to this cluster search procedure and is equal to  $O(N_{X_{\text{sub}}} \times N_{Y_{\text{sub}}})$ , where  $N_{X_{\text{sub}}}$  is number of pixels along  $X$  axis in the subimage, while  $N_{Y_{\text{sub}}}$  is number of pixels

along  $Y$  axis in the subimage. All subsequent steps (including decision-making step) are only concerned with detected clusters, making their computational complexity negligible compared to complexity of mean shift algorithm in this step.

**2.2.3. Decision-Making.** The decision-making stage comprises five steps. In the first step, large candidate regions are eliminated from subsequent processing. The elimination threshold value  $N_{th}$  is determined based on the image height, image width, and the estimated distance between the camera and the observed surface. The premise here is that if such regions were to represent a person, it would mean that the actual person is standing too close to the camera making the search trivial.

The second step is to remove all those areas inside particular candidate regions that contain only a handful of pixels. In this way, the residual noise presented by some scattered pixels left after median filtering is efficiently eliminated. Then, in the third step, the resulting image is dilated by applying a  $5 \times 5$  mask. This is done to increase the size of the connected pixel areas inside candidate regions so that the similar nearby segments can be merged together.

In the next step, the segments, belonging to the cluster with more than three spatially separated areas, are excluded from the resulting set of candidate segments, under the assumption that the image would not contain more than three suspicious objects of the same color. Finally, all the remaining segments that were not eliminated in any of the previous four steps are singled out as potential targets.

More formally, the decision-making procedure can be written as follows. An image  $I$  consists of a set of clusters  $C_i$  obtained by grouping image pixels using only color values of the chosen color model as feature vector elements:

$$I = \bigcup_{i=1}^n C_i, \quad C_i \cap C_j = \emptyset, \quad \forall i, j : i \neq j. \quad (19)$$

As mentioned before, clustering according to similarity of feature vectors is performed using mean shift algorithm. Each cluster  $C_i$  represents a set of spatially connected-component regions or segments  $S_{ik}$ .

$$C_i = \bigcup_{k=1}^m S_{ik}, \quad S_{ik} \cap S_{il} = \emptyset, \quad \forall k, l : k \neq l. \quad (20)$$

In order to accept segment  $S_{ik}$  as potential target, the following properties have to be satisfied:

$$\begin{aligned} p_1: & \text{Size}(C_i) < N_{\max}, \\ p_2: & \text{Size}(S_{ik}) > N_{\min}, \\ p_3: & m \leq N_A, \end{aligned} \quad (21)$$

where  $N_{\max}$  and  $N_{\min}$  are chosen threshold values,  $m$  is the total number of segments within a given cluster, and  $N_A$  denotes the maximum allowed number of candidate segments belonging to the same cluster. For our application,  $N_{\min}$ ,  $N_{\max}$ , and  $N_A$  are set to 10, 38,000, and 3, respectively. Please note that  $N_{\max}$  value represents 0.317% of total pixels in the image and was determined empirically (it encompasses some redundancy; i.e., objects of interest are rarely that large).

### 2.3. Performance Metrics

**2.3.1. Image Quality Metric.** Several image quality metrics are introduced and used in the experiments in order to give better overview of obtained results as well as more weight to the results. It should also be noted that it is not our intention to make conclusions about appropriateness of particular metric or to make their direct comparison but rather to make the results more accessible to wider audience.

Structural Similarity Index (SSI) [22, 23] is inspired by human visual system which is highly accustomed to extracting and processing structural information from images. It detects and evaluates structural changes between two signals (images): reference ( $x$ ) and reconstructed ( $y$ ) one. This makes SSI very consistent with human visual perception. Obtained SSI values are in the range of 0 and 1, where 0 corresponds to lowest quality image (compared to original) and 1 best quality image (which only happens for the exactly the same image). SSI is calculated on small patches taken from the same locations of two images. It encompasses three similarity terms between two images: (1) similarity of local luminance/brightness— $l(x, y)$ , (2) similarity of local contrast— $c(x, y)$ , and (3) similarity of local structure— $s(x, y)$ . Formally, it is defined by

$$\begin{aligned} \text{SSI}(x, y) &= l(x, y) \cdot c(x, y) \cdot s(x, y) \\ &= \left( \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right) \cdot \left( \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right) \\ &\quad \cdot \left( \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \right), \end{aligned} \quad (22)$$

where  $\mu_x$ ,  $\mu_y$  are local sample means of  $x$  and  $y$  image,  $\sigma_x$ ,  $\sigma_y$  are local sample standard deviations of  $x$  and  $y$  images,  $\sigma_{xy}$  is local sample cross-correlation of  $x$  and  $y$  images after removing their respective means, and  $C_1$ ,  $C_2$ , and  $C_3$  are small positive constants used for numerical stability and robustness of the metric.

It can be applied on both color and grayscale images, but for simplicity and without loss of generality it was applied on normalized grayscale images in the paper. It should be noted that SSI is widely used in predicting the perceived quality of digital television and cinematic pictures in practical application, but its performance is sometimes disputed in comparison to MSE [24]. It is used as main performance metric in the experiment due to simple interpretation and its wide usage for image quality assessment.

Peak signal to noise ratio (PSNR) [23] was used as a part of auxiliary metric set for image quality assessment which also included MSE and  $\ell_2$  norm. Missing pixels can in a sense be considered a salt-and-pepper type noise, and thus use of PSNR makes sense since it is defined as a ratio between the maximum possible power of a signal and the power of noise corrupted signal. Larger PSNR indicates better quality reconstructed image and vice versa. PSNR does not have

limited range as is the case with SSI. It is expressed in units of dB and defined by

$$\text{PSNR}(x, y) = 10 \log_{10} \left( \frac{\text{maxValue}^2}{\text{MSE}} \right), \quad (23)$$

where maxValue is maximum range of the pixel, which in normalized grayscale image is 1, and MSE is the Mean Square Error between  $x$  (referent image) and  $y$  (reconstructed image) defined as

$$\text{MSE}(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2, \quad (24)$$

where  $N$  is number of pixels in the  $x$  or  $y$  image (sizes of both images have to be the same) and  $x_i$  and  $y_i$  are normalized values of  $i$ th pixel in the  $x$  and  $y$  image, respectively. MSE is dominant quantitative performance metric for the assessment of signal quality in the field of signal processing. It is simple to use and interpret, has a clear physical meaning, and is a desirable metric within statistics and estimation framework. However, its performance has been criticized in dealing with perceptual signals such as images [25]. This is mainly due to the fact that implicit assumptions related with MSE are in general not met in the context of visual perception. However, it is still often used in the literature when reporting performance in image reconstruction and thus we include it here for comparison purposes. Larger MSE values indicate lower quality images (compared to reference one) while smaller values indicate better quality image. MSE value range is not limited as is the case with SSI.

Final metric used in the paper is  $\ell_2$  metric. It is also called the Euclidian distance and in the work we applied it to the color images. This means that  $\ell_2$  metric represents Euclidian distance between two points in RGB space: the  $i$ th pixel in the original image and the corresponding pixel in the reconstructed image. It is expressed for all pixels in the image and is defined as

$$\begin{aligned} \ell_2(x, y) \\ = \sum_{i=1}^N \sqrt{(R_{xi} - R_{yi})^2 + (G_{xi} - G_{yi})^2 + (B_{xi} - B_{yi})^2}, \end{aligned} \quad (25)$$

where  $N$  is number of pixels in the  $x$  or  $y$  image (sizes of both images have to be the same) for all color channels,  $R_{xi}$ ,  $R_{yi}$  are normalized red color channel values (0-1) of the  $i$ th pixel,  $G_{xi}$ ,  $G_{yi}$  are normalized green color channel values (0-1) of the  $i$ th pixel, and  $B_{xi}$ ,  $B_{yi}$  are normalized blue color channel values (0-1) of the  $i$ th pixel. The larger the value of  $\ell_2$  metric is, the more difference there is between two images. The  $\ell_2$  norm metric is mainly used in image similarity analysis, although there are some situations in which it has been shown that  $\ell_1$  metric can be considered as a proper choice as well [26].

**2.3.2. Detection Quality Metric.** The performance of image processing algorithm for detecting suspicious objects in images for different levels of missing pixels was evaluated

in terms of precision and recall. The standard definitions of these two measures are given by the following equations:

$$\begin{aligned} \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \end{aligned} \quad (26)$$

where TP denotes the number of true positives, FP denotes the number of false positives, and FN denotes the number of false negatives. It should be noted that all of these parameters were determined by checking whether or not a matching segment (object) has been found in the original image, without checking if it actually represents a person or an object of interest. More on accuracy (in terms of recall and precision) of the presented algorithm can be found in Section 3.3.1, where comparison with human performance on original images was made.

When making conclusions based on presented recall and precision values, it should be kept in mind that it is not envisaged as a standalone tool but as cooperative tool for human operators aimed at reducing their workload. Thus, FP values do not have a high price, since human operator can check it and dismiss it if it is a false alarm. More costly are FNs since they can potentially mislead the operator.

### 3. Results and Discussion

**3.1. Database.** For the experiment we used 16 images of 4 K resolution ( $2992 \times 4000$  pixels) obtained at 3 occasions with DJI Phantom 3's gyro stabilized camera. Camera's plane was parallel with the ground and ideally UAV was at the height of 50 m (although this was not always the case as will be explained later on in Section 3.3). All images were taken in coastal area of Croatia in which search and rescue operations often take place. Images 1–7 were taken during Croatian Mountain Rescue Service search and rescue drills (Set 1) and images 8–12 were taken during our mockup testing (Set 2), while images 13–16 were taken during actual search and rescue operation (Set 3).

All images were firstly intentionally degraded with desired level (ranging between 20% and 80%) in a manner that random pixels were set to white (i.e., they were missing). Images were then reconstructed using CS approach and tested for object detection performance.

**3.2. Image Reconstruction.** First we explore how well CS based reconstruction algorithm performs in terms of image quality metrics. Metrics were calculated for each of the images for two cases: (a) when random missing samples were introduced to the original image (i.e., for degraded image) and (b) when original image was reconstructed using gradient based algorithm. For both cases original, unaltered image was taken as a reference. Obtained results for all metrics can be found in Figure 3 in a form of enhanced Box plots.

As can be seen from Figure 3(a), the image quality before reconstruction (as measured by SSI) is somewhat low with mean value in the range of 0.033 to 0.127, and it is significantly

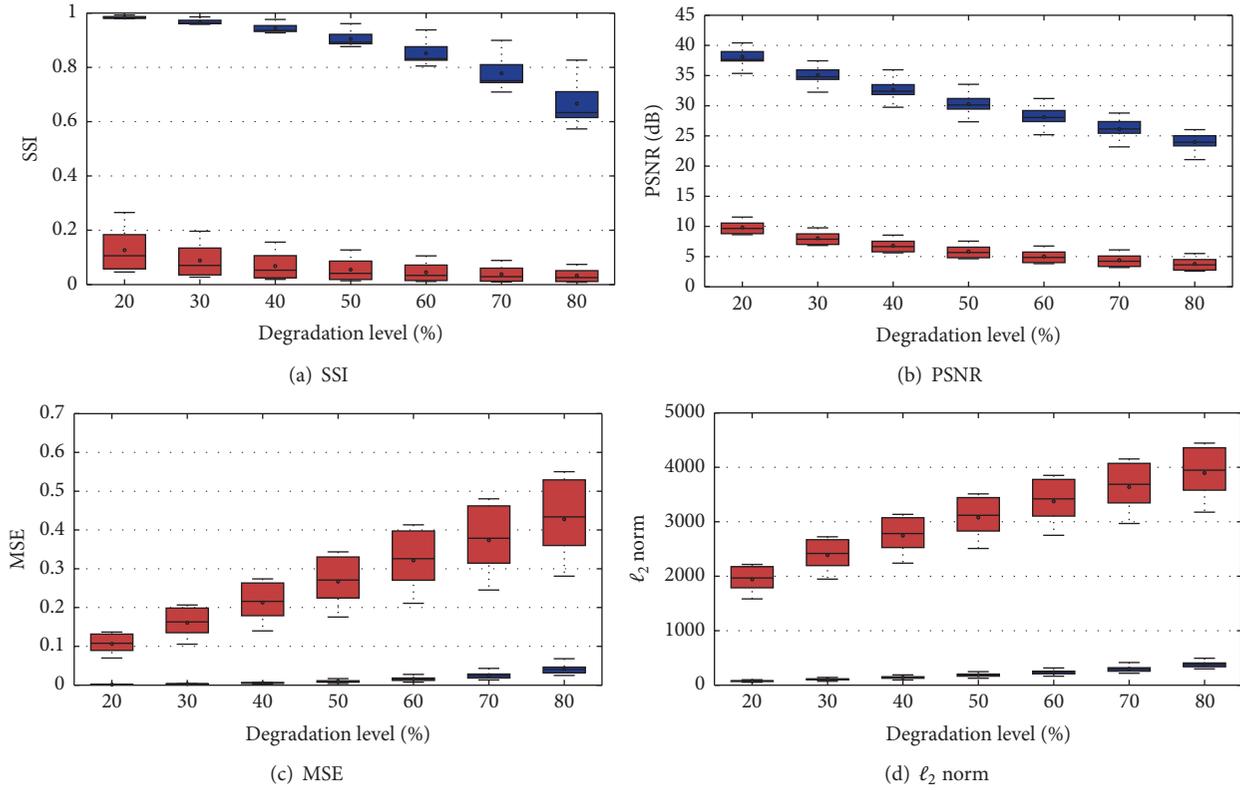


FIGURE 3: Image quality metric for all degradation levels (over all images) before and after reconstruction. Red color represents metric for degraded images, while blue represents postreconstruction values. Black dots represent mean values.

increased after reconstruction having mean values in the range 0.666 to 0.984. Same trend can be observed for all other quality metrics: PSNR (3.797 dB to 9.838 dB range before and 23.973 dB to 38.100 dB range after reconstruction), MSE ( $0.107$  to  $0.428$  range before and  $1.629e - 4$  to  $0.004$  range after reconstruction), and  $\ell_2$  norm ( $1.943e + 3$  to  $3.896e + 3$  range before and  $75.654$  to  $384.180$  range after reconstruction). Please note that for some cases (like in Figures 3(c) and 3(d)) the distribution for particular condition is very tight making its graphical representation quite small. The nonparametric Kruskal-Wallis test [27] was conducted on all metrics for postreconstruction case in order to determine statistical significance of the results with degradation level as independent variable. Statistical significance was detected in all cases with the following values: SSI ( $\chi^2(6) = 98.34$ ,  $p < 0.05$ ), PSNR ( $\chi^2(6) = 101.17$ ,  $p < 0.05$ ), MSE ( $\chi^2(6) = 101.17$ ,  $p < 0.05$ ), and  $\ell_2$  norm ( $\chi^2(6) = 101.17$ ,  $p < 0.05$ ). Tukey's honestly significant difference (THSD) *post hoc* tests (corrected for multiple comparison) were performed, which revealed some interesting patterns. For all metrics statistical difference was present only for those cases that had at least 30% degradation difference between them (i.e., 50% cases were statistically different only from 20% and 80% cases; please see Figure 2 for visual comparison). We believe this goes towards demonstrating quality of obtained reconstruction (in terms of presented metrics); that is, there is no statistical difference between 20% and 40% missing

sample cases (although their means are different in favor of 20% case). It should also be noted that, even in cases of 70% or 80% of pixels missing, reconstructed image was subjectively good enough (please see Figure 2) so that its content could be recognized by the viewer (with SSI means of 0.778 and 0.666, resp.). However it should be noted that in cases of such high image degradation (especially for 80% case) reconstructed images appeared somewhat smudged (pastel like effect) with some details (subjectively) lost.

It is interesting to explore how much gain was obtained with CS reconstruction. This is depicted in Figure 4 for all metrics. From presented figures it can be seen that obtained improvement is significant and, while its absolute value varies, detected by all performance metrics. The improvement range is between 3.730 and 81.637 for SSI, between 3.268 and 9.959 for PSNR, between  $9.612e - 4$  and 0.0205 for MSE, and between 0.029 and 0.143 for  $\ell_2$  norm. From this a general trend can be observed: the improvement gain increases with the degradation level. There are a couple of exceptions to this general rule: for SSI in 12 out of 16 images gain for 70% is larger than for 80% case. However, this phenomenon is not present in other metrics (except for the case of PSNR images #13 and #14 where they are almost identical) which might suggest this is due to some interaction between metric and type of environment in the image. Also, randomness of pixel removal when degrading image should be considered.

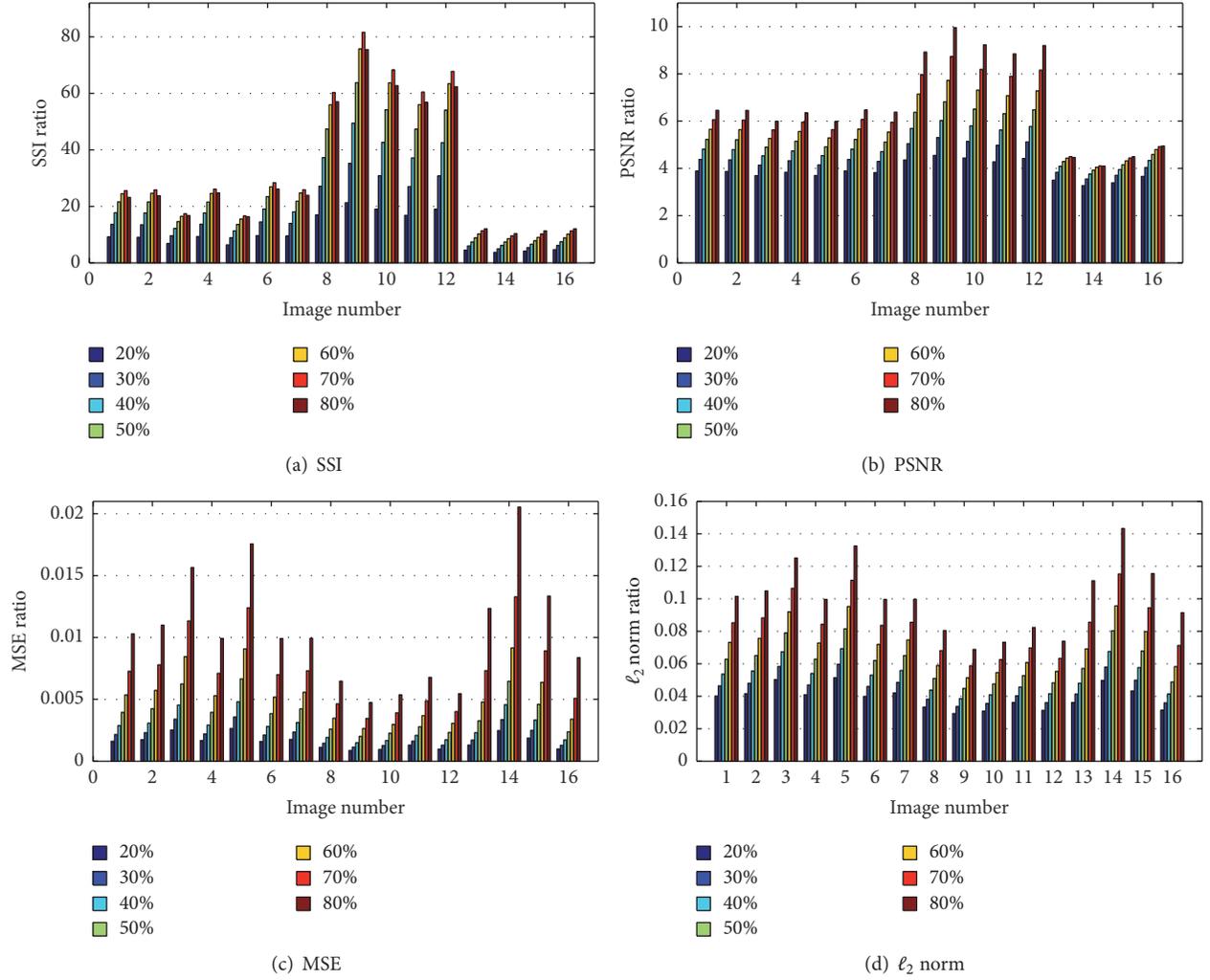


FIGURE 4: Ratios of image quality metrics for all degradation levels (over all images) after and before reconstruction.

Figure 4 also reveals another interesting observation that warrants further research: reconstruction gain and performance might depend on type of environment/scenery in the image. Taking into consideration that dataset contains images from three distinct cases (sets) as described earlier (which all have unique environment in them) and that their range is 1–7, 8–12, and 13–16, different patterns (gain amplitudes) can be seen for all metrics.

In order to detect if this observed pattern has statistical significance, nonparametric Kruskal-Wallis test (with *post hoc* THSD) was performed in a way that for particular degradation level data for images were grouped into three categories based on which set they belong to. Obtained results revealed that for all metrics and degradation levels there exists statistical significance (with varying levels of  $p$  value which we omit here for brevity) of image set on reconstruction improvement. *Post hoc* tests revealed that this difference was between image sets 2 and 3 for SSI and PSNR metrics, while it was between sets 1 and 2/3 for  $\ell_2$  norm (between sets 1 and 2-3 for degradation levels 20%, 30%, and 40% and between sets 1 and 3 for degradation levels 50%, 60%, 70%, and 80%). For

MSE metric the difference was between sets 1 and 2 for all degradation levels, with addition of difference between sets 2 and 3 for 70% and 80% cases. While all metrics do not agree with where the difference is, they clearly demonstrate that terrain type influences algorithms performance which could be used (in conjunction with terrain type classification like the one in [28]) to estimate expected CS reconstruction performance before deployment.

Another interesting analysis in line with the last one can be made so to explore if knowing image quality before the reconstruction can be used to infer image quality after the reconstruction. This is depicted in Figure 5 for all quality metrics across all used test images. Figure 5 suggests that there exists clear relationship between quality metrics before and after reconstruction which is to be expected. However, this relationship is nonlinear for all cases, although for case of PSNR it could be considered (almost) linear. This relationship enables one to estimate image quality after reconstruction and (in conjunction with terrain type as demonstrated before) select optimal degradation level (in terms of reduced data load) for particular case while maintaining image quality

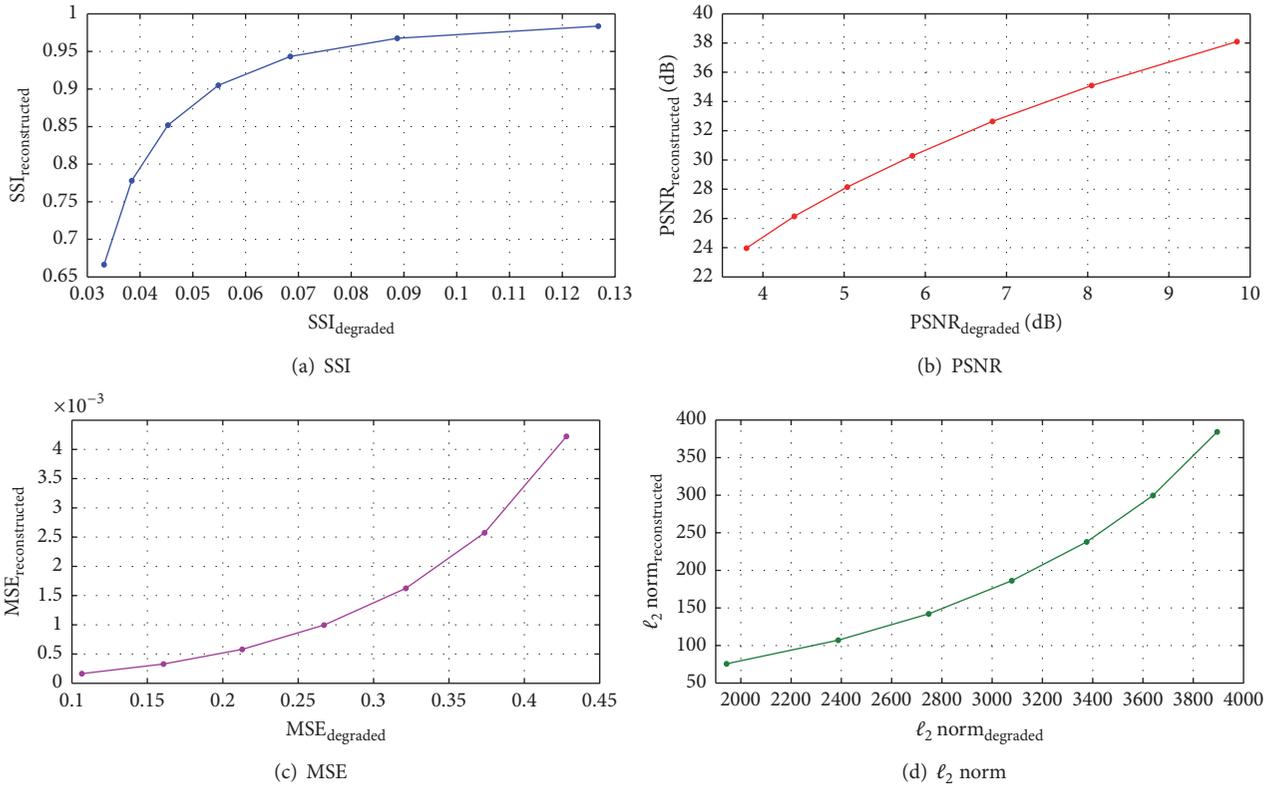


FIGURE 5: Dependency of image quality metrics after reconstruction on image quality metrics before reconstruction across all used test images. Presented are means for seven degradation levels in ascending order (from left to right).

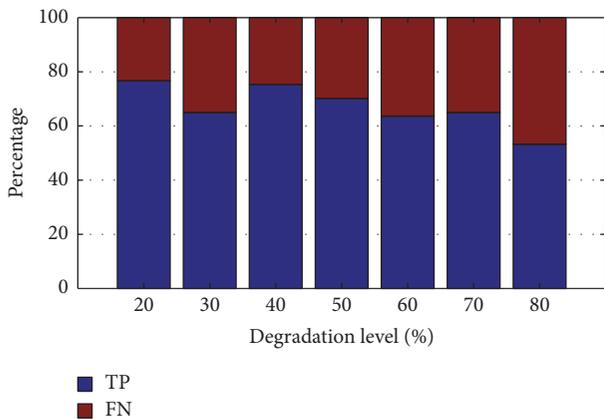


FIGURE 6: Relation between true positives (TPs) and false negatives (FNs) for particular degradation level over all images.

at desired level. Since pixel degradation is random, one should also expect certain level of deviation from presented means.

**3.3. Object Detection.** As stated before, due to intended application, FNs are more expensive than FPs and it thus makes sense to first explore number of total FNs in all images. Figure 6 presents number of FNs and TPs in percentage of total detections in the original image (since they always

sum up to the same value for particular image). Please note that reasons behind choosing detections in the original (nondegraded) images as the ground truth for calculations of FNs and TPs are explained in more detail in Section 3.3.1.

Figure 6 depicts that there is more or less a constant downward trend in TP rate; that is, FN rates increase as degradation level increases. Exception is 30% level. This dip in TP rate at 30% might be considered a fluke, but it appeared in another unrelated analysis [9] (on completely different image set with same CS and detection algorithms). Since currently reason for this is unclear, it should be investigated in the future. As expected, FN percentage was the lowest for 20% case (23.38%) and the highest for 80% case (46.74%). It has value of about 35% for other cases (except 50% case). This might be considered a pretty large cost in search and rescue scenario, but while making this type of conclusion it should be kept in mind that comparisons are made in respect to algorithms performance on original image and not ground truth (which is hard to establish in this type of experiment where complete control of environment is not possible).

Some additional insight will be gained in Section 3.3.1, where we conducted minuser study on original image to detect how well humans perform on ideal (and not reconstructed) image. For completeness of presentation, Figure 7 showing comparison of numbers of TPs and FPs for all degradation levels is included.

Additional insight in algorithm's performance can be obtained by observing recall and precision values in Figure 8,

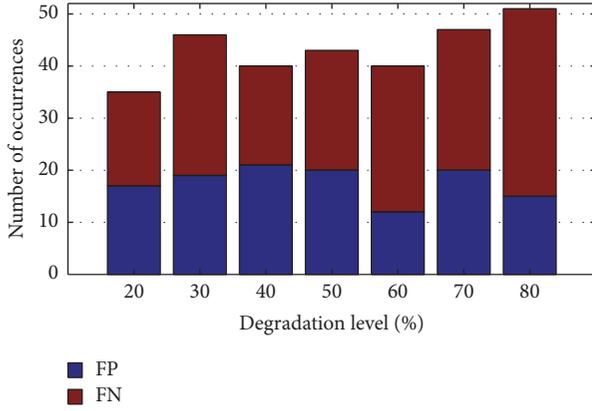


FIGURE 7: Relation of number of occurrences of false positives (FPs) and false negatives (FNs) for particular degradation level over all images.

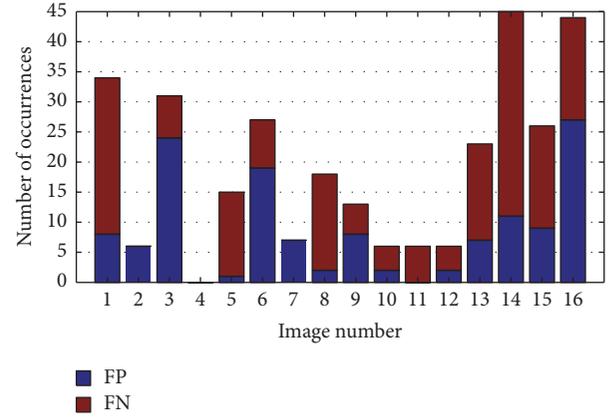


FIGURE 9: Number of occurrences of false negatives (FNs) and false positives (FPs) for all images across all degradation levels.

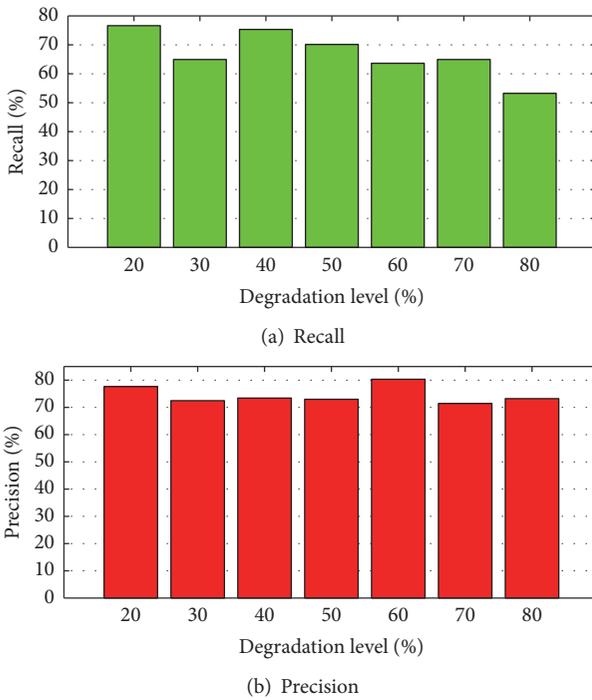


FIGURE 8: Recall and precision as detection performance metrics for all degradation levels across all images.

as defined by (26). The highest recall value (76.62%) is achieved in the case of 20% (missing samples or degradation level) followed closely by 40% case (75.32%). As expected (based on Figure 5), there is a significant dip in steady downward trend for 30% case. At the same time precision is fairly constant (around 72%) over the whole degradation range with two peaks for the case of 20% of missing samples (peak value 77.63%) and for 60% of missing samples (peak value 80.33%). No statistical significance (as detected by Kruskal-Wallis test) was detected for recall and precision in which degradation level was considered independent variable (across all images).

If individual images and their respective FP and FN rates are examined across all degradation levels, Figure 9 is obtained. Some interesting observations can be made from the figure. First, it can be seen that images #14 and #16 have unusually large number of FPs and FNs. This should be viewed in light that these two images belong to third dataset. This dataset was taken during real search and rescue operation during which UAV operator did not position UAV at the desired/required height (of 50 m) and also strong wind gusts were present. Also image #1 has unusually large number of FNs. If images #1 and #14 are removed from the calculation (since we care more about FNs and not FPs, as explained before), recall and precision values increase up to 12% and 5%, respectively. Increase in recall is the largest for 40% case (12%), while it is the lowest for 50% case (5%), while increase in precision is the largest for 80% case (5%) and the smallest for 50% case (2.5%). Second observation that can be made from Figure 9 is that there are number of cases (2, 4, 7, 10, 11, and 12) where algorithm detection performs really well with cumulative number of occurrences of FPs and FNs around 5. In case of image #4 it performed flawlessly (in image there was one target that was correctly detected in all degradation levels) without any FP or FN. Note that no FNs were present in images #2 and #7.

**3.3.1. User Study.** While we maintain that for evaluation of compressive sensing image reconstruction's performance comparison of detection rates in the original, nondegraded image (as ground truth) and reconstructed images are the best choice, baseline performance of detection algorithm is presented here for completeness. However, determining baseline performance (absolute ground truth) proved to be nontrivial, since environmental images from the wild (where search and rescue usually takes place) is hard to control and usually there are unintended objects in the frame (e.g., animals or garbage). Thus, we opted for 10-subject mini-study in which the decision whether there is an object in the image was made by majority vote; that is, if 5 or more people detected something in certain place of the image, then it would be considered an object. Subjects were from faculty

and student population and did not see the images before nor have ever participated in search and rescue via image analysis. Subjects were instructed to find people, cars, animals, cloth, bags, or similar things in the image combining speed and accuracy (i.e., not emphasizing either). They were seated in front of 23.6-inch LED monitor (Philips 247E) on which they inspected images and were allowed to zoom in and out of the image as they felt necessary. Images were randomly presented to subject to avoid undesired (learning/fatigue) effects.

On average, it took a subject 638.3 s (10 minutes and 38.3 seconds) to go over all images. In order to demonstrate inter-subject variability and how demanding the task was, we analyzed precision and recall for results from test subject study. For example, if 6 out of 10 subjects marked a part of an image as an object, this meant that there were 4 FNs (i.e., 4 subjects did not detect that object). On the other hand, if 4 test subjects marked an area in an image (and since it did not pass threshold in majority voting process), it would be counted as 4 FPs. Analysis conducted in such manner yielded recall of 82.22% and precision of 93.63%. Here again, two images (#15 and #16) accounted for more than 50% of FNs. It should be noted that these results cannot be directly compared to the proposed algorithm since they rely on significant human intervention.

#### 4. Conclusions

In the paper, gradient based compressive sensing is presented and applied to images acquired from UAV during search and rescue operations/drills in order to reduce network/transmission burden. Quality of CS reconstruction is analyzed as well as its impact on object detection algorithms in use with search and rescue. All introduced quality metrics showed significant improvement with varying ratios depending on degradation level as well as type of terrain/environment depicted in particular image. Dependency of reconstruction quality on terrain type is interesting and opens up possibility of inclusion of terrain type detection algorithms, and since then reconstruction quality could be inferred in advance and appropriate degradation level (i.e., in smart sensors) selected. Dependency on quality of degraded image is also demonstrated. Reconstructed images showed good performance (with varying recall and precision parameter values) within object detection algorithm, although slightly higher false negative rate (whose cost is high in search applications) is present. However, there were few images in the dataset on which algorithm performed either flawlessly with no false negatives or with few false positives whose cost is not big in current application setup with human operator checking raised alarms. Of interest is slight peak in performance of 30% degradation level (compared to 40% and general downward trend)—this peak was detected in earlier study on completely different image set making this find by chance unlikely. Currently no explanation for the phenomenon can be provided and it warrants future research. Nevertheless, we believe that obtained results are promising (especially in light of results of minuser study) and require further research, especially on the detection side. For example, algorithm could be augmented with terrain recognition algorithm which could give cues about terrain type to both the reconstruction

algorithm (adjusting degradation level while keeping desired level of quality for reconstructed images) and detection algorithm augmented with automatic selection procedure for some parameters like mean shift bandwidth (for performance optimization). Additionally, automatic threshold estimation using image size and UAV altitude could be used for adaptive configuration of detection algorithm.

#### Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

#### References

- [1] UK Ministry of Defense, "Military search and rescue quarterly statistics: 2015 quarter 1," Statistical Report, 2015, [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/425819/SAR\\_Quarter1\\_2015\\_Report.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/425819/SAR_Quarter1_2015_Report.pdf).
- [2] T. W. Heggie and M. E. Amundson, "Dead men walking: search and rescue in US National Parks," *Wilderness and Environmental Medicine*, vol. 20, no. 3, pp. 244–249, 2009.
- [3] M. Šuperina and K. Pogačić, "Frequency of the Croatian mountain rescue service involvement in searching for missing persons," *Police and Security*, vol. 16, no. 3-4, pp. 235–256, 2008 (Croatian).
- [4] J. Sokalski, T. P. Breckon, and I. Cowling, "Automatic salient object detection in UAV imagery," in *Proceedings of the 25th International Conference on Unmanned Air Vehicle Systems*, pp. 11.1–11.12, April 2010.
- [5] H. Turić, H. Dujmić, and V. Papić, "Two-stage segmentation of aerial images for search and rescue," *Information Technology and Control*, vol. 39, no. 2, pp. 138–145, 2010.
- [6] S. Waharte and N. Trigoni, "Supporting search and rescue operations with UAVs," in *Proceedings of the International Conference on Emerging Security Technologies (EST'10)*, pp. 142–147, Canterbury, UK, September 2010.
- [7] C. Williams and R. R. Murphy, "Knowledge-based video compression for search and rescue robots & multiple sensor networks," in *International Society for Optical Engineering: Unmanned Systems Technology VIII*, vol. 6230 of *Proceedings of SPIE*, May 2006.
- [8] G. S. Martins, D. Portugal, and R. P. Rocha, "On the usage of general-purpose compression techniques for the optimization of inter-robot communication," in *Proceedings of the 11th International Conference on Informatics in Control, Automation and Robotics (ICINCO'14)*, pp. 223–240, Vienna, Austria, September 2014.
- [9] J. Musić, T. Marasović, V. Papić, I. Orović, and S. Stanković, "Performance of compressive sensing image reconstruction for search and rescue," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 11, pp. 1739–1743, 2016.
- [10] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [11] R. Chartrand, "Exact reconstruction of sparse signals via non-convex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, 2007.
- [12] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Springer, 2013.
- [13] M. F. Duarte, M. A. Davenport, D. Takbar et al., "Single-pixel imaging via compressive sampling: building simpler, smaller,

- and less-expensive digital cameras," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.
- [14] L. Stanković, "ISAR image analysis and recovery with unavailable or heavily corrupted data," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 3, pp. 2093–2106, 2015.
- [15] M. Brajović, M. Daković, I. Orović, and S. Stanković, "Gradient-based signal reconstruction algorithm in Hermite transform domain," *Electronics Letters*, vol. 52, no. 1, pp. 41–43, 2016.
- [16] I. Stanković, I. Orović, and S. Stanković, "Image reconstruction from a reduced set of pixels using a simplified gradient algorithm," in *Proceedings of the 22nd Telecommunications Forum Telfor (TELFOR '14)*, pp. 497–500, Belgrade, Serbia, November 2014.
- [17] G. Reeves and M. Gastpar, "Differences between observation and sampling error in sparse signal reconstruction," in *Proceedings of the IEEE/SP 14th Workshop on Statistical Signal Processing (SSP '07)*, pp. 690–694, Madison, Wis, USA, August 2007.
- [18] R. E. Carrillo, K. E. Barner, and T. C. Aysal, "Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise," *IEEE Journal on Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 392–408, 2010.
- [19] H. Rauhut, "Compressive sensing and structured random matrices," in *Theoretical Foundations and Numerical Methods for Sparse Recovery*, M. Fornasier, Ed., vol. 9 of *Random Series Computational and Applied Mathematics*, pp. 1–92, Walter de Gruyter, Berlin, Germany, 2010.
- [20] G. E. Pfander, H. Rauhut, and J. A. Tropp, "The restricted isometry property for time-frequency structured random matrices," *Probability Theory and Related Fields*, vol. 156, no. 3–4, pp. 707–737, 2013.
- [21] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
- [22] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [23] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR '10)*, pp. 2366–2369, Istanbul, Turkey, August 2010.
- [24] R. Dosselmann and X. D. Yang, "A comprehensive assessment of the structural similarity index," *Signal, Image and Video Processing*, vol. 5, no. 1, pp. 81–91, 2011.
- [25] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it?" *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.
- [26] R. Russell and P. Sinha, "Perceptually-based comparison of image similarity metrics," Tech. Rep., Massachusetts Institute of Technology (MIT)—Artificial Intelligence Laboratory, 2001.
- [27] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.
- [28] A. Angelova, L. Matthies, D. Helmick, and P. Perona, "Fast terrain classification using variable-length representation for autonomous navigation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, Minneapolis, Minn, USA, June 2007.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

