

# TERRESTRIAL LASER SCANNER DATA DENOISING BY DICTIONARY LEARNING OF SPARSE CODING

E. Smigiel\*, E. Alby, P. Grussenmeyer

Icube laboratory, Photogrammetry and Geomatics, Graduate School of Science and Technology (INSA),  
24 Boulevard de la Victoire, 67084 STRASBOURG, France  
(eddie.smigiel, emmanuel.alby, pierre.grussenmeyer)@insa-strasbourg.fr

**KEY WORDS:** Terrestrial Laser Scanning, Compressive sensing, Single-pixel camera, Denoising, Accuracy

## ABSTRACT:

Point cloud processing is basically a signal processing issue. The huge amount of data which are collected with Terrestrial Laser Scanners or photogrammetry techniques faces the classical questions linked with signal or image processing. Among others, denoising and compression are questions which have to be addressed in this context. That is why, one has to turn attention to signal theory because it is susceptible to guide one's good practices or to inspire new ideas from the latest developments of this field. The literature have been showing for decades how strong and dynamic, the theoretical field is and how efficient the derived algorithms have become. For about ten years, a new technique has appeared: known as compressive sensing or compressive sampling, it is based first on sparsity which is an interesting characteristic of many natural signals. Based on this concept, many denoising and compression techniques have shown their efficiencies. Sparsity can also be seen as redundancy removal of natural signals. Taken along with incoherent measurements, compressive sensing has appeared and uses the idea that redundancy could be removed at the very early stage of sampling. Hence, instead of sampling the signal at high sampling rate and removing redundancy as a second stage, the acquisition stage itself may be run with redundancy removal. This paper gives some theoretical aspects of these ideas with first simple mathematics. Then, the idea of compressive sensing for a Terrestrial Laser Scanner is examined as a potential research question and finally, a denoising scheme based on a dictionary learning of sparse coding is experienced. Both the theoretical discussion and the obtained results show that it is worth staying close to signal processing theory and its community to take benefit of its latest developments.

## 1. INTRODUCTION

Terrestrial Laser Scanners have been used for about fifteen years and have now become a popular and affordable technology. Many impressive improvements have been achieved on the speed, precision and accuracy. On the other hand, they produce a huge amount of data which have to be stored and processed. As an imaging device, one should be aware of the latest achievements in signal and image processing to be sure to use the most efficient schemes to deal with the difficulty of handling big volumes of data.

This paper deals with recent developments in signal and image processing and the way it may shed some light on how Terrestrial Laser Scanners' data could be handled.

To illustrate how these recent developments interact with the practical usage of TLS, it will be shown that it is possible to improve denoising of the point cloud by using dictionary learning of sparse coding.

The paper is organized as follows: paragraph 2 gives a brief mathematical description of how a general signal (or an image) is considered in modern signal theory: based on this description, compression and denoising are presented as practical applications of sparsity which is an important characteristic of most natural signal. Then, a quite recent approach known as compressive sensing is explained. It is then showed that a TLS could be described as a single-pixel camera with the particularity that it is a range camera. The question of finding out how range could be the object of compressive sensing is addressed and presented as a potential research program. Paragraph 3 recalls how denoising of the point cloud can be obtained by denoising of the individual range images produced

by the scanner before registration. Then some initial experimental results of Range Image denoising are presented. Last, paragraph 4 suggests further developments since so far, the experimental results prove the interest of the principle of the method but have not allowed to obtain a complete 3D point cloud with substantial denoising.

## 2. VECTOR SPACE REPRESENTATION OF SIGNALS

For a long time, signals have been obtained by analog sensors and processed by analog electronic devices. In this case, signal description used to be done in the Hilbert space which arises naturally and frequently in mathematics, physics and engineering, typically as infinite-dimensional function spaces, when the sensed signal is analog and represented by a square-integrable function. The corresponding mathematics are difficult and subtle, not easily reached by the non-specialist. More and more, however, signals are digitized, sometimes, digital native, and the corresponding mathematics are easier to handle as a simple generalization of the classical and well-known euclidean space. As it is quite easy to understand with simple maths, it is worth giving some explanation to allow anyone to get into the issue of signal and image processing. In the discrete case, both sampled and digitized, a signal is a vector (which may be expressed a priori as a column or row vector) of an N-dimensional vector space and is written:

$$x = x_0 \delta_0 + x_1 \delta_1 + \dots + x_i \delta_i + \dots + x_{N-1} \delta_{N-1} \quad (1)$$

or

$$x = \sum_{i=0}^{i=N-1} x_i \cdot \delta_i \quad (2)$$

where:

$$\delta_0=[10\dots0], \delta_1=[010\dots0], \delta_{N-1}=[00\dots01] \quad (3)$$

Hence, in the discrete signal context, a signal, as a simple vector, may be understood as a generalization of the classical and well-known vectors of the 3-dimension euclidean space with the difference that the N-dimension may be quite high. Hence, if one works with a six million pixel image, provided that the rows are merged together in a single final row, one works in a six million-dimension vector space. However, despite the hugeness of the dimension, all the classical concepts of the simple 3-dimension euclidean space are easily generalized to any N-dimension including very high dimensions.

For instance and among others, we recall some of the basic formulas extended to the N-dimension.

The inner product of two signals is defined either by:

$$\langle x, y \rangle = \sum_{i=0}^{i=N-1} x_i \cdot y_i \quad (4)$$

or by

$$\langle x, y \rangle = \sum_{i=0}^{i=N-1} x_i \cdot \bar{y}_i \quad (5)$$

where the overline stands for the conjugate if the signals are complex. This inner product is an essential operation on discrete signals as we will see since it is the basis for signal projection and hence, the so-called transforms.

The norm of a signal is then defined as:

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{i=1}^{i=N} x_i \cdot \bar{x}_i} \quad (6)$$

in case of a complex signal (the corresponding definition for a real signal is evident).

One can also define the quadratic or euclidean distance (other definitions exist) between two signals by:

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=0}^{i=N-1} (x_i - y_i) \cdot \overline{(x_i - y_i)}} \quad (7)$$

Once these definitions are well understood, it is easy to interpret signal processing as fundamental geometric processes in this N-dimensional vector space. In particular, it becomes easy to interpret the so-called transforms and particularly, the discrete Fourier Transform in a geometric way. If we turn back to equation (1), one can read the signal as written as a linear combination of the canonical basis of the  $\delta_i$  vector set. However, it remains possible to rewrite the very same vector or signal in another basis vector set which is nothing else but a change of basis or change of frame. And this is basically the great idea of Transforms: it consists in rewriting the very same signal or vector in a basis set where the same vector has nice properties. As probably the most famous one, the Discrete Fourier Transform is defined as follows: one needs a set of N vectors (since the space is N-dimensional). The set is obtained by sampling of the N complex exponential functions:

$$y_n(k) = e^{j2\pi \frac{nk}{N}} \quad (8)$$

for  $n=0, 2, \dots, N-1$  and each exponential function is sampled on N samples ( $k=0, 1, 2, \dots, N-1$ ).

Thus, each component of the Discrete Fourier Transform (TFD) is obtained by projection of the initial vector written as the linear combination of the canonical basis on the new basis vectors and this projection is as in the classical 3D euclidean space is obtained through a mere inner product:

$$X(n) = \sum_{k=0}^{k=N-1} x(k) e^{-j2\pi \frac{nk}{N}} \quad (9)$$

which should be interpreted as the inner product between the initial vector and the nth exponential sampled function. When one makes n vary from 0 to N-1, one obtains N projections which are the N coefficients of the linear combination of the exponential vector basis which is another expression, in a different basis, of the very same initial vector. Thus, one may retrieve the initial vector as expressed in the canonical basis through its coefficients which define the Inverse Discrete Fourier Transform by:

$$x(k) = \frac{1}{N} \sum_{n=0}^{n=N-1} X(n) e^{j2\pi \frac{nk}{N}} \quad (10)$$

To get the point of these transforms and how they allow to obtain pertinent representations, let us see how a particular signal, which is not intelligible in the canonical basis because it is highly disturbed by a big amount of noise, a so-called 'white noise' whose signification will become evident below. This signal consists of a linear combination of sine signals at the frequencies, 100 Hz, 200 Hz, 300 Hz and 400 Hz with variable magnitudes. But the signal is noisy in the sense that a so-called 'white noise' has been added with a quite large amplitude to the extent that the direct representation of the vector (one thousand samples) does not allow to recognize the sine nature of its components. Figure 1 shows the initial signal/vector in the canonical basis also called direct space. Figure 2 shows the very same signal or vector after change of basis. In the signal theory, one says that figure 2 shows the Discrete Fourier Transform of the signal but, actually, it is the very same signal however expressed in another basis. In this figure, the information content becomes clearly readable. The fine and high peaks correspond to the presence of energy at the corresponding frequencies (see above) and the sharpness of these peaks indicate that the signal is composed with pure sines.

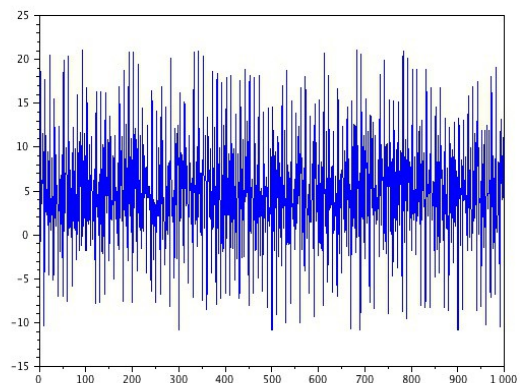


Fig. 1: a signal composed with four pure sines and white noise.

The so-called 'white noise' appears on figure 2 as the low amplitude signal spread over the whole range of the horizontal axis. This kind of noise is called white in the sense that the energy in this noise signal is equally distributed on the whole range of frequencies. Now, the practical applications become evident. Since the information content for this particular signal has been exhibited in a very pertinent way by calculating the Discrete Fourier Transform, one may instantly think about two major applications quite easy to understand. Denoising and compression. They are both directly linked with a simple but rich concept: sparsity. The initial signal has a very valuable characteristic, it is sparse in the sense that there exists a basis in which only a few coefficients of the linear combination of the basis vector set are non zero, if we reject the white noise which is a perturbation in the initial signal. Then, in the denoising scheme, one puts to zero all the coefficients which are too low

to be likely to belong to the real signal and which are likely to belong to the white noise. One retains only the few coefficients with appreciable amplitude. Hence, the noise is removed and only the actual signal remains. This removal of the great majority of coefficients is carried out in the frequency space after having applied the Discrete Fourier Transform. Then, one goes back to the canonical basis by applying the Inverse Discrete Fourier Transform according to equation (10). Of course, the removal is not perfect since there is a risk to get rid of small coefficients which are some actual signal. Hence, the quality of the method is based on a good trade-off and on finding the good threshold which separates correctly the actual signal components and the white noise.

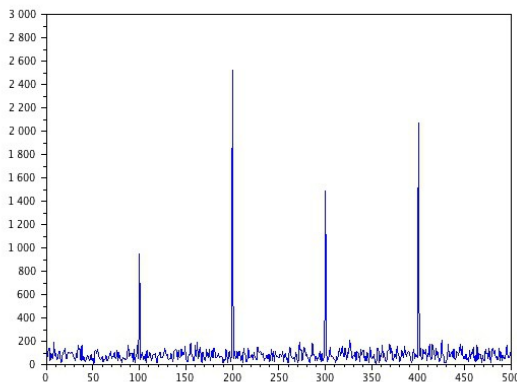


Fig. 2: the same signal as in figure 1 but expressed as a combination of complex exponential vectors. The so-called Discrete Fourier Transform of the signal in figure 1.

One may have noticed that by denoising by removal of small coefficients, one has in the same time obtained signal compression since the denoised signal is expressed by a linear combination with few coefficients. The compression is lossy as denoising was imperfect since the removal has also removed some components of the actual signal. On the other hand, by increasing the threshold, one puts to zero more coefficients which allows one to set the compression rate with the obvious rule that the bigger the rate, the lossier the compression.

The Discrete Fourier Transform based on the projection on complex exponential vectors is by far the most famous one. However, other transforms exist. The idea of sparse coding is to find out the appropriate basis, if it exists, in which few coefficients are non zero. If such a basis exists, the signal is said to be sparse or compressible. The difficulty is to determine the good basis. In the 80s, an important family of vectors has been introduced: the so-called wavelets and the corresponding Wavelet Transforms and their expansions known as curvelets or bandlets which have obtained impressive compression rates in the field of image compression. The famous JPEG algorithm uses a Discrete Cosine Transform which is a close variant of the Discrete Fourier Transform whereas JPEG 2000 uses Discrete Wavelet Transforms.

Figure 3 shows a one million pixel image (a) and its wavelet coefficients (b) obtained by projection of the image on a wavelet basis. Relatively few coefficients contain most of the signal energy and the image is said to be highly compressible. The reconstruction (c) is built after having zeroed out all the small coefficients but the 25 000 largest. At the human eye, the difference between the original image and its reconstruction is hardly noticeable.

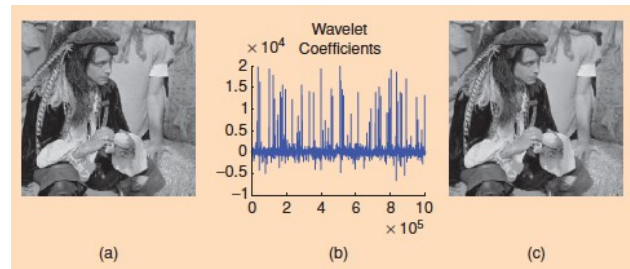


Fig. 3: the original image (a), its wavelet coefficients (b) and reconstructed image (c) after zeroing of the small coefficients.

Hence, sparsity which is a common characteristic of many natural signals is a powerful idea as regard to compression and denoising and we have seen that both processes are based on the same method: zeroing a large amount of coefficients which do not contain much information of the actual signal but rather contain noise energy. However, this sample-then-compress framework suffers from different inherent inefficiencies: first, the signal has to be sampled at a relatively high sampling rate ( $N$  samples) even if the ultimate remaining coefficients is quite low ( $K$  coefficients) with  $K \ll N$ . Second,  $N$  coefficients have to be computed by projection of the initial signal on the sparsifying basis even though all but  $K$  are to be discarded. Last, the compression encoder faces the overhead of encoding properly the large remaining coefficients. In other words, compression is based on redundancy of the  $N$ -samples representation of the initial signal but before getting rid of this redundancy, one has to acquire it.

Recent developments in advanced mathematics and signal theory have allowed to imagine an alternative known as compressive sensing or compressive sampling. It is far beyond the scope of this paper to describe the underlying mathematics which are not as simple as above. We will just give without any proof nor any theorem the general ideas leading to compressive sensing and the main astonishing results which may be obtained to show the great value of this new concept. Then, in the next paragraph, we will show how a terrestrial laser scanner could take benefit of implementing these new ideas. The complete description, theorems and proofs may be found in a very abundant literature which can be accessed through a number of introducing papers (Cadès et al., 2005, 2006, 2007, 2008; Donoho, 2006; Lustig, 2006).

We start the explanation by getting back to the first step in transform coding which represents the signal by the coefficients of an orthonormal sparsifying basis (named  $\Psi$ ) expansion.

Now the basic idea of compressive sensing is to bypass the  $N$ -sampling process and directly acquire a condensed representation using  $M < N$  linear measurements between the initial signal  $x$  and a collection of test vectors named  $\Phi$ . So, in other words, the idea of compressive sensing consists in projecting the initial signal in a  $M$ -dimensional subspace described by the  $\Phi$  basis which is independent from the initial signal and reconstruct the signal with an other basis named  $\Psi$  in which the signal is sparse. All the difficulty is to find the right pair of basis and evidences have been found that the problem may be solved by probabilistic approach. To make a long story short, in practice, we employ a pseudo-random  $\Phi$  driven by a pseudo-random number generator. Then, the signal is reconstructed by a range of alternative reconstruction techniques based on greedy, stochastic, and variational algorithms. If we turn back to figure 3, experiments have shown

that the initial image can be nicely be recovered by a little less than 100 000 measurements instead of the one million original image. Thus instead of acquiring the image redundancy before getting rid of it, the acquisition step itself inherently removes redundancy.

Intensity imaging sensors have improved dramatically in recent times thanks to the introduction of CCD and CMOS digital technology. Consumer digital cameras in the megapixel range are now ubiquitous thanks to the happy coincidence that the semiconductor material of choice for large-scale electronics integration (silicon) also happens to readily convert photons at visual wavelengths into electrons. On the contrary, imaging at wavelengths where silicon is blind is considerably more complicated, bulky, and expensive if ones intend to build a 2D array of detectors at high resolution. Recently, a new approach named single-pixel camera based on compressive sensing has been introduced. These single-pixel compressive sensing cameras are based on the computation of random linear measurements of the scene under view. The image is then recovered or processed from the measurements by a digital computer. The camera design reduces the required size, complexity, and cost of the photon detector array down to a single unit, which enables the use of exotic detectors that would be impossible in a conventional digital camera. The random compressive sensing measurements also enable a tradeoff between space and time during image acquisition. Finally, since the camera compresses as it images, it has the capability to efficiently and scalably handle high-dimensional data sets. Theory and experimental results are found in the literature and prove the efficiency of this approach (Chan et al., 2008; Duarte et al., 2009; Ma, 2009).

Somehow, a Terrestrial Laser Scanner may be understood as a single-pixel Range Image camera. However, based mainly on a time-of-flight technique (directly or through phase shift), it is not possible to imagine a 2D array of range sensors. A potential research program would then be to adapt these initial ideas successfully tried on intensity cameras also on Terrestrial Laser Scanners which can be seen as a raster Range Camera.

### 3. RANGE IMAGE DENOISING BY DICTIONARY LEARNING OF SPARSE CODING

We have been carrying out works on Range Image denoising since 2008. We have first given the general approach which consists in denoising a 3D point cloud obtained by TLS by denoising its corresponding Range Images which have led to the point cloud by registration of its different individual stations. We have first investigated the wavelet approach. We have then showed that the NL-means approach is much simpler to implement and more efficient as well (Smigiel et al, 2011). In this paragraph, we show very initial results of denoising based on dictionary learning of sparse coding. We first give a quick explanation of the method and show the initial results that we have obtained and compare them with the results we had obtained with the NL-means approach.

#### 3.1 Theory of dictionary learning of sparse coding

If we turn back to paragraph 2, we can say that dictionary learning consists mainly in finding out the sparsifying basis. In the literature, the approach is based on sparse and redundant representations over a trained dictionary. As the dictionary training algorithm is limited in handling small image patches, one extends its deployment to arbitrary image sizes by defining a global image prior that forces sparsity over patches in every location in the image. It has been shown how such Bayesian

treatment leads to a simple and effective denoising algorithm, with state-of-the-art performance, equivalent and sometimes surpassing recently published leading alternative denoising methods. For a complete survey of this approach, the reader may look at the rich literature on the subject (Duarte-Carvajalino et al., 2009; Elad et al., 2006; Kreuz-Delgado et al., 2003; Peyré, 2011).

#### 3.2 Results

In order to obtain first results of Range Image denoising, we have used a set of experimental data which have been denoised before with other approaches. In the case presented in the following lines, the data set consists in registration spheres shown on figure 4. Registration calibrated spheres were regularly distributed in the environment of an object of interest but the interesting objects in this context are the spheres themselves. They may be subject to a quantitative control which consists of modeling a mathematical sphere on the corresponding part of the point cloud then determining the standard deviation of the cloud from the spherical geometry. The best results had been obtained by NL-means denoising. Table 1 shows the obtained results. The standard deviation of the point cloud around the theoretical sphere is substantially decreased for both methods, a little more for the dictionary learning of sparse coding approach. However, a big difference exists in terms of computation time. If the NL-means approach is quick, the dictionary learning takes long, almost ten times the computation time for the NL-means approach.

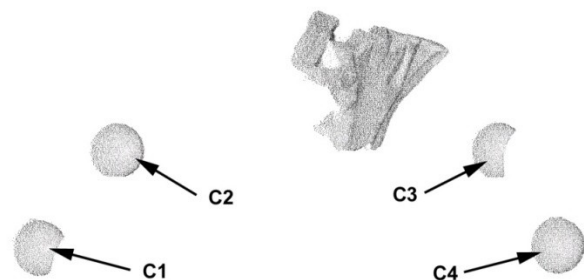


Fig. 4: the four registration spheres which have been the object of denoising through the NL-means approach and the dictionary learning of sparse coding approach.

#### 4. CONCLUSION AND FURTHER WORKS

In this paper, we have shown how modern ideas in the signal and image processing community could help developments in the photogrammetry and lasergrammetry community.

Existing methods are already capable of being applied: in particular, sparse coding may help to denoise Range Images obtained through Terrestrial Laser Scanners. Hence, one may build denoised point clouds that enhance the built-in accuracy and precision of these machines. The further works to be carried out in that direction would be to process a complete object to compare the final accuracy and precision to former approaches. The question of time computation is crucial since the herein described method has taken about ten times longer than a NL-means algorithm. Hence, the question could be asked as: is the algorithm complexity worth the computational time as regard to the signal-to-noise improvement?

The other direction is still more theoretical: it deals with the possibility of applying compressive sensing to Terrestrial Laser Scanners. A priori, a TLS is a raster single-pixel camera. Could

it be turned as a compressive sensing single-pixel camera? In other terms, the question would be to find out how to perform the  $M$  linear measurements. Because of the nature of the value to measure (either a time-of-flight or a phase shift), the answer is not obvious. However, if there should be a solution, it could be a major evolution of that kind of equipment.

	Sphere	Standard deviation (mm)
Raw data		
	C1	1.91
	C2	2.04
	C3	1.86
	C4	1.86
NL-means denoised		
	C1	0.83
	C2	0.67
	C3	1.07
	C4	0.72
Dictionary learning denoised		
	C1	0.78
	C2	0.63
	C3	1.10
	C4	0.69

Table 1: results of denoising: raw data, NL-means denoised spheres and Dictionary-learning denoised spheres.

## 5.

### REFERENCES

#### References from Journals:

- Candès, E. J., & Romberg, J. K. (2005). Signal recovery from random projections, 76-86. doi:10.1117/12.600722
- Candès, E. J., Romberg, J. K., & Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8), 1207–1223. doi:10.1002/cpa.20124
- Candès, E., & Romberg, J. (2007). Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3), 969. doi:10.1088/0266-5611/23/3/008
- Candès, E. J., & Wakin, M. B. (2008). An Introduction To Compressive Sampling. *IEEE Signal Processing Magazine*, 25(2), 21-30. doi:10.1109/MSP.2007.914731
- Chan, W. L., Charan, K., Takhar, D., Kelly, K. F., Baraniuk, R. G., & Mittleman, D. M. (2008). A single-pixel terahertz imaging system based on compressed sensing. *Applied Physics Letters*, 93(12), 121105-121105-3. doi:10.1063/1.2989126
- Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52(4), 1289-1306. doi:10.1109/TIT.2006.871582
- Duarte, M. F., Davenport, M. A., Takhar, D., Laska, J. N., Sun, T., Kelly, K. F., & Baraniuk, R. G. (2008). Single-Pixel Imaging via Compressive Sampling. *IEEE Signal Processing Magazine*, 25(2), 83-91. doi:10.1109/MSP.2007.914730
- Duarte-Carvajalino, J. M., & Sapiro, G. (2009). Learning to Sense Sparse Signals: Simultaneous Sensing Matrix and Sparsifying Dictionary Optimization. *IEEE Transactions on Image Processing*, 18(7), 1395-1408. doi:10.1109/TIP.2009.2022459
- Elad, M., & Aharon, M. (2006). Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries. *IEEE Transactions on Image Processing*, 15(12), 3736-3745. doi:10.1109/TIP.2006.881969
- Kreutz-Delgado, K., Murray, J. F., Rao, B. D., Engan, K., Lee, T.-W., & Sejnowski, T. J. (2003). Dictionary Learning Algorithms for Sparse Representation. *Neural Computation*, 15(2), 349-396. doi:10.1162/089976603762552951
- Lustig, M., Donoho, D. L., Santos, J. M., & Pauly, J. M. (2008). Compressed Sensing MRI. *IEEE Signal Processing Magazine*, 25(2), 72-82. doi:10.1109/MSP.2007.914728
- Ma, J. (2009). Single-Pixel Remote Sensing. *IEEE Geoscience and Remote Sensing Letters*, 6(2), 199-203. doi:10.1109/LGRS.2008.2010959
- Peyré, G. (2011). A Review of Adaptive Image Representations. *IEEE Journal of Selected Topics in Signal Processing*, 5(5), 896-911. doi:10.1109/JSTSP.2011.2120592
- Smigiel, E., Alby, E., & Grussenmeyer, P. (2011). TLS data denoising by range image processing. *The Photogrammetric Record*, 26(134), 171–189. doi:10.1111/j.1477-9730.2011.00631.x