Compressive-Sensing Based Codec of the Y Color Component for Point Cloud

Weiwei Wang[1], Hui Yuan[2(✉)], Hao Liu[1], and Qi Liu[1]

[1] School of Information Science and Engineering, Shandong University, Qingdao, Shandong, China
wangweiwei@mail.sdu.edu.cn, liuhaoxb@gmail.com, sdqi.liu@gmail.com

[2] School of Control Science and Engineering, Shandong University, Jinan, Shandong, China
yuanhui0325@gmail.com

Abstract. The point cloud obtained by the 3D laser scanner contains a very large amount of data, in order to transmit the point cloud data as much as possible with the limited bandwidth, the effective compression of point cloud data has become a problem that needs to be solved urgently nowadays. In this paper, we use the compressive sensing theory to compress and reconstruct one of the point features, that is, the Y color component, served as the signal. We also use the K-SVD algorithm to explore the signal's sparsity according to its unique structural features, the K-SVD algorithm can learns a sparse basis matrix that is common to all point cloud models used in our experiments. For experimental results, we use rate-distortion metric. The results show that for each point cloud model, our method can achieve a higher probability to reconstruct the original data after compressed.

Keywords: Compressive sensing · K-SVD algorithm · Pointcloud

1 Introduction

In the field of computer vision, it has become more important to represent data in 3D. In recent years, point clouds have become popular to represent 3D data, as the scanners that can capture 3D data increasingly ubiquitous, 3D point clouds have widely been used in different kinds of fields in modern society, like robotics [1], autonomous driving [2], virtual/augmented reality [3], vehicular networking technology [4], internet of things [5] etc. A point cloud is a collection of points that can describe the surface of an object which can obtained by the 3D scanners, and can represented as a set of 3D points $\{p_i|i = 1, ...n\}$, where each point contains a position vector (x, y, z), and its features such as color (R, G, B), surface normal, etc. However, with the development of the scanners, point cloud can be created at very high rates which allow for efficient and compact storage as well as transfer of this data, compression of point cloud has therefore

B. Li et al. (Eds.): IoTaaS 2019, LNICST 316, pp. 249–258, 2020.
https://doi.org/10.1007/978-3-030-44751-9_22

been Chinese and foreign academia hot research field, which is the key driving force in the fields of immersive communication and automatic driving.

The compression of point cloud can be started from two aspects: geometry and texture. At first, the compression of geometry information gained the most concerned. Among them, the octree decomposition method [6] has been used extensively because of its efficiency and low-complexity. Given a point cloud P, we can get a cube based on its geometric information that is able to surround the entire point cloud model, then, an octree O is constructed with a maximum number of level L and the points in P are sorted into the cells of the octree. We can divide the cube by the mean value from the three directions of x, y, z, each subdivision produces eight child cells. If the current cell is occupied or does not meet the maximum number of level that we already set, it continues to be divided. At the process of encode, the existent child cells are specified in a single byte per cell subdivision, i.e. each bit specifies the occupancy of a child cell. This is the most basic principle for encoding geometric information of the point cloud with an octree.

In recent years, with the development of 3D video panorama and other technologies, for better visual effects, and recently the color feature of the point cloud received an increasing amount of attention in particular. However, for data represented in point cloud form, which is irregular, the compression of the color also faces enormous challenges. Instead of compressing the irregular data directly, [7–9] map the irregular data into regular data for convenient data processing. Mekuria *et al.* [7] traversed each point's color with a depth first order from the octree and then used the zig-zag scan to map them to a 8×8 blocks of a 2D grid. After that, they compressed the grids with JPEG by using the correlations between the colors. Almost the same idea with Mekuria, Tu *et al.* [8] map the point cloud into range images and then compressed them with either JPEG or MPEG-4. Cui *et al.* [9] compressed the 2D grids data by selecting two redefined models. And in this paper, we compressed one of the color attributes by exploiting compressive sensing theory.

In 2006, the theory of compressive sensing (CS) proposed by Candès and Donoho [10–13] pointed out that for signals that are sparse themselves or sparse under a certain transform basis, they can be observed by non-linear downsampling. And then the low-dimensional observations can be used by the measurement matrix which satisfy the Restricted Isometry Property (RIP) with transform basis to perform a high probability reconstruction of the original signal. Different from the traditional Nyquist sampling theorem, the compressive sensing theory combines the sparse characteristics of the signal, and uses the measurement matrix to observe the signal, so that the sampling process of the signal does not depend on the bandwidth of the signal, but the content and structure of the signal. Therefore, the theory of compressive sensing opens a new path for the compressing and coding theories of signals. In the past decade, the CS algorithm has made great progress, especially the development of its reconstruction algorithm [14–16]. Wang *et al.* [17] applied this theory to the deep network of image reconstruction, which not only achieves good reconstruction effect, but also reduces the computational complexity. In order to avoid unauthorized access to multimedia content, Athira *et al.* [18] even added this theory to Encryption technology. In the automobile sensor system, radar sensors are often used to image in order to provide better visual aids for drivers. A new radar signal processing technology based on compressive sensing theory is proposed by Baselice *et al.* [19] which can image two or more targets in the same line of sight. The performance of radar DAS (DAS: Driver Assistance Systems) is greatly improved. It can be seen that the theory is widely used in various fields.

In this paper, we use compressive sensing theory to compress and reconstruct one of each point's color component, that is, Y color component. In the early data preprocessing stage, we will use the geometric information to spatially decompose the point cloud with octree firstly, after that we can obtain the Y color component values for each point in the same cell. The Y color component values are then applied as a signal in our compressive sensing theory. Simultaneously, we use the K-SVD algorithm to learn a sparse basis that can be universally used by these five models: Longdress_vox10_1300, Loot_vox10_1200, Queen_0200, Redandblack_vox10_1550 and Soldier_vox10_0690, these five models we used in the experiments are all from the new test models presented at the MPEG 125th conference.

The following contents are organized as follows: in the second part we will introduce how to use the geometric spatial characteristics and the local similarity characteristics of the point cloud to obtain the training data; the third part, as the acquired training data has some similarities, we propose to use the K-SVD algorithm to train a dictionary to obtain the sparse basis of signals; in the fourth part, we will give the experimental results and have a brief explanation; in the fifth part, we concludes the paper and provides pointers to future directions.

2 Acquisition of Training Data

Generally speaking, the points collected by the 3D laser scan are out of order, if we want to use the compressive sensing theory to have a good compression and reconstruction effect on the point cloud data, the higher the similarity of the point's Y color component value, the better, this will provide important guarantees for our subsequent dictionary learning method to obtain a good sparse basis, so we need to do some preprocessing on the point cloud data. In this paper, we use a signal matrix to train the sparse basis matrix $\mathbf{D}$ with each column of the signal matrix to be one signal. In the experiment, we take the Y color component of each point in the same cell as a signal. To ensure that the dimensions of each signal can be the same, we first decompose each point cloud model with an octree. The octree decomposition process only uses the geometric information of the point cloud model: given a point cloud model, a cuboid can be constructed according to the maximum and minimum values of its coordinate information, which can then be divided from the mean of the three coordinate axes, with each cell be divided into eight childcells. We finally want to extract the Y color component of the points in the same cell as our signal, and also try to ensure that the number of points in each cell can be the same, that is, the dimensions of our signals are the same. In experimenting we set the dimension of a signal to 512. For each cell subdivision, the number of points in each cell can be known and only cells with more than 512 points are further subdivided. However, this does not absolutely guarantee that the number of points contained in each cell is the same as 512.

Through the above octree decomposition, we can know which cell each point is located. After simple processing, we can get an index matrix. The value in each column of this matrix represents the index value of the point in the same cell. The number of columns of the index matrix is the number of cells in which the point cloud is divided by the octree. (Index value: each point cloud model is a $\mathrm{N} \times 6$ matrix $\mathbf{I} = \{i_0, ..., i_{N-1}\}$ after reading with Matlab, regardless of the normal vector, N represents the total number of points in the model, with $i_n\,(n = 0, ..., N-1)$ represents the geometric and color information of a point. n is what we said the index value). After the index matrix is obtained, to make a better use of the similarity between signals, we slice the points in the same cell, and perform a raster scan for each point on the slice. The slice operation changes only the order of the index value in the index matrix **I**. Then we extract the Y color component of each point according to the index matrix and subsequently get the signal matrix **S**. What needs to be mentioned is that each value in the signal matrix is de-averaged. For signals with less than 512 values, zero-padding operations is performed in the vacant place. The disadvantages of this is that it is equivalent to an increase in the amount of data, and the coding efficiency will be very low.

In our experiment, the signal matrix is obtained from five point cloud models by the method of obtaining training data proposed above. The signal matrix is then used as our training data in the K-SVD algorithm. Among the five models, the training data of the Longdress_vox10_1300 model has a dimension of 512×4777, the Loot_vox10_1200 model is 512×3824, the Queen_0200 model is 512×6131, the Redandblack_vox10_1550 model is 512×3784, and the Soldier_vox10_0690 model is 512×7385, so the dimension of the large training data **S** composed of these five models is 512×25901. The purpose of using the five point cloud models to form a large training data is to learn a dictionary **D** that is common to the five point cloud models. Under the dictionary **D**, each point cloud model can have a good sparse representation.

3 Over-Complete Dictionary Training of Point Cloud Based on K-SVD Algorithm

In the existing sparse representation theory, there are usually two methods used to create a sparse representation dictionary: (1) Based on mathematical models. E.g. Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT) [20], Wavelet Transform [21,22] etc. (2) Training dictionary for training sets with distinct characteristics.

For signals that do not have the properties of digital image signals, they do not necessarily have good sparsity in the DCT transform domain, so if the

DCT domain transform is still performed, the sparse coefficients obtained cannot guarantee the sparsity. In the preliminary stage of the experiment, we used the DCT transform basis for the signal, but the effect was not satisfactory. In order to get the sparsity of the more general signals, we later modified it to obtain the sparse basis matrix of the signal through the dictionary learning method. The dictionary learning learns a matrix, which is equivalent to the transformation matrix in the DCT transform, and is often called over-complete dictionary, its number of rows is much smaller than the number of columns, which is generally represented by $\mathbf{D}$. In the signal encoding and decoding algorithm based on compressive sensing, constructing a good dictionary is a very important part. The quality of the dictionary directly affects the quality of the final reconstruction effect.

The commonly used dictionary learning algorithms are K-SVD [23] and K-Means [24]. The K-SVD algorithm obtains the over-complete dictionary that is most suitable for the training set through continuous training update. Since it is adaptively obtained through training update, the signals can be decomposed according to its excellent structural features in the over-complete dictionary for better exploration of the signal's sparsity.

3.1 K-SVD Algorithm

In the K-SVD algorithm, $\mathbf{Y} = \mathbf{DX}$, with $\mathbf{Y}$ is the sample signal matrix, $\mathbf{D}$ is the dictionary, $\mathbf{X}$ is the sparse coefficient matrix, the goal of the algorithm is to train a dictionary $\mathbf{D}$ so that the product of the dictionary $\mathbf{D}$ and the sparse coefficient matrix $\mathbf{X}$ can be as close as possible to the matrix $\mathbf{Y}$, while each column of the sparse coefficient matrix $\mathbf{X}$ is as sparse as possible. In our experiment, the sample signal matrix $\mathbf{Y}$ is our training set $\mathbf{S}$ which we obtained above. During training, according to the suggestions in [23], for atoms that are used less frequently and have similarities, we replaced them with normalized vectors in the sample signal matrix that maximizes the error. In this experiment, the training set $\mathbf{S}$ is composed of five point cloud models, Longdress_vox10_1300, Loot_vox10_1200, Queen_0200, Redandblack_vox10_1550, Soldier_vox10_0690, and the learned dictionary can be universally used by this five models, then, each model is compressed through the compressive sensing methods by using the common dictionary $\mathbf{D}$ as there sparse basis.

3.2 Compressive Sensing Theory

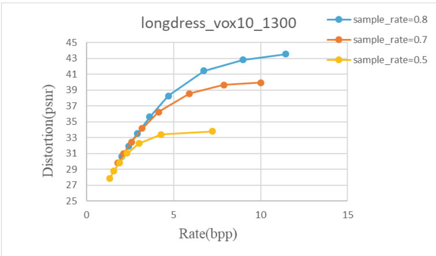
As we stated earlier, the theory of compressive sensing (CS) proposed by Candès and Donoho [10–13] pointed out that for signals that are sparse themselves or sparse under a certain transform basis, almost exact reconstruction of the signal can be achieved by unknown observations. The focuses of this theory are measurement matrix, sparse basis matrix and the reconstruction algorithm. The measurement matrix acts as a down-sampling, reducing the original high-dimensional data to a low-dimensional, reducing the amount of data, which in turn can be encoded with fewer bits-streams. There are two requirements for the measure matrix: it can play the role of down-sampling; and together with the sparse basis matrix satisfies the Restricted Isometry Property (RIP). Therefore, the dimension of the measurement matrix must ensure that the number of rows is less than the number of columns. In our experiments, the measurement matrix $\Phi \in R^{m \times n}$, m = a × n, obviously, a is the sample rate. We set the values of the two sample rates to 0.7, 0.8 and 0.5 respectively for comparison experiments, n is 512. The measurement matrix can be divided into deterministic measurement matrix and random measurement matrix. The common deterministic measurement matrix has partial orthogonal matrix, polynomial deterministic matrix, Toeplitz matrix, partial Fourier matrix, etc. and the commonly used random measurement matrix is Gaussian random measure matrix. In this experiment, a more common Gaussian random measurement matrix is adopted. Since the generation of the matrix is random, in order to ensure that the same measurement matrix can be used under different quantization steps, we first generate a m ∗ n-dimensional Gaussian random measurement matrix. The measurement matrix is saved and then loaded directly into each run.

The discrete point cloud data are almost non-sparse in numerical representation, so it is impossible to directly observe and reconstruct point cloud data. Therefore, we need to find the sparse representation of point cloud data under a certain transformation basis. The so-called sparse representation of a signal means that the signal can be linearly represented by a small number of basis vectors in its space, and the basic idea of sparse representation theory based on an over-complete dictionary can be considered in a condition of reconstructing the original signal as much as possible to replace the traditional orthogonal basis by an over-complete basis. The number of rows of the over-complete dictionary is much smaller than the number of columns. The K-SVD algorithm gradually trains the redundant dictionary that is most suitable for the training set through iterative operations. That is, we can get the sparse basis of point cloud data through the K-SVD algorithm described above.

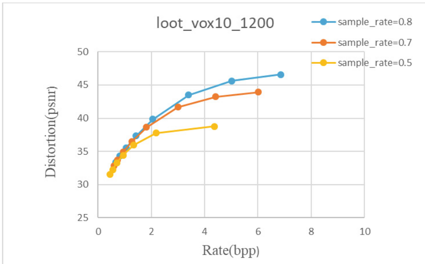
After compressing and encoding the point cloud data by the compressive sensing method in the encoder, we need to use the reconstruction algorithm at the decoder to recover the original data as much as possible. In recent years, many reconstruction algorithms have appeared in the field of compressive sensing. Among many reconstruction algorithms, the orthogonal matching pursuit (OMP) algorithm is widely used in the field of sparse representation because of its simple and easy to use. The input of this algorithm is a one-dimensional signal, that is to say, the one-dimensional signal OMP algorithm can perform sparse representation better. In this experiment, although the input is the sample signal matrix, each column of the matrix is a signal, so we adopt this reconstruction method.

4 Experiment Result

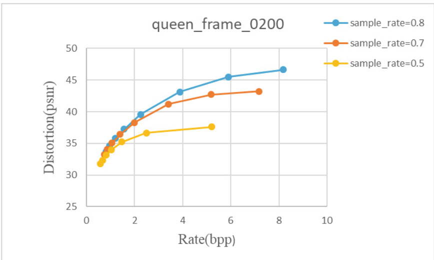
A brief review of the entire experiment process: Five point cloud models, after acquiring the training data, combine the training data of the five models to obtain a large training data, and then use the K-SVD algorithm to train a common sparse basis for the five models and finally applied to each point cloud model by compressive sensing. The models used in the experiment is the new test models given by MPEG 125th conference: Longdress_vox10_1300 with 857966 points, Loot_vox10_1200 with 805285 points, Queen_0200 with 1000993 points, Redandblack_vox10_1550 with 757691 points, Soldier_vox10_0690 with 1089091 points. In the case of compressive sensing, different sample rates and quantization steps were applied to compare experiments. The specific experimental data and RD curves are as Fig. 1:



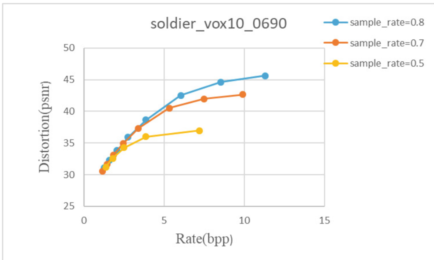
(a) Longdress



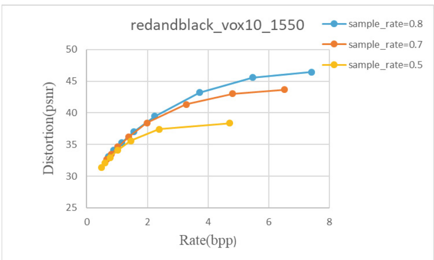
(b) Loot



(c) Queen



(d) Soldier



(e) Rendandblack

Fig. 1. Rate and distortion curves of these five models after encoding and decoding of the compressive sensing

From the experimental results, we can see that in the case of high code rate, the reconstruction effect is better. The higher the sample rate, the better the reconstruction effect.

5 Conclusion and Prospect

The rapid development of 3D scanning technology makes it possible to efficiently obtain massive point cloud data to represent real-world objects, but usually the amount of point cloud data is very large, so compressed encoding and reconstruction of 3D point cloud has become one of the research hotspots. How to perform high-efficiency compressed encoding and reconstruction without degrading the quality of the point cloud model has been a problem that the field is trying to solve. The compressive sensing theory proposed by Candès *et al.* in 2006 provides a new way for compression algorithms.

After studying the theory of compressive sensing and sparse representation theory, this paper analyzed and processed the scattered and disordered point cloud data in space, studied its sparse representation method, and used the octree to fully exploit the information inside its data. The models were spatially decomposed and finally obtained point cloud data suitable for the theory of compressive sensing. The work done in this paper can be summarized as follows:

1. Using the octree to achieve the spatial decomposition of the point cloud, and slicing the points in each cell, which improves the numerical similarity of the Y-color components of the 3D point cloud data and provides an important guarantee for sparse representation of three-dimensional point clouds.
2. Using the five models of Longdress_vox10_1300, Loot_vox10_1200, Queen_0200, Redandblack_vox10_1550, Soldier_vox10_0690 in the test models given by the MPEG 125th conference, an over-complete dictionary of these five models is trained using the K-SVD algorithm, and by using the dictionary in compressive sensing, both five point cloud models can achieve higher probability reconstruction.

The next research work mainly includes:

1. Research how to divide the point cloud surface into k-nearest neighbors. In this paper, we use the octree method to decompose the point cloud, and can't guarantee the points in each cell has the same number, and we increase the amount of data artificially for the cells with insufficient points.
2. For the K-SVD algorithm to train the over-complete dictionary part, research whether an adaptive optimization algorithm can be used to judge if the iteration has reached convergence. In this paper, we just set the number of iterations to a fixed value and did not judge whether convergence has been reached.

Acknowledgement. This work was supported in part by the National Natural Science Foundation of China under Grants 61871342; in part by the National Key R&D Program of China under Grants 2018YFC0831003; in part by the Shandong Natural Science Funds for Distinguished Young Scholar under Grant JQ201614; in part by the Shandong Provincial Key Research and Development Plan under Grant 2017CXGC1504; in part by the open project program of state key laboratory of virtual reality technology and systems, Beihang University, under Grant VRLAB2019B03; and in part by the Young Scholars Program of Shandong University (YSPSDU) under Grant 2015WLJH39.

References

1. Garrote, L., Rosa, J., Paulo, J., et al.: 3D point cloud downsampling for 2D indoor scene modelling in mobile robotics. In: IEEE International Conference on Autonomous Robot Systems & Competitions. IEEE (2017)
2. Mousavian, A., Anguelov, D., Flynn, J., et al.: 3D bounding box estimation using deep learning and geometry (2016)
3. Mekuria, R., Blom, K., Cesar, P.: Design, implementation and evaluation of a point cloud codec for tele-immersive video. IEEE Trans. Circuits Syst. Video Technol. **27**, 828–842 (2016)
4. Du, Y., ShangGuan, W., Chai, L.: Particle filter based object tracking of 3D sparse point clouds for autopilot. In: 2018 Chinese Automation Congress (CAC), Xi'an, China, pp. 1102–1107 (2018)
5. Wang, H., et al.: The research of chemical plant monitoring base on the internet of things and 3D visualization technology. In: 2013 IEEE International Conference on Information and Automation (ICIA), Yinchuan, pp. 860–864 (2013)
6. Schnabel, R., Klein, R.: Octree-based point-cloud compression. In: Eurographics. Eurographics Association (2006)
7. Mekuria, R., Blom, K., Cesar, P.: Design, implementation, and evaluation of a point cloud codec for tele-immersive video. IEEE Trans. Circuits Syst. Video Technol. **27**(4), 828–842 (2017)
8. Tu, C., Takeuchi, E., Miyajima, C., Takeda, K.: Compressing continuous point cloud data using image compression methods. In: Proceedings of the IEEE Intelligent Transportation Systems (ITSC), Rio de Janeiro, pp. 1712–1719 (2016)
9. Cui, L., Xu, H., Jang, E.S.: Hybrid color attribute compression for point cloud data. In: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, pp. 1273–1278 (2017)
10. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory **52**(2), 489–509 (2006)
11. Donoho, D.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
12. Candès, E.: Compressive sampling. In: International Congress of Mathematics, Madrid, Spain, pp. 1433–1452 (2006)
13. Baraniuk, R.: Compressive sensing. IEEE Signal Process. Mag. **24**(4), 118–121 (2007)
14. Metzler, C.A., Maleki, A., Baraniuk, R.G.: From denoising to compressed sensing. IEEE Trans. Inf. Theory **62**(9), 5117–5144 (2014)
15. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. Image Process. **15**(12), 3736–3745 (2006)
16. Ma, L., Bai, H., Zhang, M., Zhao, Y.: Edge-based adaptive sampling for image block compressive sensing. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **99–A**(11), 2095–2098 (2016)

17. Wang, Y., Bai, H., Zhao, Y.: Image reconstruction from patch compressive sensing measurements. In: 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM). IEEE Computer Society (2018)
18. Athira, V., George, S.N., Deepthi, P.P.: A novel encryption method based on compressive sensing. In: International Multi-conference on Automation. IEEE (2013)
19. Baselice, F., Ferraioli, G., Matuozzo, G., et al.: Compressive sensing for in depth focusing in 3D automotive imaging radar. In: 2015 3rd International Workshop on Compressed Sensing Theory and Its Applications to Radar, Sonar and Remote Sensing (CoSeRa). IEEE (2015)
20. Bai, H., Wang, A., Zhang, M.: Compressive sensing for DCT image. In: International Conference on Computational Aspects of Social Networks. IEEE (2010)
21. Wang, D.W.D., Zhang, L.Z.L., Vincent, A., et al.: Curved wavelet transform for image coding. IEEE Trans. Image Process. **15**(8), 2413–2421 (2006)
22. Cheng, G.Q., Cheng, L.Z.: A new image compression via adaptive wavelet transform. In: International Conference on Wavelet Analysis & Pattern Recognition. IEEE (2007)
23. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. **54**(11), 4311–4322 (2006)
24. Sahoo, S.K., Makur, A.: Dictionary training for sparse representation as generalization of K-means clustering. IEEE Signal Process. Lett. **20**(6), 587–590 (2013)