Depth Image Inpainting method based on sparse gradient prior

J Wu¹, and Z Li¹, ²

¹ School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan, Hubei 430081, China
² School of Physics and Electronic Engineering, Xinxiang College, Xinxiang, Henan 453000, China

Abstract: This paper solves the inpainting problem of single depth images. depth images are regarded as natural images without texture. Because of the sparsity property of natural images and the textureless property of depth images, we propose a similar group-based sparse model with sparse gradient regularization. For one thing, the similar group-based sparse model can better represent the local smooth and nonlocal self-similarity. For another, the sparse gradient regularization can better represent the textureless properties. The proposed algorithm takes advantage of the properties of depth images. The experimental results show the effect of the proposed algorithm.

1. Introduction
Depth images are a key research topic in the field of 2D-to-3D technology. However, the acquired depth images are prone to loss of information and require repairing.

At present, image inpainting technologies [1-3] are mostly focused on natural and medical images, such as inpainting of old photos and restoration of CT images. Most image inpainting methods can be applied to depth images directly, and relatively few scholars have paid attention to depth images inpainting.

Depth images can be viewed as natural images without texture. Due to the limitations of hardware and scenes, there exist black holes in depth images. With the extensive application of depth images, certain scholars have begun to focus on depth image inpainting. A weighted analysis representation model was proposed that advanced the conventional method in two aspects, namely, task-driven learning and dynamic guidance [3]. In [4], a modification to the morphological closing by a reconstruction algorithm was proposed without correlative color image information, which improved the depth image quality. In [5], Xue et al. proposed the integration of low gradient regularization with the low rank regularization method for depth image inpainting, the proposed method was effective.

The mathematical expression of the depth image inpainting problem is:

$$x = \arg \min_x \frac{1}{2} \| Hx - y \|^2_2 + \lambda \cdot \psi(x) \quad (1)$$

where $x$ and $y$ are representations of the original and degraded depth image, respectively; $H$ is a binary template; $n$ usually is additive Gaussian white noise; $\| Hx - y \|^2_2$ is the data fidelity term; $\psi(x)$ is the regularization term and $\lambda$ is the weight parameter, We aim to obtain the original depth image $x$.
Because the prior of depth images plays an important role in inpainting, we should take advantage of it. Depth images are approximated as natural images with textureless [5], so we apply a similar group-based sparse model assumption to obtain the rough inpainting result. Each similar group of depth images can be accurately represented based on an effective self-adaptive dictionary learning method. Due to the textureless property, we choose $L_0$ gradient regularization as the sparse gradient constrain, which is the most suitable measure for textureless[6-8]. In summary, we propose a similar group-based sparse model with $L_0$ gradient regularization for the inpainting of single depth images.

2. Similar Group-Based Sparse Model

In [9], the basic unit of sparse representation is a similar group. The construction of a similar group is shown below.

As shown in Fig. 1, image $x$ is first divided into overlapped pixel blocks of size $\sqrt{s} \times \sqrt{s}$, and each pixel block is in vector form $x_k$, where $k=1, 2, 3...n$. Then, in the red training window, for each patch $x_k$ denoted by a blue mark, its $c$ similar patches are determined, which comprise set $S_{x_k}$. Finally, all patches in set $S_{x_k}$ are arranged in a matrix to obtain similar groups $x_{c_k} \in R^{n \times c}$. Therefore, according to equation (1), depth image inpainting can be formulated as follows:

$$
\hat{x} = \arg\min_{x} \frac{1}{2} \|HD_G \circ a_G - y\|_2^2 + \lambda_1 \|a_G\|_p + \lambda_2 \|\nabla \hat{x}\|_0
$$

where $D_G$ is the dictionary, $a_G$ is the sparse vector, $\lambda$ is the weight parameter and $\|a_G\|_p$ represents the norm. In this paper, $p$ is 0.

The similar group-based sparse model can be directly applied to depth image inpainting. Because of the textureless property, it is reasonable to define the $L_0$ gradient regularization as the regular term.

3. Similar Group-Based Sparse Model with $L_0$ Gradient Regularization

As mentioned in section 2, this section introduces the $L_0$ gradient regularization[11-13] in the similar group-based sparse model and describes the determination of a suitable solution. The proposed model is formulated as:

$$
\hat{x} = \arg\min_{x} \frac{1}{2} \|HD_G \circ a_G - y\|_2^2 + \lambda_1 \|a_G\|_0 + \lambda_2 \|\nabla \hat{x}\|_0
$$

where $\lambda_1$ and $\lambda_2$ are the weight parameters; $\|\nabla \hat{x}\|_0$ represents the norm of the gradient and can be used to control the number of non-zero gradients.

We use the split Bregman iteration (SBI) method to divide the complex problem into subproblems, which are easy to solve. The first subproblem is updated by:

$$
x^{k+1} = \arg\min_{x} \frac{1}{2} \|HD_G \circ a_G - y\|_2^2 + \lambda_1 \|a_G\|_0 + \lambda_2 \|\nabla x\|_0 + \frac{\mu}{2} \|x - D_G \circ a_G - b_k\|_2^2
$$

The second subproblem is updated by:

$$
a^{k+1}_G = \arg\min_{a_G} \lambda_1 \|a_G\|_0 + \frac{\mu}{2} \|x^{k+1} - D_G \circ a_G - b_k\|_2^2
$$

The update of $b^{k+1}$ is as follows:

$$
b^{k+1} = b^k - (x^{k+1} - D_G \circ a^{k+1}_G)
$$

For clarity, $k$ is ignored in the follow discussion.
3.1. The x Subproblem
Given $a_G$, the solution is as follows:

$$x = \arg\min_x \frac{1}{2} ||Hx - y||^2 + \lambda_2 ||Vx||_0 + \frac{\mu}{2} ||x - D_G \cdot a_G - b||^2$$  \hspace{1cm} (7)

The TV norm is usually a relaxation of the $L_0$ norm, but this operation will smooth the boundary of the regions [13, 14]. In this paper, we introduce the region fusion criterion [5,10] to solve the x subproblem. With $M = D_G \cdot a_G$, equation (7) can be written as:

$$x = \arg\min_x \sum_{i=1}^L \frac{1}{2} ||x_i - y_i||^2 + \mu \sum_{i=1}^L ||x_i - M_i \cdot b_i||^2 + \frac{\mu}{2} \sum_{i=1}^L ||x_i \cdot x_j||_0$$  \hspace{1cm} (8)

where $\Omega$ represents the missing area in depth images, $L$ represents the number of pixels, $N_i$ represents the four neighborhoods of the $i^{th}$ pixel and $x_i$, $y_i$, $M_i$ and $b_i$ represent the values of $i^{th}$ pixel in $x$, $y$, $M$ and $b$, respectively.

According to [10], the objective function is defined as follows:

$$f = \min_{x_i, y_i} \frac{\alpha_0}{2} ||x_i - y_i||^2 + \frac{\beta}{2} ||x_j - y_j||_\Omega + \beta \nabla c_y ||x_i - x_j||_0$$ \hspace{1cm} (9)

where $\beta$ is a parameter and $\alpha_0$ represents the number of pixels in $i^{th}$ region. Similarly, a is obtained. Whether the two regions should be fused becomes:

$$\{x_i, y_j\} = \begin{cases} \{A, A\} & \text{if } f_A \leq f_B \\ \{B_i, B_j\} & \text{otherwise} \end{cases}$$  \hspace{1cm} (10)

where:

$$A = \frac{\tilde{\alpha}_i y_i + \mu \alpha g (M_i + b_i)}{(1 + \tilde{\alpha}_i + \alpha g)}, \quad B_i = \frac{\tilde{\alpha}_i y_i + \mu \alpha g (M_i + b_i)}{\tilde{\alpha}_i + \mu \alpha g}, \quad B_j = \frac{\tilde{\alpha}_j y_j + \mu \alpha g (M_j + b_j)}{\tilde{\alpha}_j + \mu \alpha g}$$  \hspace{1cm} (11)

where $\tilde{\alpha}_i$ represents the number of pixels in the $i^{th}$ region but not in the missing region $\Omega$. Similarly, $\alpha g$ is obtained.

We introduce the region fusion method to solve the $L_0$ gradient minimization. This method ensures that the solution converges rapidly and is accurate.

3.2. The $a_G$ Subproblem
Given $x$, the solution is as follows:

$$a_G = \arg\min_{a_G} \frac{1}{2} ||a_G||_0 + \frac{\mu}{2} ||x - D_G \cdot a_G - b||^2$$  \hspace{1cm} (11)

Let $u = D_G \cdot a_G, r = x - b$. Then, Eqn. (11) becomes:

$$a_G = \arg\min_{a_G} \frac{1}{2} ||a_G||_0 + \frac{\mu}{2} ||r - u||^2$$  \hspace{1cm} (12)

In each iteration, estimates $r_{e_i}$ of similar groups $x_{e_i}$ are expressed using SVD.

$$r_{e_i} = U_{G_e} \text{diag}(\sigma_{G_e}) V_{G_e}^T = \sum_{j=1}^m \gamma_{G_{e_i}} (u_{G_{e_i}} \otimes v_{G_{e_i}}) = D_{G_e} \gamma_{G_{e_i}}$$  \hspace{1cm} (13)

So the formula for solving each similar group is as follows:

$$a_{G_{e_i}} = \arg\min_{a_{G_{e_i}}} \frac{1}{2} ||a_{G_{e_i}} - \gamma_{G_{e_i}}||^2 + \tau ||a_{G_{e_i}}||_0$$  \hspace{1cm} (14)
where \( \tau = \frac{i K}{\mu N} \). So we can obtain each similar group \( \mathbf{a}_{c_i}^\wedge \) as follows:

\[
\mathbf{a}_{G_i}^\wedge = \text{hard}(\gamma_{\mathbf{a}_{G_i}}, \sqrt{2\tau})
\]

(15)

where \( \text{hard}(\cdot) \) denotes hard thresholding. Each similar group is solved according to equation(15), which can be integrated to obtain \( \mathbf{a}_{c_i}^\wedge \).

4. Experiments and Results

In this paper, our experimental platform is MATLAB version 7.0. We use subjective visual effects and objective parameters to compare the results.

In this experiment, we use two datasets: Middlebury datasets[15-17] and NYUv2 datasets [18]. As comparative algorithms, we use the traditional algorithms: FOE[19], BPFA[20] and GSR[9].

In experiment 1, there are no perfect depth images for comparison. The objective metrics can be assessed on the basis of one objective parameters: Natural image quality evaluator(NIQE). In experiment 2, the objective metrics can be assessed on the basis of two objective parameters: PSNR and FSIM.

![Figure 2. Experimental test images (1)](image1)

(a) Cloth depth image (b) Art depth image (c) Lampshade depth image

Figure 2. Experimental test images (1)

![Figure 3. Experimental test images (2)](image2)

(a) Bedroom depth image (b) Lamp depth image (c) Kitchen depth image (d) Corrupted Bedroom (e) Corrupted Lamp (f) Corrupted Kitchen

Figure 3. Experimental test images (2).
Figure 4. Visual contrasting of the inpainting results (1)

(a) FOE            (b) BPFA           (c) GSR           (d) Proposed

(e) FOE            (f) BPFA          (g) GSR           (h) Proposed

(i) FOE            (j) BPFA           (k) GSR           (l) Proposed

Figure 5. Visual contrasting of the inpainting results (2).

(a) FOE            (b) BPFA           (c) GSR           (d) Proposed

(c) FOE            (f) BPFA          (g) GSR           (h) Proposed

(i) FOE            (j) BPFA           (k) GSR           (l) Proposed

Table 1. NIQE in Experiment (1)

| Image   | Algorithm (NIQE) | FOE  | BPFA | GSR  | Proposed |
|---------|------------------|------|------|------|----------|
|         |                  |      |      |      |          |
| Cloth   |                  | 13.5199 | 8.5898 | 8.4712 | 8.4011 |
| Art     |                  | 14.1812 | 13.0225 | 9.7045 | 9.5476 |
| Lampshade |                | 15.9002 | 12.4485 | 11.1468 | 8.7534 |
Table 2. PSNR and FSIM in Experiment(2)

| Image   | Algorithm (PSNR/FSIM) | FOE   | BPFA  | GSR   | Proposed |
|---------|------------------------|-------|-------|-------|----------|
| Kitchen |                        | 32.4552/0.9885 | 32.6152/0.9917 | 33.1509/0.9912 | 33.2441/0.9918 |
| Bedroom |                        | 32.5153/0.9932 | 37.0632/0.9925 | 33.2441/0.9918 | 33.2441/0.9918 |
| Lamp    |                        | 32.7676/0.9876 | 36.8944/0.9928 | 37.0632/0.9925 | 33.2441/0.9918 |

Subjectively, as shown in figure 4 and 5, all four algorithms can meet the visual requirements. The effect of the FOE algorithm is fuzzy, particularly in Experiment (2). The BPFA algorithm can repair the images smoothly, but the algorithm also easily blurs boundaries. The GSR algorithm can improve the effect on boundaries, but the algorithm readily causes blurring in areas where the grayscale value is not very different. To a certain extent, the proposed algorithm improves the inpainting effect.

Objectively, our proposed algorithm is superior to the other three algorithms, as shown in Tables 1 and 2. For different types of images, the objective parameters of our proposed algorithm result in improvements.

5. Conclusions
The main research problem is the inpainting of single depth images. Similar group-based sparse representation with an $L_0$ gradient regularization can repair depth images better due to the properties of the depth images. Then, we extend adaptive dictionary learning and region fusion to solve two subproblems. The experiments show that the proposed algorithm obtains improved inpainting effects in terms of subjective visual effects and objective parameters because the algorithm fully utilizes the properties of depth image.

Acknowledgements
This work has been supported by the National Natural Science Foundation of China under grant nos. U1704132 and 11747089.

References
[1] Zhang J, Zhao D B, and Xiong R Q 2014 IEEE Transactions on circuits and systems for video technology. 24 915.
[2] Afonso M, Bioucas-Dias J, and Figueiredo M 2010 IEEE Transactions on Image Processing. 19 2345.
[3] Gu S H, Zuo W M, and Guo S 2017 Proc. IEEE Conf. on computer vision and pattern recognition (New Jersey: Piscataway/American IEEE) p 712.
[4] MA G R 2017 Journal of visual communication and image representation. 47 36.
[5] Xue H Y, Zhang S M, and Cai D 2017 IEEE Transactions on Image Processing. 26 4311.
[6] Yao J W, Xu Z, and Huang X L 2018 Medical image analysis. 44 14.
[7] Wang Y L, Yang J F, and Yin W 2008 SIAM Journal on Imaging Sciences. 1 248.
[8] Zhang H L, Tang L M, and Fang Z 2018 Signal Processing. 143 69.
[9] Zhang J, Zhao D B, and Gao W 2014 IEEE Transactions on Image Processing. 23 3336.
[10] Nguyen R M H, and Brown M S 2015 Proc. IEEE Conf. on on Computer Vision p 208.
[11] Beck A, and Teboulle M 2009 IEEE Transactions on Image Processing. 18 2419.
[12] Li Z, Malgouyres F, and Zeng T 2017 Journal of Mathematical Imaging and Vision. 59 296.
[13] Xu L, Lu C W, and Xu Y 2011 ACM Transactions on Graphics. 30 1.
[14] Xu L, Zheng S C, and Jia J Y 2013 Proc. IEEE Conf. on computer vision and pattern recognition (New Jersey: Piscataway/American IEEE) p 1107.
[15] Scharstein D, and Szeliski R 2003 Proc. IEEE Conf. on computer vision and pattern recognition (New Jersey: Piscataway/American IEEE) p 195.
[16] Scharstein D, and Pal C 2007 Proc. IEEE Conf. on computer vision and pattern recognition (New Jersey: Piscataway/American IEEE) p 17.
[17] Hirschmüller H, and Scharstein D 2007 Proc. IEEE Conf. on computer vision and pattern
recognition (New Jersey: Piscataway/American IEEE) p 1.

[18] Nathan S, Derek H, Pushmeet K and Rob F 2012 *Proc. European Conf. on computer vision* (Germany: Berlin / Germany Springer) p 746.

[19] Roth S, and Black M J 2009 *International Journal of Computer Vision*, 82 205.

[20] Zhou M, Chen H, and Paisley J 2012 *IEEE Transactions on Image Processing*, 21 130.