

A COMPRESSION METHOD BASED on COMPRESSIVE SENSING for 3-D LASER RANGE SCANS of INDOOR ENVIRONMENTS

Oguzcan Dobrucali and Billur Barshan

Department of Electrical and Electronics Engineering, Bilkent University
06800, Bilkent, Ankara, Turkey
{dobrucali, billur}@ee.bilkent.edu.tr

Abstract. When 3-D models of environments need to be transmitted or stored, they should be compressed efficiently to increase the capacity of the communication channel or the storage medium. We propose a novel compression technique based on compressive sensing, applied to sparse representations of 3-D range measurements. We develop a novel algorithm to generate sparse innovations between consecutive range measurements along the axis of the sensor's motion, since the range measurements do not have highly sparse representations in common domains. Compared with the performances of widely used compression techniques, the proposed method offers the smallest compression ratio and provides a reasonable balance between reconstruction error and processing time.

1 Introduction

Many techniques have been developed for extracting 3-D models of environments, which allow describing objects with undefined or arbitrary shapes or patterns [1]. One approach to constructing 3-D models is to use laser range finders that measure the distance between the sensor and the objects within the field of view. The model is acquired by using either a conventional 3-D laser scanner, which is an expensive device, or a number of translating and/or rotating 2-D laser scanners [2].

In this study, we consider an indoor environment scanned in 3-D with a single 2-D laser range finder, rotating around a horizontal axis above ground level. The device used in this study is SICK LMS200 with its maximum range 80 m, field of view 180° , range resolution 1 mm, and angular resolution 0.5° [3]. Since the 3-D model is composed of a considerable number of 2-D scans that are themselves comprised of a large number of range measurements, the measurements need to be compressed when they are transmitted or stored.

The *compression ratio* (CR), which is the ratio of the size of the compressed output to the size of the original data, and the *speed of compression* are two important criteria for measuring compression performance. The CR is between zero and one (or zero and 100%), such that the closer the CR is to zero, the

greater the amount of compression. In terms of the CR, a compression method can be considered efficient when the size of the original data is reduced by more than one half, so that the capacity of the communication channel or data storage medium is at least doubled [4].

2 Review of Compressive Sensing

Compressive sensing is a technique that samples a signal in \Re^N , where N is very large, at a rate lower than the signal's Nyquist rate, using a linear sampling model with an optimization procedure for reconstructing the sampled signal [5].

The sampling model is composed of the *sparsifying basis* and the *measurement model* that satisfy *sparsity* and *incoherence* properties, respectively. Sparsity requires the signals to have sparse projections onto the sparsifying basis in which only a small number of the coefficients (K) will have large values, whereas the majority ($N - K$) will be close to zero. The sparsifying basis is an orthonormal basis denoted by $\Psi = [\psi_1, \dots, \psi_N]$ which is spanned by $\{\psi_i\}_{i=1}^N$. Thus, the sampled signal can be represented as $\mathbf{x} = \sum_{i=1}^N s_i \psi_i = \Psi \mathbf{s}$, where $\mathbf{s} = [s_1, \dots, s_N]^T$ in which $s_i = \langle \mathbf{x}, \psi_i \rangle$. Notice that, \mathbf{x} and \mathbf{s} are different representations of the same signal in time and Ψ domains, respectively. The measurement model determines M measurements, where $M \ll N$, using a linear operator $\Phi = [\phi_1^T, \dots, \phi_M^T]^T$ composed of $\{\phi_i\}_{i=1}^M$, each of which is in \Re^N . The measurement model should be chosen so that $\{\phi_i\}_{i=1}^M$ cannot sparsely represent $\{\psi_i\}_{i=1}^N$, which is a requirement of the incoherence property. Baraniuk suggests in [6] that the measurement model, in which each entry is chosen from a Gaussian distribution with zero mean and $\frac{1}{N}$ variance, is incoherent with any sparsifying basis with high probability. Given N and K , the lower bound on M is determined by:

$$M \geq cK \ln \left(\frac{N}{K} \right) \quad (1)$$

where c is a small positive constant [6]. Eventually, the measurement vector denoted by $\mathbf{y} = [y_1, \dots, y_M]^T$, where $y_i = \langle \mathbf{x}, \phi_i \rangle$, is obtained such that $\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \mathbf{s} = \Theta \mathbf{s}$. The signal is then reconstructed by determining \mathbf{s} , given \mathbf{y} and Θ . Since Θ is an $M \times N$ matrix with $M \ll N$, there is no unique solution to $\mathbf{y} = \Theta \mathbf{s}$. Therefore, the optimal solution is found by [7]:

$$\hat{\mathbf{s}} = \arg \min \|\mathbf{s}\|_1 \text{ such that } \mathbf{y} = \Theta \mathbf{s}. \quad (2)$$

Finally, the original signal is approximated from $\hat{\mathbf{x}} = \Psi \hat{\mathbf{s}}$ with little distortion.

3 The Proposed Method

In compressive sensing, although determining the measurement model is straightforward, determining the sparsifying basis is not so simple. One of the main objectives of this study is to obtain sufficiently sparse representations of the 2-D scans that form the 3-D model.

The experimental data set [8] used in this study is composed of 29 3-D scans from different indoor environments, each of which is acquired by taking 2-D

scans as the sensor is rotated in 471 steps around a horizontal axis above ground level. As a consequence, every 3-D scan in the data set constitutes 471 2-D scans, sequentially acquired as vectors in \mathfrak{R}^{361} (i.e., $N = 361$).

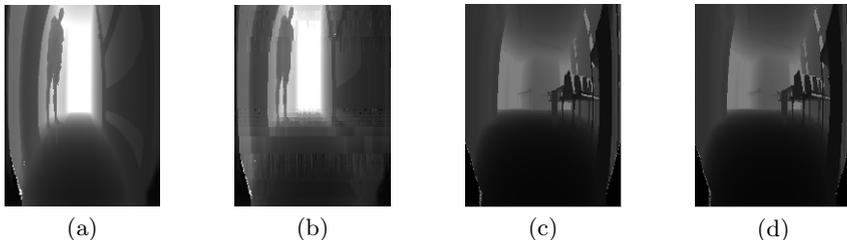


Fig. 1: (a) and (c): sample 3-D scans, (b) and (d): their reconstructions.

To apply the sampling model described in Sect. 2, we first consider the projections of a 3-D scan from the data set, illustrated in Fig. 1(a), onto some of the well-known sparsifying bases. The 2-D scans forming the 3-D scan are projected one at a time onto $N \times N$ sparsifying bases formed by using Fourier [9], Gabor [9], and Haar [10] dictionaries. After lowering the small values to zero, the average number of non-zero coefficients in these projections are around 270, 220, and 320, respectively. The projections are not sufficiently sparse, so both the CR and the distortion on the reconstruction would be high, if compressive sensing were used with one of these sparsifying bases and the measurement model mentioned in Sect. 2 [7]. Therefore, we propose a novel technique to generate more sparse innovations with approximately 40 non-zero coefficients on the average, for the same scan data.

The proposed method involves sparsifying, measurement, and reconstruction stages: The sparsifying model generates sparse innovations for each scan, and then the measurement model samples the innovations with the minimum number of samples. Finally, the reconstruction model rebuilds each scan from the samples encoded by the measurement model. In the following subsections, these three stages are described in more detail.

3.1 The Sparsifying Model

In the sparsifying model, we generate innovations between the currently acquired scan \mathbf{x} , and the previous scan $\tilde{\mathbf{x}}$. First, $\tilde{\mathbf{x}}$ is generated at the encoder by employing the reconstruction procedure that the decoder follows, to adapt the sparsifying parameters according to the reconstructions at the decoder. Then, $\tilde{\mathbf{x}}$ is approximated to \mathbf{x} by shifting $\tilde{\mathbf{x}}$ along the vertical and horizontal axes by amplitude (ϵ) and phase (δ) shifts, respectively.

We define an error function $\mathcal{E}^2 = \sum_{k=1}^N [x[k] - (\tilde{x}[k + \delta] + \epsilon)]^2$, where k is the discrete-time index, and set its partial derivatives with respect to ϵ and δ to zero, to find the optimal ϵ and δ . Ignoring the δ term in $\frac{\partial \mathcal{E}^2}{\partial \epsilon}$, we determine ϵ as:

$$\epsilon = \frac{1}{N} \sum_{k=1}^N (x[k] - \tilde{x}[k]) \quad (3)$$

which corresponds to the average amplitude difference between \mathbf{x} and $\tilde{\mathbf{x}}$. Since δ is assumed to be very small compared to N , $\tilde{x}[k + \delta]$ is expressed using the first two terms of its *Taylor series* expansion around k . Then, we find δ as:

$$\delta = \frac{\sum_{k=1}^N \tilde{x}'[k] (x[k] - \tilde{x}[k] - \epsilon)}{\sum_{k=1}^N \tilde{x}'[k]^2} \quad (4)$$

where $\tilde{x}'[k]$ is the first-order difference of the sequence $\tilde{\mathbf{x}}$ at k .

Shifting $\tilde{\mathbf{x}}$ along the horizontal and vertical axes by ϵ and δ , respectively, we obtain an approximation $\hat{\mathbf{x}}$ to \mathbf{x} . Then, the innovation is defined as $\mathbf{v} = \mathbf{x} - \hat{\mathbf{x}}$ and is vertically shifted in either positive or negative direction by the offset Δ , to bring its average value to the zero level. We eventually obtain a highly sparse innovation $\hat{\mathbf{v}}$ after nulling very small variations around zero. Consequently, \mathbf{x} is represented with ϵ , δ , Δ , and $\hat{\mathbf{v}}$, respectively. When the *mean square error* (MSE) between \mathbf{x} and $\hat{\mathbf{x}}$ is very low, $\hat{\mathbf{v}}$ becomes very small, so \mathbf{x} is represented without $\hat{\mathbf{v}}$. When the MSE is greater than a preset threshold, $\hat{\mathbf{v}}$ becomes not as sparse as we would like, so \mathbf{x} is not encoded.

3.2 The Measurement Model

The measurement model gets the minimum number of samples from $\hat{\mathbf{v}}$ by using either simple coding (SC) or compressive sensing (CS). Simple coding encodes $\hat{\mathbf{v}}$ using the location and amplitude of the non-zero components. The measurement size M is $2K$, and the reconstruction error is zero. Compressive sensing measures arbitrary linear combinations of the components in $\hat{\mathbf{v}}$. In this case, M is determined using (1), and the reconstruction error increases with K .

When M required by SC is lower than M required by CS, using SC is advantageous over using CS, in terms of the zero reconstruction error and lower M . Therefore, we obtain the measurements \mathbf{m} applying either SC when $K \leq K^*$, or CS, otherwise. Here, K^* is the value of K that makes M for SC equal to M for CS. We include a special character (i.e., π) at the beginning of \mathbf{m} when SC is applied to inform the decoder that we are using SC instead of CS. When $K > \frac{N}{2}$, $\hat{\mathbf{v}}$ cannot be considered sparse, since the reconstruction error would be very high if $\hat{\mathbf{v}}$ were sampled using CS. In that case, \mathbf{x} is not encoded.

At the output of the measurement model, \mathbf{x} is represented with $\{\epsilon, \delta, \Delta, \mathbf{m}\}$ if it is encoded. Otherwise, \mathbf{x} is left as is.

3.3 The Reconstruction Model

The reconstruction model rebuilds \mathbf{x} from the output generated by the encoder. When \mathbf{x} is encoded, the output is $\{\epsilon, \delta, \Delta, \mathbf{m}\}$ with length $(M + 3)$ less than N . Otherwise, the output is \mathbf{x} with length N . Therefore, the output is stored directly as the reconstruction of \mathbf{x} if its length is N . Otherwise, the reconstruction model is applied to the output.

The reconstruction model first decomposes the output into ϵ , δ , Δ , and \mathbf{m} . After this step, to obtain $\hat{\mathbf{x}}$, $\tilde{\mathbf{x}}$ is shifted along the vertical and horizontal axes by ϵ and δ , respectively. Afterwards, \mathbf{v} is rebuilt from \mathbf{m} and Δ . In this step, if the first value of \mathbf{m} is π , then $\hat{\mathbf{v}}$ is rebuilt, decoding the rest of \mathbf{m} with respect to the SC scheme, which involves filling an empty signal in \mathfrak{R}^N with the values of

location and amplitude pairs given in the measurements. Otherwise, $\hat{\mathbf{v}}$ is rebuilt, decoding \mathbf{m} with respect to the CS scheme, which involves solving (2), where $\Theta = \Phi$, by following the procedure in [7]. Then, \mathbf{v} is acquired by shifting the amplitude of $\hat{\mathbf{v}}$ by Δ . Eventually, \mathbf{x} is reconstructed by adding \mathbf{v} to $\hat{\mathbf{x}}$.

The reconstruction model is used at the decoder, as well as at the encoder, to estimate the reconstructions generated by the decoder.

4 Comparing the Proposed Method with Some Well-Known Compression Techniques

In this section, we compare the compression performance of the proposed method with some well-known and widely used lossless and lossy compression techniques that are applied to every 2-D scan independently in all of the 3-D scans in the data set. Thus, for each technique in the comparison, we compare the CR, the average of the *root mean square error* between the 2-D scans and their reconstructions (E), and the time required for encoding (t_{enc}) and decoding (t_{dec}). These values are found by averaging over the values obtained for the whole data set, including 4,930,899 ($= 29$ 3-D sets \times 471 2-D scans \times 361 measurements) range measurements in total.

We first compress the scan data using four of the lossless techniques: *Huffman* [11], *arithmetic* [11], *ZLIB* [12], and *GZIP* [13]. Besides, we also apply two of the lossy compression methods to the data set: *JPEG* [11] and 3-level *wavelet transform* using the *Haar dictionary* [14]. The results are given in [Table 1](#).

	method	CR (%)	E (cm)	t_{enc} (s)	t_{dec} (s)
lossless	Huffman coding	41.7	0	165.6	610.6
	arithmetic coding	11.1	0	37.6	48.9
	ZLIB	65.3	0	0.4	0.2
	GZIP	76.7	0	0.5	0.3
lossy	JPEG	25.4	184.2	0.8	0.2
	wavelet transform	12.7	37.3	0.1	0.1
	proposed	10.9	12.9	15.3	14.5

Table 1: Table of CR, E , t_{enc} , and t_{dec} for compression using different methods.

Finally, the data set is encoded using the proposed method. In this method, since the measurement model in CS is determined arbitrarily in each trial, small fluctuations in the compression performance are observed ($\pm 2\%$ in CR). Therefore, CR, E , t_{enc} , and t_{dec} are obtained as in [Table 1](#), after the whole data set is encoded 10 times, and the results are averaged. On the average, the data set is compressed by 89% with about 13 cm distortion in the reconstructions. In this case, 57% of the 2-D scans are encoded with ϵ , δ , and Δ ; 24% are encoded with ϵ , δ , Δ , and \mathbf{m} obtained using SC; 16% are encoded with ϵ , δ , Δ , and \mathbf{m} obtained using CS. Only 3% are not encoded. When the 3-D scans illustrated in [Fig. 1\(a\)](#) and (c) are compressed by 89% and 85%, the resulting average distortions are 12 and 16 cm, respectively. The reconstructed scans are shown in [Fig. 1\(b\)](#) and (d) to allow comparison with their originals.

The proposed method compresses the experimental data more than all the lossy and lossless techniques we have considered. In terms of speed, the proposed

method is much faster than Huffman and arithmetic coding, but much slower than ZLIB, GZIP, JPEG, and the wavelet transform. However, the proposed method compresses more than the latter four. Moreover, the proposed method provides much less reconstruction error than the latter two.

5 Conclusion

In this study, we consider 3-D modelling of indoor environments employing the SICK LMS200 laser range finder. The 2-D range scans forming the 3-D model are compressed so that they can be stored or transmitted efficiently. From this perspective, we propose a novel compression technique based on compressive sensing for sequentially acquired 2-D scans.

According to the criteria described in Sect. 1, the proposed method is fast and efficient in terms of CR, and provides a reasonably good balance between reconstruction accuracy and speed [11]. It is recommended for applications where both CR and speed are crucial. However, a lossless compression technique can be used in applications where the accuracy of the range measurements is more important. Our future work involves improving the compression performance of the proposed method, and extending its application to 3-D range measurements of outdoor environments.

References

1. C. Brenneke, O. Wulf, B. Wagner: Using 3-D laser range data for SLAM in outdoor environments. *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Syst.* (2003) 188–193
2. D. Borrmann, J. Elseberg, K. Lingemann, A. Nüchter, J. Hertzberg: Globally consistent 3-D mapping with scan matching. *Robot. Auton. Syst.* **56** (2007) 130–142
3. SICK AG: Quick Manual for LMS Communication Setup (March 2002) Ver. 1.1.
4. D. Salamon: *A Guide to Data Compression Methods*. Springer, New York, U.S.A. (2002)
5. E. J. Candes and M. B. Wakin: An introduction to compressive sampling. *IEEE Signal Proc. Mag.* **25** (2008) 21–30
6. R. G. Baraniuk: Compressive sensing. *IEEE Signal Proc. Mag.* **24** (2007) 118
7. E. Candes, J. Romberg: Signal recovery from random projections. *Proc. SPIE*. Vol. 5674. (2005)
8. Nüchter, A.: Osnabruck University and Jacobs University Knowledge-based Systems Research Group Repository (2009) <http://kos.informatik.uni-osnabrueck.de/3Dscans/>
9. S. S. Chen: Basis Pursuit. PhD thesis, Stanford University, Department of Statistics, California, U.S.A. (1995)
10. I. Daubechies: *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania (1992)
11. K. Sayood: *Introduction to Data Compression*. Academic Press, San Diego, U.S.A. (2000)
12. G. Roelofs, M. Adler: The ZLIB Homepage (August 2009) <http://www.zlib.net/>
13. J. Gailly, M. Adler: The GZIP Homepage (July 2003) <http://www.gzip.org/>
14. G. Strang and T. Nguyen: *Wavelets and Filterbanks*. Wellesley-Cambridge Press, Wellesley MA, U.S.A. (1997)