Research Article

Hand Depth Image Denoising and Superresolution via Noise-Aware Dictionaries

Huayang Li,[1] Dehui Kong,[1] Shaofan Wang,[1] and Baocai Yin[2]

[1]*Beijing Key Laboratory of Multimedia & Intelligent Software Technology, College of Metropolitan Transportation, Beijing University of Technology, Beijing 100124, China*
[2]*Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China*

Correspondence should be addressed to Shaofan Wang; wangshaofan@bjut.edu.cn

This paper proposes a two-stage method for hand depth image denoising and superresolution, using bilateral filters and learned dictionaries via noise-aware orthogonal matching pursuit (NAOMP) based K-SVD. The bilateral filtering phase recovers singular points and removes artifacts on silhouettes by averaging depth data using neighborhood pixels on which both depth difference and RGB similarity restrictions are imposed. The dictionary learning phase uses NAOMP for training dictionaries which separates faithful depth from noisy data. Compared with traditional OMP, NAOMP adds a residual reduction step which effectively weakens the noise term within the residual during the residual decomposition in terms of atoms. Experimental results demonstrate that the bilateral phase and the NAOMP-based learning dictionaries phase corporately denoise both virtual and real depth images effectively.

1. Introduction

With the development of 3D range imaging devices such as laser scanner, Kinect sensor, and Time-of-Flight (ToF) cameras, depth images are widely used in various research fields including computer vision, computer graphics, virtual reality, and human computer interaction. While laser scanners provide 3D measurements with precise accuracy, Kinect sensor and ToF cameras provide a convenient way to accomplish 3D range imaging in faithful time, which facilitate many applications with high requirement of efficiency and convenience.

However, depth images provided by Kinect sensor or ToF cameras, either in structured light principle or in ToF principle, suffer from lower quality and resolutions because of the deficiency of received light speckles and the noise incurred from ranging environment. Typically, depth images produce holes, missing regions, or unstable boundaries and nonzero-mean Gaussian noise (see Figure 1).

The research work devoted to enhancement of depth images, including superresolution of depth images and denoising depth images, can be roughly divided into three categories: filtering methods, probabilistic methods, and sparse representation methods. In general, filtering methods [1–6] perform depth enhancement by using filters, based on the assumption that faithful depth data and noise be separable in frequency domains; probabilistic methods [7–12] formulate the depth enhancement as the uncertainty problem of depth measurement and use probabilistic graphical models to resolve; sparse representation methods [13–18] model the depth enhancement problem as a sparse optimization by assuming that faithful depth data have an underlying sparse or low-rank structure.

Hand depth image denoising and superresolution are important for hand-based human-machine interaction. Although many research works have been proposed for RGB/depth image denoising and superresolution, conventional approaches do not work well for hand depth images. This is because the resolutions of depth images captured from Kinect sensor are 512×424, where the hand takes only a very small subregion (typically 170×150 in our experiments). Thus traditional approaches usually confuse depth with noise and are unsuitable for such small scale depth data.

This paper proposes hand depth image denoising and superresolution using bilateral filters and NAOMP-based dictionaries. The bilateral filtering phase recovers singular
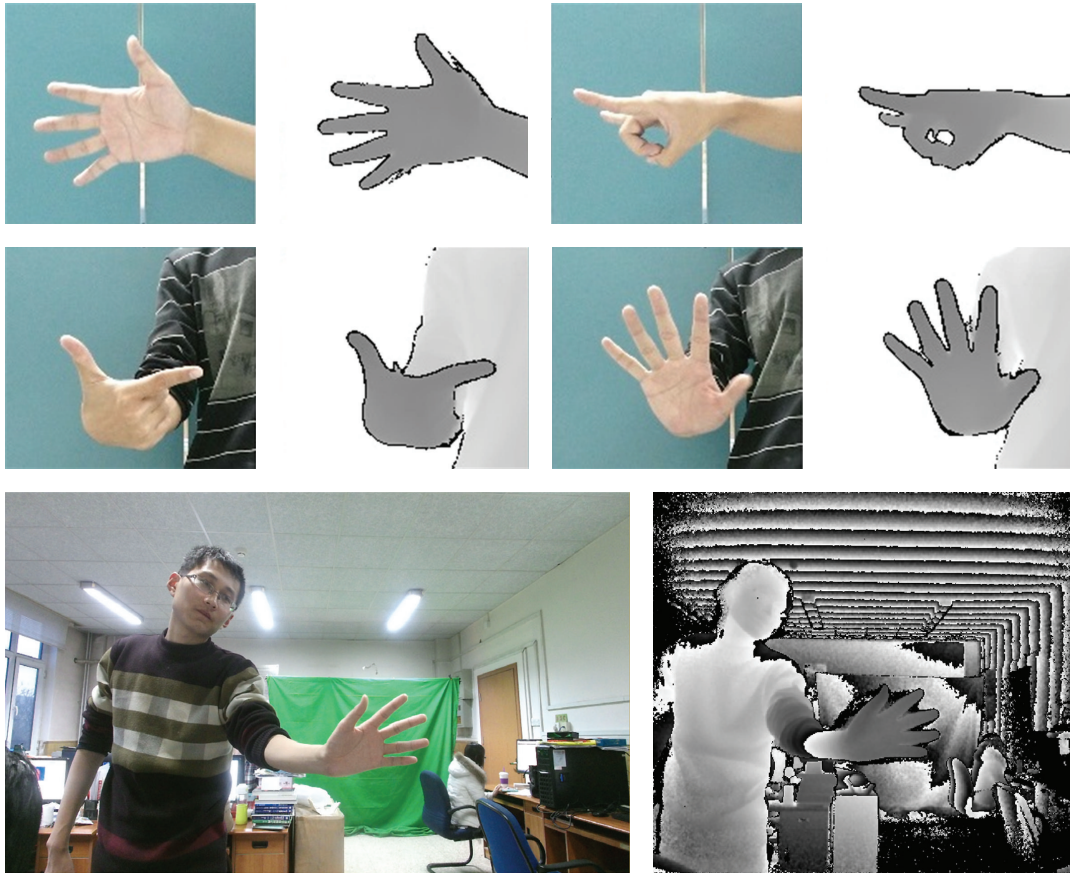
Figure 1: Color/depth images of human hands captured from Kinect sensor v2.

points and removes artifacts on silhouettes by averaging depth data with restrictions of both depth difference and RGB similarity. The dictionary learning phase uses the K-SVD method with the NAOMP for training dictionaries to separate faithful depth from noisy data. While traditional orthogonal matching pursuit (OMP) works well for training dictionaries to denoise RGB images, the performance of denoising depth images deteriorates badly as depth images involve nonzero-mean Gaussian noise. As a result, traditional dictionary learning algorithm (e.g., OMP-based K-SVD) cannot prevent noisy data from penetrating dictionaries which results in an unsatisfied denoising effect. To improve traditional OMP-based K-SVD for fitting nonzero-mean Gaussian noise which frequently appears in depth data, we propose NAOMP for replacing traditional OMP in K-SVD, where the noise term within residuals is weakened in each atom updating step. Such a modification gives dictionaries capable of representing faithful depth data in a more precise fashion. Experimental results show that the proposed bilateral filters and NAOMP-based dictionaries corporately give promising results for denoising both virtual and real depth data, compared with traditional bilateral filters and traditional OMP-based dictionaries.

This paper is organized as follows. Section 2 reviews previous work on enhancing depth images. Section 3 proposes the bilateral filtering phase using RGBD data, which first

recovers singular points and then removes incorrect points over silhouettes and finally removes nonsingular points over nonsilhouette regions. Sections 4 and 5 propose depth image denoising and superresolution, respectively, both using learning dictionaries via NAOMP-based K-SVD. Section 6 shows experimental results with both virtual and real hand depth images.

2. Previous Work

Previous work on enhancing depth images is reviewed in this section, including superresolution of depth images and denoising depth images in the following three categories.

2.1. Filtering Methods. Yang et al. [1] construct a 3D volume of depth probability, impose a bilateral filter over the volume iteratively, and obtain high-resolution depth images by taking the winner-takes-all approach on the weighted volume and a subpixel refinement afterward. Huhle et al. [2] present a two-stage depth enhancement method, which first removes outliers from depth data and then performs smoothing via a nonlocal means filter which uses the similarity of both color and intrapatch of depth. Wasza et al. [3] propose a GPU-based depth image preprocessing, including a normalized convolution for restoring depth images, a bilateral temporal

averaging for dynamic scenes, and a guided filter for edge-preserving denoising. Min et al. [4] propose a joint histogram of depth maps for measuring color similarity between reference and neighboring pixels and find a global mode solution via ℓ_1-norm minimization for depth video enhancement. Fu et al. [5] propose a spatial-temporal denoising algorithm which exploits both the intraframe spatial correlation and the interframe temporal correlation to fill the depth hole and suppress the depth noise. Camplani and Salgado [6] propose a joint-bilateral filtering framework for denoising depth images, by evaluating missing depth values with a filter to neighboring pixels involving both spatial and temporal information, with the filter weights selected as a function of a photometric similarity measure of the neighbor pixels.

2.2. Probabilistic Methods. Mac Aodha et al. [7] explore the height field of patches of low-resolution depth images and select high-resolution candidate depth patches by solving a Markov Random Field (MRF) labeling problem. Shen and Cheung [8] use depth layers to account for the differences between foreground objects and background scene, the missing depth value phenomenon, and the correlation between color and depth channels and consider the depth layer labeling as a maximum a posteriori estimation problem. Wang et al. [9] evaluate the confidence of depth map for adaptive weighting of MRF energy terms and introduce a guided depth recovering method in the framework of MRF optimization for handling large holes across multiple image regions. Li et al. [10] segment the input low-resolution depth image into several regions with different labels which correspond to a high-resolution counterpart on training images and formulate the depth superresolution as an MRF-based patchwork assembly problem. Hui and Ngan [11] propose a variational-based depth map enhancement by fusing the depth maps from the active sensor of a moving RGBD system and the depth cues from an induced optical flow. Yang et al. [12] propose a regression method for enhancing depth images using RGB-D data, which first fits the regression model for depth images and then designs pixel-wise regression predictors using the similarity of depth images and the accompanied color images.

2.3. Sparse Representation Methods. Schuon et al. [13] combine several low-resolution noisy depth images of a static scene from slightly displaced viewpoints and merge them into a high-resolution depth image with ToF calibration data. The depth superresolution problem is formulated as an optimization of a data reconstruction term plus a sparsity term of spatial gradient for separating noise from features. Li et al. [14] propose a novel joint example-based depth map superresolution method which reconstructs high-resolution depth images by learning a mapping function from a set of training samples of an image database. Kiechle et al. [15] propose a joint intensity and depth cosparse model for depth map superresolution, by assuming that the cosupports of corresponding intensity and depth image structures be aligned. Zheng et al. [16] propose constructing multiple dictionaries with different structures and different number of atoms for sparse representing each low-resolution patch of

depth images. Xie et al. [17] learn a coupled dictionary with local coordinate constraints and incorporate an adaptively regularized shock filter to sharpen the edges and implement both depth superresolution and depth denoising. Lu et al. [18] assemble similar RGBD patches into a low-rank matrix in order to prevent the noise or weak correlation between color and depth.

3. RGBD-Based Bilateral Filters

This section proposes the bilateral filtering phase, which preprocesses depth images by using bilateral filters with both depth and RGB restrictions. The phase includes the following three steps. The depth of each singular point of depth images is corrected by averaging depths over the neighborhood according to a rule of both RGB comparison and depth histogram difference; the depth of each point over silhouettes is corrected by averaging depths over the neighborhood according to a rule of both RGB comparison and depth difference; the depth of other points is corrected using traditional bilateral filters, that is, by averaging depths over spatial neighborhood.

Let I_q be the intensity values/depth value at the pixel q of an intensity/depth image. Traditional filters at the pixel p with respect to its neighborhood Ω_p are given by

$$J_p = \frac{1}{k_p} \sum_{q \in \Omega_p} I_q f(p, q) g\left(\left\|I_p - I_q\right\|_2\right),$$

$$k_p = \sum_{q \in \Omega_p} f(p, q) g\left(\left\|I_p - I_q\right\|_2\right),$$

(1)

where $\Omega_p = \{q \in \mathbb{N}^2 : |p(1) - q(1)| \leq \epsilon, |p(2) - q(2)| \leq \epsilon\}$ denotes an $\epsilon \times \epsilon$ spatial neighborhood of p with $p(1)$, $p(2)$ representing the row index, the column index of p within images, respectively, k_p denotes the normalization term, $f(p, q)$ is the 2D Gaussian smoothing kernel (known as the domain term) which measures the closeness of the pixels, and $g(\|I_p - I_q\|_2)$ is the 1D Gaussian smoothing kernel (known as the range term) which measures the similarity of RGB/depth values of the pixels in RGB/depth images.

3.1. Recovering Singular Points. The singular points of a depth image are referred to as the points whose depth is undetected. For each singular point p of depth images, we first set an initial depth value D_p^{initial} at p as the average of depths which are no smaller than $(1/2)D_{\max}^{\Omega}$ in the neighborhood Ω_p of p (choosing such an initial value because experimental results indicate that the depths over the neighborhood of a singular point belong to two regions generally: an interval determined by the depths of the target hand and an interval determined by the depths of the background. Experimental results show that $(1/2)D_{\max}^{\Omega}$ separates two intervals well and hence gives an initial approximation of depth at p), where $D_{\max}^{\Omega}$ is the maximum depth over Ω_p, and then choose a suitable subset of Ω_p whose histogram of depth is much greater than the histogram of depth at p (choosing such a subregion of the neighborhood of p overcomes directly choosing a spatial
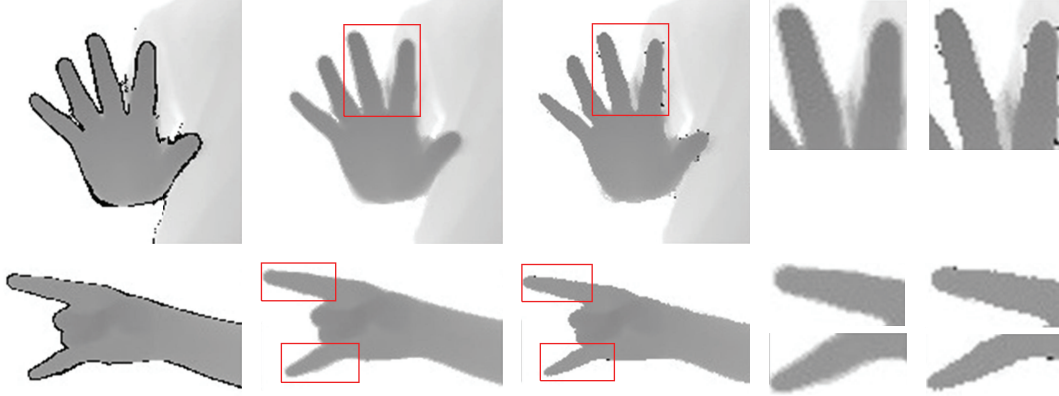
FIGURE 2: Left to right: noisy depth images, images preprocessed by traditional filters, images preprocessed by the proposed filters (2)–(4), and comparison details. One can check from details that the proposed filters (2)–(4) make artifacts on silhouettes either removed or remained with a small number while traditional filters turn artifacts into a shadow effect. Although those shadow artifacts give smaller depth values, the continuous region where the artifacts locate make them difficult to be removed in the dictionary learning phase.

neighborhood of p, because the RGB comparison improves the confidence of the pixels and because the histogram of the depth avoids producing shadows within the silhouette. We illustrate this improvement in Figure 2). Finally we select the filtered value of the depth at the pixel p by averaging the depth data with respect to both the domain term and the range term provided that the confidence pixels are enough; otherwise the singular points remain as the same value (such untreated singular points are much less than before and are dispersed within the depth image and can hence easily be treated in the second phase). That is,

$$J_p = \frac{1}{k_p} \sum_{q \in \widehat{\Omega}_p} D_q f(p,q) g\left(\left\|I_p - I_q\right\|_2\right) \quad \text{if } \left|\widehat{\Omega}_p\right| \geq \epsilon_1, \quad (2)$$

where I_q, D_q denote the RGB, depth values at the pixel q of an image, respectively, k_p denotes the normalization term, and $\widehat{\Omega}_p = \{q \in \Omega_p : \text{hist}_{\Omega_p}(D_q) \geq \text{hist}_{\Omega_p}(D_p^{\text{initial}}) + \epsilon_2, \|I_p - I_q\|_2 \leq \epsilon_3\}$ denotes the confidence subset of a square neighborhood Ω_p of p with both a depth difference restriction and an intensity similarity restriction, with $\text{hist}_\Omega(d) := |\{(i, j) \in \Omega : D_{i,j} = d\}|$ denoting the number of pixels within Ω whose depth value is equal to d.

3.2. Removing Incorrect Points on Silhouettes. For all nonsingular points p located the silhouettes of depth images, we modify its depth value at the pixel p by first choosing a suitable subset of Ω_p whose depth is much greater than the depth of p and whose RGB values are similar to p's and then averaging the depth data within the background domain (corresponding to the restriction of depth difference) with similar RGB values at p, provided that such suitable pixels are enough. That is,

$$J_p = \frac{1}{k_p} \sum_{q \in \widehat{\Omega}_p} D_q f(p,q) g\left(\left\|I_p - I_q\right\|_2\right) \quad \text{if } \left|\widehat{\Omega}_p\right| \geq \epsilon_4, \quad (3)$$

where $I_q, D_q, k_p, f(\cdot, \cdot)$, and $g(\cdot)$ have the same meanings as in (2) and $\widehat{\Omega}_p = \{q \in \Omega_p : D_q \geq D_p + \epsilon_5, \|I_p - I_q\|_2 \leq \epsilon_6\}$ denotes the confidence subset of a square neighborhood Ω_p of p with both a depth difference restriction and an intensity similarity restriction.

3.3. Removing Other Incorrect Points. For all other nonsingular points p of depth images, we treat them with a smoothing filter based on the similarity of both depth value and RGB values. That is,

$$J_p = \frac{1}{k_p} \sum_{q \in \Omega_p} D_q f(p,q) g\left(\left|D_p - D_q\right|\right) g\left(\left\|I_p - I_q\right\|_2\right), \quad (4)$$

where $I_q, D_q, k_p, f(\cdot, \cdot)$, and $g(\cdot)$ have the same meanings as in (2) and $\Omega_p = \{q \in \mathbb{N}^2 : |p(1) - q(1)| \leq \epsilon_7, |p(2) - q(2)| \leq \epsilon_7\}$ denotes an $\epsilon_7 \times \epsilon_7$ square neighborhood of p (different from subtle filtering (2) and (3) where both of the neighborhoods are determined by a depth difference restriction and an intensity similarity restriction, the filtering (4) simply selects a square neighborhood without additional restrictions. This is because the filtering (2) and (3) treats singular points and incorrect points on silhouettes which require careful treatment, whereas the filtering (4) performs a smoothing effect over the whole square neighborhood of the point which needs smoothing.).

Figures 2 and 3 show some comparison results of depth images using traditional filters [6] and the proposed filters (2)–(4). In a word, the proposed filters recover singular points using neighboring points with a quantity restriction (2) and correct artifacts using neighboring points with an intensity restriction (3), while traditional bilateral filters accomplish such tasks by directly averaging points within spatial neighborhood. Therefore, traditional filters always produce new artifacts while the proposed filters either correct the artifacts over silhouettes or suppress them in a small number without creating new ones, so that they can be easily treated in the next dictionary learning phase.
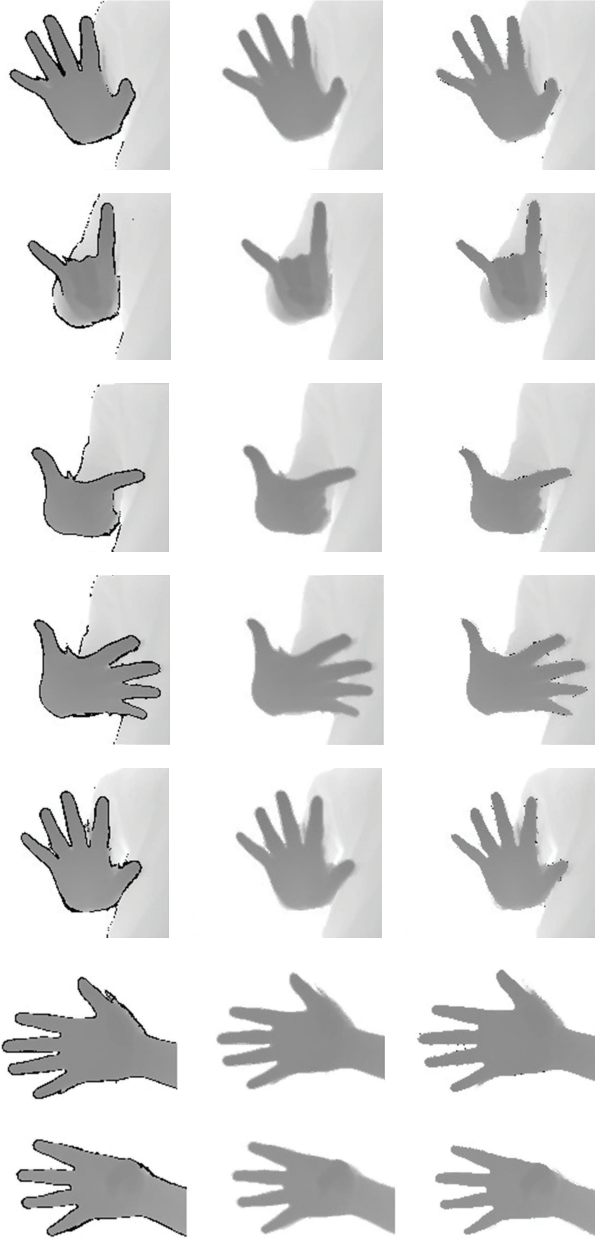
Figure 3: Bilateral filters on hand depth images. Left: noisy depth images. Middle: depth images with bilateral filters (1). Right: depth images with bilateral filters (2)–(4).

4. Hand Depth Denoising Using Noise-Aware Dictionaries

This section proposes the dictionary-based denoising phase, followed by the bilateral filtering phase given in Section 3. Recovering the original image from a degraded image with additive white Gaussian noise can be modelled by solving the ill-posed system $\mathbf{Y} = \mathbf{X} + \mathbf{v}$, where $\mathbf{Y}$ denotes the degraded image, $\mathbf{X}$ denotes the original image, and $\mathbf{v}$ denotes the noise term. Sparse representation is an important tool for solving such an ill-posed system. According to sparse representation, natural images can be represented by a linear combination of

a series of overcomplete basis (known as a *dictionary*) with very few nonzero combinational coefficients. Therefore, by denoting $\mathbf{D} \in \mathbb{R}^{n \times K}$ to be such a dictionary whose columns are basis vectors (known as *atoms*) with $K \gg n$, the original image can be obtained by imposing an ℓ_0-norm constraint of the coefficients α over the system $\mathbf{Y} = \mathbf{D}\alpha + \mathbf{v}$. In particular, we select image patches of size $\sqrt{n} \times \sqrt{n}$ pixels randomly from $\mathbf{Y}$ as training data and order each patch lexicographically as column vectors $\mathbf{Y}_{ij} \in \mathbb{R}^n$. Then we obtain a trained dictionary $\mathbf{D}$ and sparse coefficients α_{ij} simultaneously via the following ℓ_0-minimization of all coefficients of training data:

$$\min_{\mathbf{D}, \{\alpha_{ij}\}} \quad \sum_{i,j} \left\| \alpha_{ij} \right\|_0$$
$$\text{s.t.} \quad \mathbf{D}\alpha_{ij} \approx \mathbf{Y}_{ij} \tag{5}$$
$$\forall (i, j) \in \Omega_{\text{train}},$$

where Ω_{train} denotes the training set of indices of patches of $\mathbf{Y}$ we randomly select. The original image $\mathbf{X}$ is finally reconstructed by assembling all image patches $\mathbf{X}_{ij} = \mathbf{D}\alpha_{ij}$ with α_{ij} given by

$$\min_{\alpha_{ij} \in \mathbb{R}^K} \quad \left\| \alpha_{ij} \right\|_0$$
$$\text{s.t.} \quad \mathbf{D}\alpha_{ij} \approx \mathbf{Y}_{ij} \tag{6}$$
$$\forall (i, j).$$

Because (5) and (6) are both nonconvex systems involving the ℓ_0-norm minimization, developing efficient algorithms for solving them is an important task. Elad and Aharon [19] propose the K-SVD algorithm to solve (5), by constructing a dictionary iteratively from training signals with a sparse coding phase and an atom update phase. Within the sparse coding phase of K-SVD and the optimization problem (6), OMP is a greedy algorithm frequently used for solving the ℓ_0-norm minimization.

While traditional K-SVD denoises traditional images well for zero-mean Gaussian additive noise, it is not well suited for denoising depth images with nonzero-mean noise because traditional OMP can hardly separate noise data from noiseless image when the amplitude of noise has an irregular distribution. To remove such noise from depth images in a more effective fashion, traditional OMP is improved by modifying the amplitude of entries of residual whenever such entries have large amplitude and a small quantity. Figure 4 helps readers understand how this idea works. The y-axis represents the value of each component of residuals while the x-axis represents the index of each component of residuals. Let r_{s-1} be the residual obtained in an iteration of OMP. The left subfigure of Figure 4 shows how traditional OMP represents residuals when the noise is zero-mean Gaussian. While r_{s-1} contains four components heavily contaminated by noise (denoted by blue cubes, i.e., the original training data $\mathbf{Y}$ has noise terms in the positions of those four components), the least square fitting, obtained when computing the sparse coefficients, approximates noiseless components
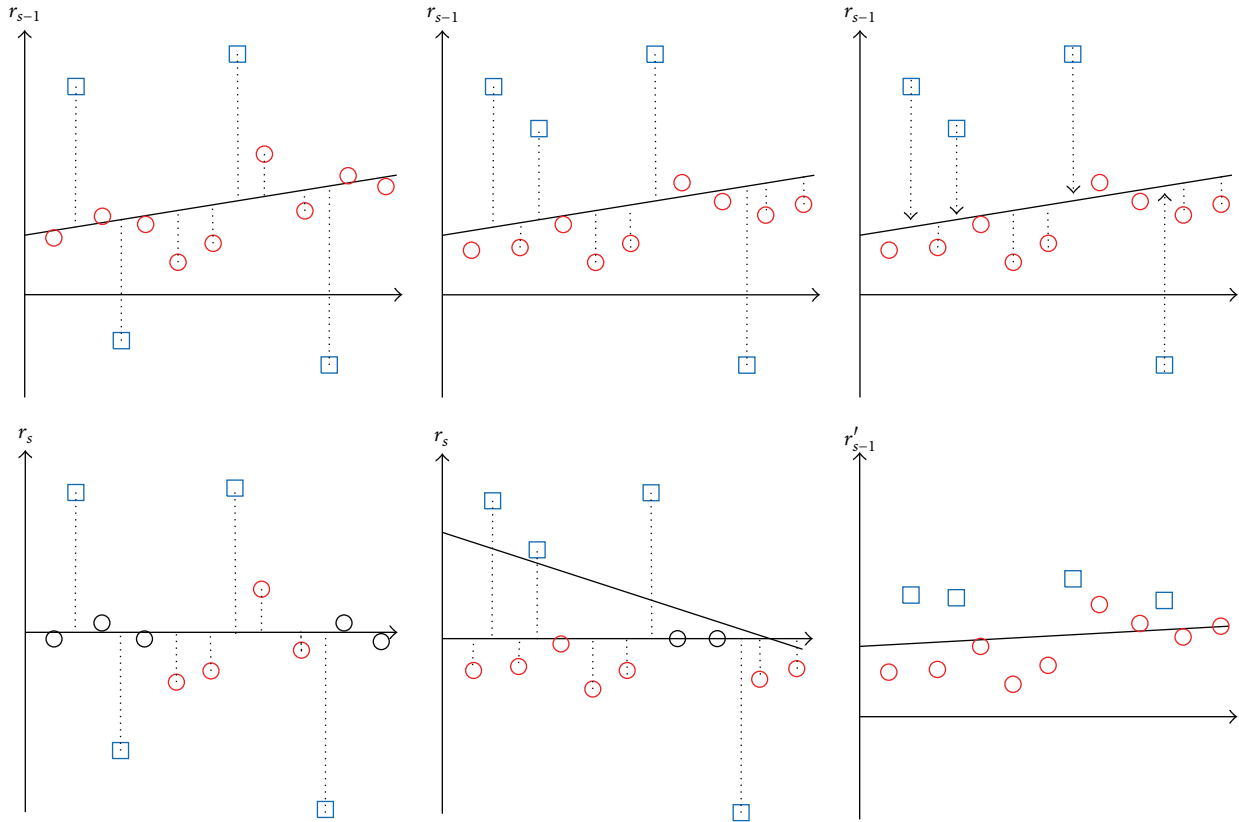
FIGURE 4: The difference between how OMP and NAOMP work on decompositions of residuals with nonzero-mean noise. Left: traditional OMP for zero-mean Gaussian noise. Middle: traditional OMP for nonzero-mean noise. Right: NAOMP for nonzero-mean noise. The x-axis: the index of each component of the residual; the y-axis: the value of each component of the residual. The residual in this figure is a vector of length sixteen. From top to bottom: a procedure for updating a residual and an atom.

well and hence makes most of noiseless components vanish at the next iteration before the decomposition of the noise terms with respect to current atoms, because the stopping criteria of the number of sparse coefficients are satisfied. The middle subfigure of Figure 4 shows how traditional OMP fails to remove noise from depth images when the noise is nonzero-mean. In this case, the least square fitting deviates from most noiseless terms and hence the noisy terms begin decomposition in the next iteration, because neither the criteria of the number of sparse coefficients nor the criteria of the residual amplitude are satisfied. From this step, the new atoms obtained shall be contaminated by noise. The right subfigure of Figure 4 shows how NAOMP improves this issue. When the current residual r_{s-1} contains greater-magnitude entries and those entries have a relatively small number, it is reasonably believed that those entries correspond to the noisy components of training data. In this case, each of those components of r_{s-1} is modified by reevaluating it as the amplitude of corresponding component within the fitting line. Then the updated residual is redecomposed over atoms. By doing so, the least square fitting approximates noiseless terms of residuals and weakens the effect of noisy terms. We illustrate the idea using numerical examples in Appendix.

The NAOMP algorithm is illustrated in Algorithm 3. After performing traditional OMP (lines (4)–(9)), the number of components which is relatively greater than others is checked. Once the number is smaller, the current residual is reevaluated (line (12)) so that the values of greater components approach those of other components. Then the atom and the residual are reupdated as in OMP (lines (13)–(17)). Detailed parameter setting shall be given in Section 6.

5. Hand Depth Image Superresolution Using Noise-Aware Dictionaries

We apply NAOMP to hand depth image superresolution, where we use NAOMP for joint dictionary training. The main idea is similar to the work of [20]; hence we only give the different part of our work (which is the dictionary training algorithm) without giving details. We randomly select patches of virtual hand depth images as a training set, each of which is stacked as a vector, denoted by X_i, $i \in \Omega_{\text{train}}$. Then we obtain its corresponding downsampled version Y_i and form pairwise training sets $\{(X_i, Y_i) : i \in \Omega_{\text{train}}\}$. We denote $\mathbf{D}_h$ and $\mathbf{D}_l$ to be the joint dictionary, which is the sparse representations for high-resolution and low-resolution image

patches, respectively. The joint dictionary is given by the following optimization:

$$\min_{\mathbf{D}_h, \mathbf{D}_l, \{z_i\}} \sum_{i \in \Omega_{\text{train}}} \left(\frac{1}{N} \|X_i - \mathbf{D}_h z_i\|_2^2 + \frac{1}{M} \|Y_i - \mathbf{D}_l z_i\|_2^2 + \lambda \left(\frac{1}{N} + \frac{1}{M} \right) \|z_i\|_0 \right). \tag{7}$$

By rewriting the first and the second terms in a single term of 2-norm, we obtain the optimization model similar to the last section and solve it using the NAOMP-based K-SVD algorithm. To recover the high-resolution patch of an input image $\mathbf{Y}$, we find a sparse representation of each patch of $\mathbf{Y}$ with respect to $\mathbf{D}_l$ and then obtain the corresponding high-resolution patch by combining the high-resolution atoms $\mathbf{D}_h$ with the same sparse coefficients with respect to $\mathbf{D}_l$.

6. Experimental Results

The experimental results are given in this section. The experiments are run on Core™ 2 Quad Q6600 2.4 GHz machine with 2 GB RAM using Visual Studio 2010. The proposed filters are given by (2)–(4), and the NAOMP is given by Algorithm 3. All the parameters in traditional filters [6], the proposed filters (2)–(4), OMP-based K-SVD [19] (Algorithms 1 and 2), and NAOMP-based K-SVD (Algorithms 1 and 3) are given in Table 1, respectively. One can see from Table 1 that traditional OMP-based K-SVD has to select different values for the residual threshold ε according to the intensity of noise, while NAOMP-based K-SVD selects a single value. In fact, according to [19], the denoising effect heavily depends on the choice of the residual threshold. Such a threshold is difficult to determine for OMP-based K-SVD when no a priori information of noise is given, while this threshold is fixed and easily determined in NAOMP-based K-SVD.

6.1. Denoising Virtual Hand Depth Images. Artificial Gaussian noise is added on three virtual hand models and the images are denoised using OMP-based K-SVD and NAOMP-based K-SVD, both without filter preprocessing. Qualitative results are given in Figure 5 (with 0.5%, 2%, and 5% of noise) and quantitative results are given in Table 2. We see that traditional OMP fails to recover the wrist part of models while NAOMP treats them well. Moreover, within all cases, the proposed method provides higher PSNR than OMP-based K-SVD except for the first example with 0.5% noise.

6.2. Denoising Hand Depth Images Obtained from Kinect v2. We show qualitative results of denoising and superresolution of six hand depth images obtained from Kinect sensor v2 in Figure 6 and give the average running time of different approaches in Table 3. The comparison results include five approaches: traditional bilateral filters [6], OMP-based K-SVD [19], NAOMP-based K-SVD, the bilateral filters (2)–(4) plus OMP-based K-SVD, and the bilateral filters (2)–(4) plus NAOMP-based K-SVD.

In general, both bilateral filters preprocess depth images well in that singular points are removed (the different effect of two filters is given in Section 3 using Figures 2 and 3). Moreover, one can see from the fifth column and the sixth column that NAOMP-based K-SVD produces clearer silhouettes than OMP-based K-SVD, as the trained dictionaries from NAOMP remove noisy terms well.

6.3. Discussions. The two-stage depth image denoising and superresolution enhance hand depth images well mainly because of the following two reasons. For one thing, the proposed bilateral filter functions choose suitable neighborhood pixels with exquisite depth and RGB restrictions than traditional filters. For another, the NAOMP modifies the noisy terms of the residual in each atom updating step so that they are reevaluated with closed values to noiseless terms. This enables the residual to decompose in a more noiseless fashion and results in dictionaries which are less contaminated by noise.

It should be noted that the proposed method may fail in denoising and superresolution of depth images of objects other than human hands. This is because while the skin of human hands shares RGB data in a small range, other objects do not show such an advantage, which make the proposed bilateral filters fail to select suitable neighborhood for filter functions.

In future work, the proposed image denoising and superresolution framework shall be developed for enhancing depth images of more complex objects or scenes. More accurate RGB/depth restrictions can be designed for filter functions to preprocess depth images, so that the singular points and artifacts remain in a more dispersing fashion. Furthermore, the modification of residuals within NAOMP can be improved so that the updated atom can represent the noiseless data more precisely.

Appendix

A. OMP versus NAOMP: The First Example

Let

$$\mathbf{Y} = \mathbf{X} + \mathbf{v} = (k + 1, k, k)^\top,$$
$$\mathbf{X} = (k, k, k)^\top \tag{A.1}$$

be an input signal, a faithful signal, respectively, where $\mathbf{v} = (1, 0, 0)^\top$ denotes the noise term, $k \gg 1$, and let the atom number be one. Without loss of generality we assume that both traditional K-SVD and NAOMP-based K-SVD return the same DCT dictionary (the expression of the trained dictionary depends on the training set and residual

---

input: training data $\mathbf{Y}_{ij} \in \mathbb{R}^n, (i, j) \in \Omega_{\text{train}}$, residual threshold ε, and iteration number J
output: a dictionary $\mathbf{D} \in \mathbb{R}^{n \times K}$
(1) Initialize $\mathbf{D}^{(1)}$ as overcomplete DCT dictionary;
(2) **for** $m = 1, \ldots, J$ **do**
(3)   **foreach** $(i, j) \in \Omega_{\text{train}}$ **do** /* *sparse coding phase* */
(4)   Using the OMP (Algorithm 2 or Algorithm 3) to solve
                  $\min_{\alpha_{ij}} \|\alpha_{ij}\|_0$ s.t. $\|\mathbf{Y}_{ij} - \mathbf{D}^{(m)}\alpha_{ij}\|_2 \leq \varepsilon$;
(5)   **end**
(6)   **for** $k = 1, \ldots, K$ **do**   /* *atom update phase* */
(7)   Find the set of patches which use the kth atom
                  $\omega_k \leftarrow \{(i, j) \in \Omega_{\text{train}} : \alpha_{ij}(k) \neq 0\}$;
(8)   **foreach** $(i, j) \in \omega_k$ **do**
(9) $\mathbf{e}_{ij}^k \leftarrow \mathbf{Y}_{ij} - \sum_{l \neq k} \mathbf{d}_l \alpha_{ij}(l)$;   /* $\mathbf{d}_l$ *denotes the lth column of* $\mathbf{D}^{(m)}$ */
(10)   **end**
(11)   Set $\mathbf{E}_k \in \mathbb{R}^{n \times |\omega_k|}$ as the matrix whose columns are $\{\mathbf{e}_{ij}^k\}_{(i,j) \in \omega_k}$;
(12)   Apply SVD decomposition $\mathbf{E}_k = \mathbf{U}\Delta\mathbf{V}^\top$;
(13)   Update the kth column of dictionary $\mathbf{D}^{(m)}$ by $\mathbf{d}_k \leftarrow$ the first column of $\mathbf{U}$;
(14)   Update the coefficient values $\{\alpha_{ij}(k)\}_{(i,j) \in \omega_k}$ to be the first column of $\mathbf{V}$ multiplied by $\Delta(1, 1)$;
(15)   **end**
(16) **end**

ALGORITHM 1: The K-SVD algorithm [19].

---

input: vector $\mathbf{Y} \in \mathbb{R}^n$, dictionary $\mathbf{D} \in \mathbb{R}^{n \times K}$, sparsity S
output: sparse coefficient $\alpha \in \mathbb{R}^K$ of $\mathbf{Y}$ with respect to $\mathbf{D}$
(1) Initialize residual $r_0 \leftarrow \mathbf{Y}$, index set $\Lambda_0 \leftarrow \varnothing$, $\mathbf{D}^{(0)} \leftarrow \varnothing$;
(2) **for** $s = 1, \ldots, S$ **do**
(3)   Denote $\mathbf{d}_j, j = 1, \ldots, K$ to be all columns of $\mathbf{D}$;
(4) $\lambda_s \leftarrow \arg\max_{j=1,\ldots,K} |\langle r_{s-1}, \mathbf{d}_j \rangle|$;
(5)   Update the index set $\Lambda_s \leftarrow \Lambda_{s-1} \cup \{\lambda_s\}$;
(6)   Update the atom matrix $\mathbf{D}^{(s)} \leftarrow [\mathbf{D}^{(s-1)}, \mathbf{d}_{\lambda_s}]$;
(7) $\alpha \leftarrow \arg\min_x \|\mathbf{Y} - \mathbf{D}^{(s)} x\|_2$;
(8)   Update residual $r_s \leftarrow \mathbf{Y} - \mathbf{D}^{(s)}\alpha$;
(9) **end**

ALGORITHM 2: The orthogonal matching pursuit.

---

threshold we choose. Here we make this assumption so that the dictionary can exactly represent the faithful signal):

$$\mathbf{D} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}. \tag{A.2}$$

After the dictionary is obtained, we recover the signal using OMP and NAOMP, respectively. OMP obtains sparse coefficients and the reconstruction signal as follows:

$$\alpha = \arg\min_{x \in \mathbb{R}} \|r_0 - \mathbf{D}x\|_2 = \arg\min_{x \in \mathbb{R}} \|\mathbf{Y} - \mathbf{D}x\|_2$$

$$= k + \frac{1}{3}, \tag{A.3}$$

$$\mathbf{X}_{\text{OMP}} = \mathbf{D}\alpha = \left(k + \frac{1}{3}, k + \frac{1}{3}, k + \frac{1}{3} \right)^\top.$$

Nevertheless, NAOMP obtains the same sparse coefficients $\alpha = k + 1/3$ and finds that the first component of the residual $r_1 = (2/3, -1/3, -1/3)^\top$ is relatively greater than the second and the third ones. As a result, the residual is modified as $r_0'(1) \leftarrow r_0(1) - r_1(1)$ (noticing that $r_0 = \mathbf{Y}$ holds in the first step). Thus the sparse coefficients are recomputed by

$$\alpha = \arg\min_{x \in \mathbb{R}} \|r_0' - \mathbf{D}x\|_2 = k + \frac{1}{9}, \tag{A.4}$$

where $r_0' = (k + 1/3, k, k)^\top$ is the modified residual term, and the reconstruction signal is given by

$$\mathbf{X}_{\text{NAOMP}} = \mathbf{D}\alpha = \left(k + \frac{1}{9}, k + \frac{1}{9}, k + \frac{1}{9} \right)^\top \tag{A.5}$$

which approximates the faithful signal compared with $\mathbf{X}_{\text{omp}}$.

---

input: vector $\mathbf{Y} \in \mathbb{R}^n$, dictionary $\mathbf{D} \in \mathbb{R}^{n \times K}$, sparsity S, residual threshold $\hat{r}$, noise index threshold $\hat{L}$
output: sparse coefficient $\alpha \in \mathbb{R}^K$ of $\mathbf{Y}$ with respect to $\mathbf{D}$

(1) Initialize residual $r_0 \leftarrow \mathbf{Y}$, index set $\Lambda_0 \leftarrow \emptyset$, $\mathbf{D}^{(0)} \leftarrow \emptyset$;
(2) **for** $s = 1, \ldots, S$ **do**
(3)   Denote $\mathbf{d}_j$, $j = 1, \ldots, K$ to be all columns of $\mathbf{D}$;
(4) $\lambda_s \leftarrow \arg\max_{j=1,\ldots,K} |\langle r_{s-1}, \mathbf{d}_j \rangle|$;
(5)   Update the index set $\Lambda_s \leftarrow \Lambda_{s-1} \cup \{\lambda_s\}$;
(6)   Update the atom matrix $\mathbf{D}^{(s)} \leftarrow [\mathbf{D}^{(s-1)}, \mathbf{d}_{\lambda_s}]$;
(7) $\alpha \leftarrow \arg\min_x \|\mathbf{Y} - \mathbf{D}^{(s)} x\|_2$;
(8)   Update residual $r_s \leftarrow \mathbf{Y} - \mathbf{D}^{(s)}\alpha$;
(9) $L \leftarrow \{1 \le l \le n : |r_s(l)| \ge \hat{r}\}$; /* *collect all indices of great components of residual r_s* */
(10)   **if** $|L| \le \hat{L}$ **then** /* *when r_s has only a small number of great components, we re-update the atom by removing those*
             components from r_{s-1} */
(11)   **foreach** $l \in L$ **do** $r_{s-1}(l) \leftarrow r_{s-1}(l) - r_s(l)$;
(12) $\lambda'_s \leftarrow \arg\max_{j=1,\ldots,K} |\langle r_{s-1}, \mathbf{d}_j \rangle|$;
(13)   Re-update the index set $\Lambda_s \leftarrow \Lambda_{s-1} \cup \{\lambda'_s\}$;
(14)   Re-update the atom matrix $\mathbf{D}^{(s)} \leftarrow [\mathbf{D}^{(s-1)}, \mathbf{d}_{\lambda'_s}]$;
(15) $\alpha \leftarrow \arg\min_x \|\mathbf{Y} - \mathbf{D}^{(s)} x\|_2$;
(16)   Re-update residual $r_s \leftarrow \mathbf{Y} - \mathbf{D}^{(s)}\alpha$;
(17)   **end**
(18) **end**

ALGORITHM 3: The noise-aware orthogonal matching pursuit.

B. OMP versus NAOMP: The Second Example

Let

$$\mathbf{Y} = \mathbf{X} + \mathbf{v} = \left((k+1)c^2 + c, c^2 + c, c^2, c^2\right)^\top,$$
$$\mathbf{X} = \left(c^2 + c, c^2 + c, c^2, c^2\right)^\top = c^2\mathbf{d_1} + c\mathbf{d_3} \tag{B.1}$$

be an input signal, a faithful signal, respectively, where $\mathbf{v} = (kc^2, 0, 0, 0)^\top$ denotes the noise term and where $\mathbf{d_1}$, $\mathbf{d_3}$ are atoms of the following dictionary:

$$\mathbf{D} = (\mathbf{d_1}, \mathbf{d_2}, \mathbf{d_3}, \mathbf{d_4}) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{B.2}$$

According to the expression of $\mathbf{X}$, the best choice of atoms is $\{\mathbf{d_1}, \mathbf{d_3}\}$. Suppose that $2 < kc < 10$, $c > 0$. Let us recover the signal using OMP and NAOMP, respectively, both with two iteration steps. For OMP, in the first step, the atom which takes the maximum inner product with $r_0 = \mathbf{Y}$ is $\mathbf{d_1}$, and the residual is accordingly given by

$$r_1 = r_0 - \alpha_1 \mathbf{d_1} = \left(\frac{3}{4}kc^2 + \frac{1}{2}c, -\frac{1}{4}kc^2 + \frac{1}{2}c, -\frac{1}{4}kc^2\right.$$
$$\left. -\frac{1}{2}c, -\frac{1}{4}kc^2 - \frac{1}{2}c\right)^\top, \tag{B.3}$$

where $\alpha_1 = \arg\min_{\alpha \in \mathbb{R}} \|r_0 - \alpha \mathbf{d_1}\|_2$. In the second step, using simple calculation and the inequalities $kc > 2$, $c > 0$, the atom

which takes the maximum inner product with r_1 is $\mathbf{d_4}$. Finally we recover the signal by

$$\mathbf{X}_{\text{OMP}} = \left(c^2 + \frac{1}{3}c\right)\mathbf{d_1} + \left(kc^2 + \frac{2}{3}c\right)\mathbf{d_4} \tag{B.4}$$

with the coefficients given by

$$\left[c^2 + \frac{1}{3}c, kc^2 + \frac{2}{3}c\right] = \arg\min_{\alpha, \beta \in \mathbb{R}} \|\mathbf{Y} - \alpha\mathbf{d_1} - \beta\mathbf{d_4}\|_2. \tag{B.5}$$

Thus the reconstruction error is given by

$$\|\mathbf{X}_{\text{OMP}} - \mathbf{X}\|_2 = \sqrt{k^2 c^4 + \frac{2}{3}c^2}. \tag{B.6}$$

For NAOMP, the first step finds the same atom $\mathbf{d_1}$ and the same residual r_1 as in OMP. Using $kc > 2$ we can see that the first component of r_1 is greater than the other three components and hence we modified the residual by $r'_0(1) \leftarrow r_0(1) - r_1(1)$, which gives

$$r'_0 = \left(\left(\frac{1}{4}k + 1\right)c^2 + \frac{1}{2}c, c^2 + c, c^2, c^2\right)^\top. \tag{B.7}$$

We reselect the atom taking the maximum inner product with r'_0 which is still $\mathbf{d_1}$ and the residual is given by

$$r_1 = \left(\frac{3}{16}kc^2 + \frac{1}{8}c, -\frac{1}{16}kc^2 + \frac{5}{8}c, -\frac{1}{16}kc^2 - \frac{3}{8}c,\right.$$
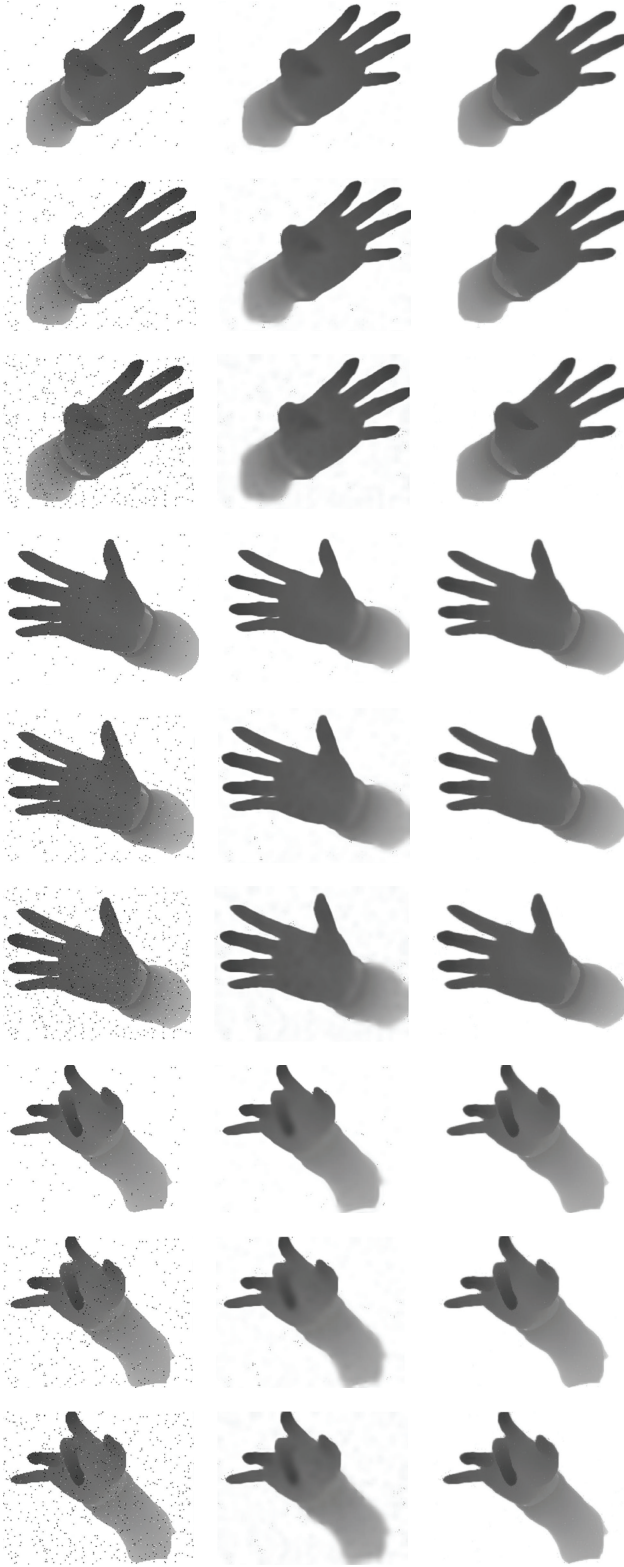$$\left. -\frac{1}{16}kc^2 - \frac{3}{8}c\right)^\top. \tag{B.8}$$

FIGURE 5: Denoising additive Gaussian noise using traditional OMP and NAOMP (both without the filtering phase). Left: hand depth images with artificial noise. Middle: denoised images using traditional OMP. Right: denoised images using NAOMP. Rows 1–3: 0.5%, 2%, and 5% of noise for *hand 1*; rows 4–6: 0.5%, 2%, and 5% of noise for *hand 2*; rows 7–9: 0.5%, 2%, and 5% of noise for *hand 3*.

TABLE 1: Parameter setting in filters [6], filters (2)–(4), OMP-based K-SVD, and NAOMP-based K-SVD.

Stages	Parameters	Values
Traditional bilateral filters [6]	σ for $f(\cdot, \cdot)$	4
	σ for $g(\cdot)$	15
	ϵ	3
Bilateral filters (2)–(4)	σ for $f(\cdot, \cdot)$	4
	σ for $g(\cdot)$	15
	ϵ_1	5
	ϵ_2	30
	ϵ_3	5
	ϵ_4	25
	ϵ_5	30
	ϵ_6	6
	ϵ_7	3
OMP-based K-SVD ([19], Algorithms 1 and 2)	n	64
	K	512
	J	10
	S	30
	ε for Figure 6	10^5
	ε for Figure 5 (0.5% noise)	$5 * 10^4$
	ε for Figure 5 (2% noise)	$1.2 * 10^5$
	ε for Figure 5 (5% noise)	$2 * 10^6$
NAOMP-based K-SVD (Algorithms 1 and 3)	n	64
	K	512
	J	10
	S	30
	ε	10^3
	$\hat{r}$	30
	$\hat{L}$	15
NAOMP-based superresolution (7)	N	25
	M	100
	λ	0.15

In the second step, using simple calculation and the relationship $kc < 10$, the atom taking the maximum inner product with r_1 is $\mathbf{d_3}$. Then we update the residual by

$$r_2 = \left(\frac{1}{2}kc^2, -\frac{1}{2}kc^2, 0, 0 \right)^\top. \tag{B.9}$$

Because no single component has a relatively great magnitude, we do not implement the modification step and hence return the reconstruction signal by

$$\mathbf{X}_{\text{NAOMP}} = c^2\mathbf{d_1} + \left(\frac{1}{2}kc^2 + c \right)\mathbf{d_3} \tag{B.10}$$

with the coefficients given by

$$\left[c^2, \frac{1}{2}kc^2 + c \right] = \arg\min_{\alpha, \beta \in \mathbb{R}} \left\| \mathbf{Y} - \alpha\mathbf{d_1} - \beta\mathbf{d_3} \right\|_2. \tag{B.11}$$
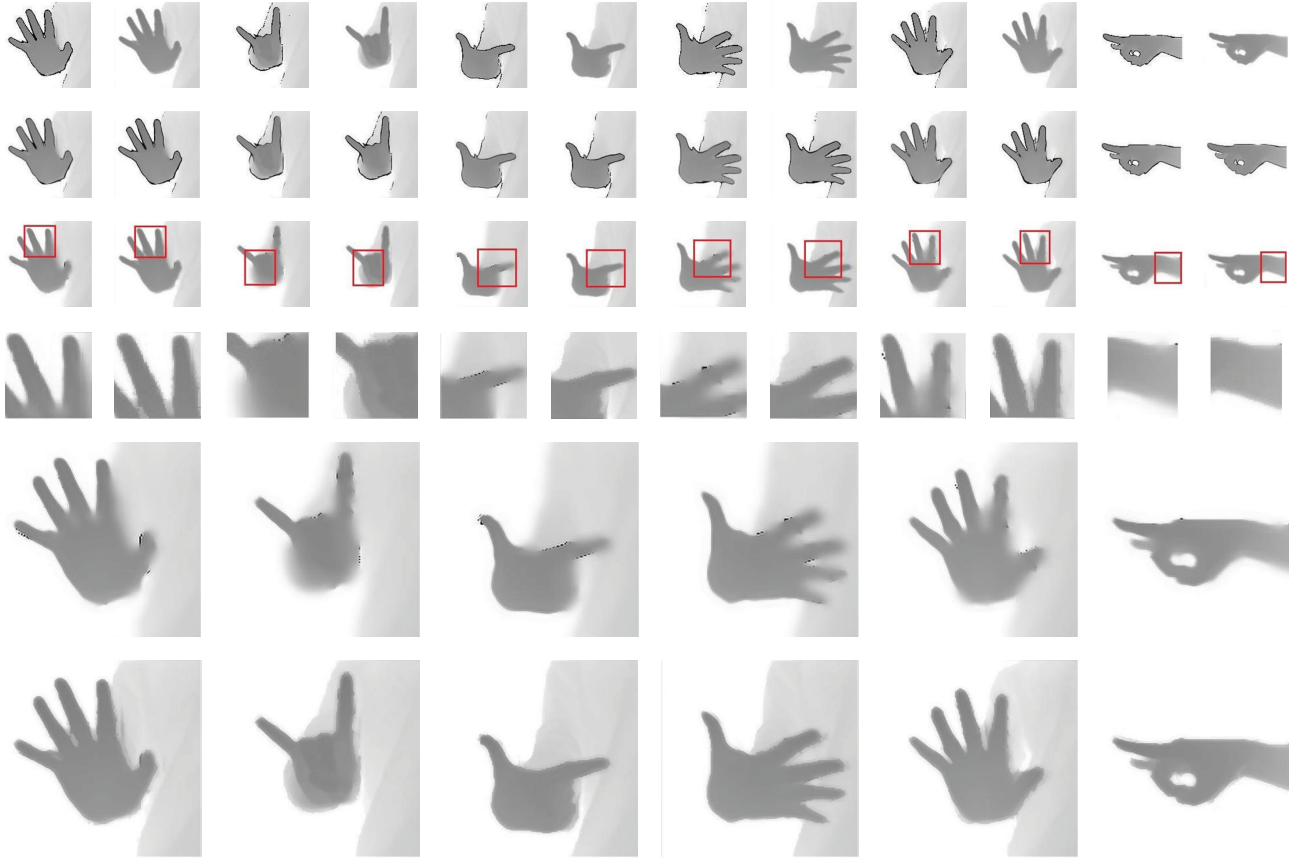
Figure 6: Denoising Gaussian noise from hand depth images captured by Kinect sensor v2. Columns 1–3 are three examples of hand depth images. Row 1 left: noisy depth images of resolutions 150×150. Row 1 right: denoised images using bilateral filters [6]. Row 2 left: denoised images using OMP-based K-SVD [19]. Row 2 right: denoised images using NAOMP-based K-SVD. Row 3 left: denoised images using bilateral filters (2)–(4) plus OMP-based K-SVD [19]. Row 3 right: denoised images using the proposed filters (2)–(4) plus NAOMP-based K-SVD. Row 4: amplified comparison details of images of row 3. Row 5: superresolution of denoised images of left subfigure of row three via OMP-based K-SVD. Row 6: superresolution of denoised images of right subfigure of row three via NAOMP-based K-SVD.

Table 2: PSNR (unit: dB) comparison of three models in Figure 5.

Noise percent	PSNR (hand 1)		PSNR (hand 2)		PSNR (hand 3)	
	OMP	NAOMP	OMP	NAOMP	OMP	NAOMP
0.5%	40.2939	39.8312	39.7280	39.7308	39.7076	41.4870
2%	37.6736	39.6812	37.5300	39.5911	37.0001	41.1889
5%	35.4332	39.4074	35.5310	39.3932	35.0824	40.9597

Table 3: Average running time comparison of six models in Figure 6.

Methods	Time (s)
Filters [6]	19.644
Filters (2)–(4)	9.444
K-SVD [19]	85.848
NAOMP K-SVD (Algorithms 1 and 3)	95.163
Filters [6] + K-SVD [19]	105.492
Filters (2)–(4) + NAOMP K-SVD	104.607

The reconstruction error is given by

$$\|\mathbf{X}_{\text{NAOMP}} - \mathbf{X}\|_2 = \frac{1}{\sqrt{2}} k c^2 \tag{B.12}$$

which is smaller than the error of OMP. Moreover, NAOMP selects the correct atoms $\{\mathbf{d}_1, \mathbf{d}_3\}$ while OMP selects the atoms $\{\mathbf{d}_1, \mathbf{d}_4\}$.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

References

[1] Q. Yang, R. Yang, J. Davis, and D. Nistér, "Spatial-depth super resolution for range images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, Minneapolis, MN, USA, June 2007.

[2] B. Huhle, T. Schairer, P. Jenke, and W. Straßer, "Robust nonlocal denoising of colored depth data," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '08)*, pp. 1–7, Anchorage, Alaska, USA, June 2008.

[3] J. Wasza, S. Bauer, and J. Hornegger, "Real-time preprocessing for dense 3-D range imaging on the GPU: defect interpolation, bilateral temporal averaging and guided filtering," in *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV '11)*, pp. 1221–1227, Barcelona, Spain, November 2011.

[4] D. Min, J. Lu, and M. N. Do, "Depth video enhancement based on weighted mode filtering," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1176–1190, 2012.

[5] J. Fu, S. Wang, Y. Lu, S. Li, and W. Zeng, "Kinect-like depth denoising," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '12)*, pp. 512–515, Seoul, South Korea, May 2012.

[6] M. Camplani and L. Salgado, "Efficient spatio-temporal hole filling strategy for kinect depth maps," in *Three-Dimensional Image Processing (3DIP) and Applications II*, vol. 8920 of *Proceedings of SPIE*, Burlingame, Calif, USA, 2012.

[7] O. Mac Aodha, N. D. Campbell, A. Nair, and G. J. Brostow, "Patch based synthesis for single depth image super resolution," in *Proceedings of the 12th European Conference on Computer Vision (ECCV '12)*, pp. 71–84, Florence, Italy, October 2012.

[8] J. Shen and S.-C. S. Cheung, "Layer depth denoising and completion for structured-light RGB-D cameras," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 1187–1194, Portland, Ore, USA, June 2013.

[9] Y. Wang, F. Zhong, Q. Peng, and X. Qin, "Depth map enhancement based on color and depth consistency," *Visual Computer*, vol. 30, no. 10, pp. 1157–1168, 2014.

[10] J. Li, Z. Lu, G. Zeng, R. Gan, and H. Zha, "Similarity-aware patchwork assembly for depth image super-resolution," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 3374–3381, Columbus, Ohio, USA, June 2014.

[11] T.-W. Hui and K. N. Ngan, "Motion-depth: RGB-D depth map enhancement with motion and depth in complement," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 3962–3969, IEEE, Columbus, Ohio, USA, June 2014.

[12] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive autoregressive model," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3443–3458, 2014.

[13] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "Lidarboost: depth superresolution for TOF 3D shape scanning," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops '09)*, pp. 343–350, Miami, Fla, USA, June 2009.

[14] Y. Li, T. Xue, L. Sun, and J. Liu, "Joint example-based depth map super-resolution," in *Proceedings of the 13th IEEE International Conference on Multimedia and Expo (ICME '12)*, pp. 152–157, IEEE, Melbourne, Australia, July 2012.

[15] M. Kiechle, S. Hawe, and M. Kleinsteuber, "A joint intensity and depth co-sparse analysis model for depth map super-resolution," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 1545–1552, Sydney, Australia, December 2013.

[16] H. Zheng, A. Bouzerdoum, and S. L. Phung, "Depth image super-resolution using multi-dictionary sparse representation," in *Proceedings of the IEEE International Conference on Image Processing*, pp. 957–961, IEEE, Melbourne, Australia, September 2013.

[17] J. Xie, C.-C. Chou, R. Feris, and M.-T. Sun, "Single depth image super resolution and denoising via coupled dictionary learning with local constraints and shock filtering," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '14)*, pp. 1–6, Chengdu, China, July 2014.

[18] S. Lu, X. Ren, and F. Liu, "Depth enhancement via low-rank matrix completion," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 3390–3397, Columbus, Ohio, USA, June 2014.

[19] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[20] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3467–3478, 2012.

Journal of
Engineering

The Scientific
World Journal

International Journal of
Rotating
Machinery

Journal of
Sensors

International Journal of
Distributed
Sensor Networks

Advances in
Civil Engineering

Journal of
Control Science
and Engineering

Journal of
Robotics

Journal of
Electrical and Computer
Engineering

Advances in
OptoElectronics

VLSI Design

International Journal of
Navigation and
Observation

Modelling &
Simulation
in Engineering

International Journal of
Aerospace
Engineering

International Journal of
Chemical Engineering

International Journal of
Antennas and
Propagation

Active and Passive
Electronic Components

Shock and Vibration

Advances in
Acoustics and Vibration

Hindawi

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