Depth Enhancement with Improved Inpainting Order and Smoothing Method

To cite this article: Kang Yi et al 2019 J. Phys.: Conf. Ser. 1187 042065

View the article online for updates and enhancements.
Depth Enhancement with Improved Inpainting Order and Smoothing Method

Kang Yi*, Yuting Zhao, Yiyun Lei, Jian Pan
School of Information and Communication, National University of Defence Technology, Xi’an, China

*Corresponding author e-mail: penfangs@163.com

Abstract. Holes and noises are two main defects existing in Kinect’s depth images, which severely restrict the development of depth-based applications. This paper proposes a novel depth enhancement algorithm with improved propagation order and smoothing methods, based on the classic inpainting ideology. The propagation order is optimized by dividing the structural holes into separate connected regions, which is processed one by one according to definition of the pixel priority with confidence term and directional term, which is calculated with the isophote and the pseudo-normal. Furthermore, the improved smoothing method is presented in two distinctive ways: the linear combination of BF and JBF utilizes the different intensity similarity evaluation, while the joint trilateral filter exploits the properties of Gaussian kernel function, in which the proper combination of the colour similarity, depth similarity and geography distance are considered. Experimental results show that our proposed method obtains outstanding performance over other state-of-the-art methods and stage-tests also demonstrate the advantages of our proposed method over their alternatives. This work can benefit a lot for various depth-based applications, especially for the low cost Kinect’s popularity.

1. Introduction

Robot applications, such as map building, path planning and environmental perception, are becoming highly research topics recently. These tasks previously subject to images from conventional camera, however, depth sensors have made a new type of data available. In 2010, Microsoft released an imaging device called Kinect, whose lower price and relatively higher resolution wins great popularity. Although Kinect has made high-resolution real-time depth maps available at a lower cost, there are numerous noises and holes with invalid values. In fact, there are three main reasons leading to the formation of these holes. Firstly, the range is limited, which means objects nearer or farther cannot been measured correctly; secondly, some areas cannot be seen by both the projector and the camera, which is called occlusion in image processing literature; and thirdly, the material properties and reflectivity of the object also make difference. If the surface is too smooth, specular reflection may occur and if the material absorbs all light patterns projected on it by the infrared projector, the camera will capture nothing too. Both conditions will result in holes in the captured depth maps.

As described in [1], the performance of depth-based algorithms can be improved significantly if holes and noises are removed. Hence, the enhancement of depth maps which is aiming at removing noises and filling in the holes, is necessary and practically meaningful.

Researchers have paid great efforts to eliminate these defects and improve quality of depth images captured by state-of-the-art range sensors. Chen et al. [2] detected and removed the pixels with wrong
values using a region growing method, with the help of its corresponding RGB image, and filled the holes with a joint bilateral filter. They also proposed an adaptive bilateral filter to effectively reduce the noise. Yang et al. [3] presented another method and adopted an adaptive color-guided auto-regressive (AR) model, which was based on the observation that AR model tightly can fit depth maps of generic scenes for high quality depth recovery from low quality measurements. The task was formulated into a minimization of AR prediction errors subject to measurement consistency, and the predictor for each pixel was constructed according to both the local correlation in the initial depth map and the non-local similarity in the accompanied high quality color image. However, in this method, recovery of an image sized $MN \times MN$ will have a predictor matrix sized $MN \times MN$, a matrix $P$ sized $p \times MN$ ($p$, number of non-zero elements in the original image), and multiple $MN \times MN$ temporary matrices which consumed high computing and storage. Liu et al. [4] formulated the guided depth enhancement problem based on the heat diffusion framework, referred to as GAD. Meanwhile, Le et al. [5] proposed an adaptive directional filters to fill in the holes and suppress the noises in depth maps, whose main contribution lied in the window shapes adaptively adjusted based on the edge direction of the color image.

In our proposed method, conventional inpainting method, which is frequently used in optical images, is adopted for holes filling because of their similarity to a certain degree. We extend the traditional inpainting methods by optimizing its propagation order and introducing the exemplar-based filling method. Filtering with higher color similarity is further applied for noises and artifacts removal.

1.1 Conventional Inpainting Techniques

Inpainting, modification of images in a way that is non-detectable for an observer, can be dated back to the Renaissance, when medieval artwork started to be restored by filling in gaps [6, 7]. Aim of inpainting is to reconstitute the missing or damaged portions of the work, i.e. repairing unknown regions with reference of known regions in an image [8].

The key problems are how to propagate and how to inpaint, namely the inpainting method and its order. Telea A. [9] made a distance transform before inpainting instead of iterative confidence computation, which is based on the FMM (Fast Marching Method). For a region in the image to be inpainted, the algorithm starts from the boundary and goes inside gradually. More weightage is given to those pixels lying near to the point, near to the normal of the boundary and those lying on the boundary contours. Once a pixel is inpainted, it moves to next nearest pixel using FMM. FMM can ensure those pixels near the known are inpainted first, so that it just works like a manual heuristic operation. Another method [10], based on fluid dynamics, utilized partial differential equations, traveled along the edges (meant to be continuous) from known regions to unknown regions and continued isophotes while matching gradient vectors at the boundary of the inpainting region.

Inpainting has been applied to many practical applications with some improvements. Criminisi et al. [11] removed the object and filled in the hole left behind in a visually plausible way, by combining the advantage of texture synthesis and inpainting, which replicated texture and structure information simultaneously and improved computational efficiency by a block-based sampling process. Liu et al. [12] extended the original fast marching method, whose results were limited to only using the information itself, by incorporating another image as guidance to reconstruct the unknown regions or damaged portions. We hence refer to it as a guided FMM, quoted with GFMM later. Furthermore, Gong et al. [13] proposed a new inpainting approach based on the FMM, by extending the inpainting model and the propagation strategy of FMM to incorporate color information for depth inpainting. Zhang et al. [14], inspired by the above-mentioned methods [11, 12], added a level set distance to improve the propagation order and proposed an edge-preserving filtering method to further remove noises and artifacts, which is referred to as ImprovedInpainting later.

1.2 Conventional Inpainting Techniques Noise-removing Strategies

Noise-removing or smoothing has been a much-talked-about topic in image processing literature. Various methods have been proposed and applied to different practical applications.
Simple linear translation-invariant (LTI) filters with explicit kernels, such as the mean, Gaussian, Laplacian and Sobel filters [15] have been widely used in noise-removing community. Low-pass filtering may be the most obvious method, like box filter [16] (also called mean filter), in which the value of \( i_{\text{th}} \) pixel is replaced by the weighted mean of pixels in the window centered at \( i_{\text{th}} \) pixel. According to the weighting way, smoothing methods are divided into two categories: isotropic filtering and anisotropic diffusion. Isotropic filtering methods adopt a uniform attitude to all elements, the regions and the edges (including texture and details), which may lead to blurring.

Otherwise, many anisotropic diffusion methods are proposed, like the bilateral filter [17], joint bilateral filter [18], guided filter [19] and so on. They treat regions and edges respectively to preserve details as much as possible and avoid blurring. These efforts have made some difference in the image smoothing community.

2. Material and Methods

In order to remove noises and fill in holes existing in Kinect’s captured depth images in a more visually plausible way, thus accelerating development of the depth-based applications, a novel depth enhancement method based on the conventional inpainting method with improved propagation order and smoothing method is proposed in this paper. Figure 1 depicts its framework, which is composed of four stages: the pre-processing, selection of the holes with the highest priority, filling in the holes with Exemplar-based inpainting and smoothing the noises and artifacts caused by inpainting. Holes selection is completed in two steps: drawing the scope of pixels taken into account and then selecting a pixel among them in a certain way. Detailed description of each stage is presented below.

![Figure 1. The Framework of the proposed method.](image)

2.1 Pre-processing

As described in Section 1, holes are inevitable in Kinect’s captured depth images. Microsoft researchers are also quite clear of this, hence they mark pixels with invalid values using the tag “No-available”, which is transferred to zero when the maps are rendered as 8-bit images [20, 21]. Thus, ‘zero’ does not mean the zero depth, but the absence of a value. Therefore, the holes-removing goal is to replace the pixels value zero with a reasonable value and produce a more visually plausible depth map.

Holes on the captured depth maps can be divided into stochastic holes and structural holes. Stochastic holes consist of discrete single pixels with absent values, while structural holes are made up of regions of holes. As shown in Figure 1, the Kinect’s captured depth maps are firstly pre-processed to remove the stochastic holes which may severely influence the inpainting result and the running speed of subsequent processing.

Generally, the mathematical morphological closing operation with structure element size of one pixel is applied to obtain outlier-removed depth maps. Morphological closing operation completed with expansion followed by contraction is efficient to remove the outliers.

Images processed with the above-mentioned method look like Figure 2, in which stochastic holes are removed while remaining structural holes appear black.
2.2 Propagation and Inpainting
Our holes-filling procedure is highly inspired by the object-removal method in [11], which aims at removing the target object and filling in the hole left behind in a visually plausible way. It shares similar challenge on holes filling with the depth enhancement. Exemplar-based texture synthesis contains the essential process required to replicate both texture and structure; the success of structure propagation, however, is highly dependent on the order in which the filling proceeds.

2.2.1 Propagation. As described in Figure 1, selection of the hole to be dealt with is completed in two steps: drawing the scope of pixels to consider of and selecting a hole from the candidate pixels in a certain way. Different methods can be employed to fulfill these tasks.

**Drawing the scope of pixels to take into account**
Holes in the pre-processed images are comprised by several connected holes regions, shown as Figure 2. In our proposed method, holes in one connected holes region are filled in after another, whose effectiveness has been proved in Section 3.1.1. Due to the fact that a pixel only has relationships with its neighboring known pixels, two separate connected holes regions have little impact on each other. Hence, connected holes regions are processed according to their geometry location, for example, from left-top corner to right-bottom of the image.

To make sure all the connected regions are selected in closed areas and the subsequent processing more easily, the image resulted from pre-processing are expanded with 1 pixel on both sides. For example, the original depth map is denoted with A in matrix form, then the expansion result looks like

\[
\begin{bmatrix}
1 & 1 & 1 \\
1 & A & 1 \\
1 & 1 & 1 \\
\end{bmatrix}
\]

As shown in Figure 2, connected holes regions in pre-processed Art are marked with green closed curves. Image on the left is obtained before expansion and the right-hand one is got after expansion. It’s clear that all the holes in Figure 2 (b) is circled in closed green curves, while some holes in Figure 2 (a) is surrounded by unclosed red curves.

![Figure 2. Connected holes regions in pre-processed Art marked with green closed curves.](image-url)

**Picking out the target pixel**
Before inpainting, another problem arises after one connected holes region specified, namely the filling order of pixels in the selected region. For ease of description, we adopted notions similar to that used in inpainting literature. The regions of holes, whose values are absent, are denoted as the target regions. Correspondingly, the regions with known values, which will remain fixed and provide reference in the filling process, are defined as source regions.

In our proposed method, each target patch, window size 9*9 centered at a selected pixel, maintains a pixel priority determined by the priority evaluation function. Two components are combined to model the pixel priority evaluation function, which are the confidence level and a direction term respectively.
The confidence level gives out the reliability of information in the target patch, and is calculated with recursion. The original known pixels, namely pixels in the source regions, are given 100% trust, i.e. the conference value is 1. And the value decreases as we go further into the target regions.

As the same time, the direction term is given by inner product of two vectors, whose directions are respectively the same as the isophote’s direction and the gradient of a pseudo-normal. Isophote is defined as the line of the same gray level and is computed by gradient with 90° rotation, for the reason that the gradient is thought to be the fastest changing direction. Meanwhile, pseudo-normal is similar but different to the normal. Directions of lines and textures are not what we care about and noises may make things more complex, therefore, computing the normal direction is not a good choice. What we rarely concern is the direction from the source region pointing into the target region. Therefore, a counterpart of the original depth image, in which pixels located at the corresponding of holes in the original depth image value 0 while others value 1, is created. The pseudo-normal direction is calculated by gradient of the just-created counterpart.

The pixel priority evaluation function is defined as $P(p)$ in Equation (1).

$$P(p) = C(p) \ast \text{Dir}(p)$$

(1)

Where $C(p)$ defines confidence level and $\text{Dir}(p)$ denotes direction term respectively, which are defined in Equation (2, 3).

$$C(p) = \frac{\sum_{q \in P_t \cap \Psi} C(q)}{|P_t|}$$

(2)

$$\text{Dir}(p) = \frac{\nabla I_p \cdot n_p}{\alpha}$$

(3)

Where $P_t$ is the target patch centered at pixel $p$, and $|P_t|$ is its area. As shown in the Fig 3, $P_t$ is presented with a red rectangular frame with central pixel $p$. $\Psi$ defines the source regions, which means that $q$ belongs to the intersection of $P_t$ and $\Psi$. $\nabla I_p$ represents the isophote direction of pixel $p$ and $n_p$ is its pseudo-normal direction, which is shown in Fig3 with arrow lines. Their inner product gives priority to the pixels whose isophote and pseudo-normal go in the similar direction, namely, when they go in the same direction, the direction term gains its maximal value; otherwise, the value decreases as the angle of isophote and pseudo-normal increases till orthogonal to each other. When the image is 8-bit, $\alpha$ is set 255 to ensure the value of $\text{Dir}(p)$ range between $[0,1]$.

![Figure 3. Notations used in definition of priority evaluation function.](image)

2.2.2 Inpainting. The target patch centered at the pixel picked out is filled in with the exemplar-based inpainting method, which is fulfilled in two steps. Firstly, the target patch is used as a sliding window which scans through the whole source regions, to find out its best-match patch. The similarity is measured with the value difference of known pixels in both patches. Then, values of holes in the target patch are replaced with values in the corresponding place of the best-match patch.
The exemplar-based Inpainting method propagates the texture and structure simultaneously and obtains great performance.

2.3 Smoothing

Smoothing is indispensable after holes filled in due to noises existing in the original images and artifacts caused by exemplar-based inpainting algorithm. As described in [22], noises in Kinect’s depth maps are a deterministic function of distance added to the random noises present in all systems, modeled as the deterministic noises and random noises. Although the exemplar-based inpainting propagates texture and structure information well, it also brings about artifacts and bumpiness on the surface.

As described in [23, 24], Kinect measures the distance between object and the sensor plane. It’s a distance between a set of points and a specified plane, which means that the depth values vary with the geometrical shape of the object. As we all known, any object in the real world varies continuously with no interruption. Therefore, despite the holes, the variation trend is known to us, which is very useful for our depth enhancement.

Thus, the depth value of each pixel in the depth maps has relationship with its neighbors because of the continuity, and only with its neighbors. Therefore, in our proposed method, Gaussian kernel is adopted and non-isotropic-filtering is used to ensure that more details are preserved.

Moreover, pixels are affected by its neighbors, but in structural holes regions, depth information around a hole may be not quite believable. However, information in the corresponding color image may be more believable. Hence, higher weightage is given to the color similarity.

In our proposed method, the smoothing is modeled in two ways, which can obtain similar performance, as shown in Section 3.2.2. Their goal is accordant, though appear different, to combine the influences of three items and enlarge the weight of color similarity at the same time. The first model, defined as Eq.4, is given by a linear combination of the classical bilateral filter [17] and a joint bilateral filter [18]. Meanwhile, the second smoothing model, denoted as Eq.5, is referred to as the joint trilateral filtering.

\[
JBF_BF[I]_p = \frac{1}{\omega_p} \sum_{q \in S} G_{\sigma_d} \left( \left| p - q \right| \right) \left( \alpha G_{\sigma_c} \left( I_p - I_q \right) + \beta \cdot G_{\sigma_d} \left( I_{\phi_p} - I_{\phi_q} \right) \right) I_q
\]  

\[
JiontTF[I]_p = \frac{1}{\omega_p} \sum_{q \in S} G_{\sigma_d} \left( \left| p - q \right| \right) \cdot G_{\sigma_c} \left( I_{\phi p} - I_{\phi q} \right) \cdot G_{\sigma_d} \left( I_{d p} - I_{d q} \right) \cdot I_q
\]

Where, \( G_{\sigma_d} \left( \left| p - q \right| \right) \), \( G_{\sigma_c} \left( I_{\phi p} - I_{\phi q} \right) \) and \( G_{\sigma_d} \left( I_{d p} - I_{d q} \right) \) respectively define the weights of distance, color similarity and depth similarity on the central element \( p \), which are defined as Equation (6, 7, 8) respectively. They are all computed between \( p \) and \( q \), which belongs to \( p \)'s neighborhood \( S \).

And \( \omega_p \) is the normalization term, which is used to make sure that range of \( I \) is constant.

\[
G_{\sigma_d} \left( \left| p - q \right| \right) = \exp \left( -\frac{\left| p - q \right|^2}{2\sigma_d^2} \right)
\]

\[
G_{\sigma_c} \left( I_{\phi p} - I_{\phi q} \right) = \exp \left( -\frac{\left| I_{\phi p} - I_{\phi q} \right|^2}{2\sigma_c^2} \right)
\]

\[
G_{\sigma_d} \left( I_{d p} - I_{d q} \right) = \exp \left( -\frac{\left| I_{d p} - I_{d q} \right|^2}{2\sigma_d^2} \right)
\]
3. Results And Discussion
We experimented with our depth completion and denoising method on Middlebury stereo dataset[25]. The same as [26], Zero-Mean-White-Gaussian noise with standard deviation 25 and 5 are added to the color and depth images respectively, and around 13% pixels with unknown depth values are added to the original Middlebury stereo dataset.

For ease of comparison, all the experiments are conducted on the 30 images in [26]. One typical example among them is the Aloe, presented in Figure 4. And the control variables are used to demonstrate the advantage of each step with its possible alternatives.

![Example of the depth map and its aligned colour image, the Aloe from the Middlebury Dataset, to be enhanced.](image)

Figure 4. Example of the depth map and its aligned colour image, the Aloe from the Middlebury Dataset, to be enhanced.

![Results obtained with exemplar-based inpainting with selection method ScopeSelectionMethod1 or ScopeSelectionMethod2, Baby2, Baby3 and Bowling1 are given in row (a) and (b).](image)

Figure 5. Results obtained with exemplar-based inpainting with selection method ScopeSelectionMethod1 or ScopeSelectionMethod2, Baby2, Baby3 and Bowling1 are given in row (a) and (b).

3.1 Comparison of Propagation methods
As illustrated in Figure 1, the second stage, selecting the hole to be dealt with, is accomplished by two steps: drawing the scope of pixels taken into account and then selecting a pixel among them in a certain way. Possible ways to draw the scopes and select a pixel among them are discussed in the next two sections.
3.1.1 Drawing the scope of pixels to take into account. There are two ways to draw the scope of pixels to take into account: taking all the holes in the pre-processed images into account (referred to as ScopeSelectionMethod1) or dealing with connected holes regions one by one (quoted with ScopeSelectionMethod2). Results of the two scope selection methods are evaluated quantitatively and qualitatively.

As shown in Figure 5, sample results after inpainting, referred to as Baby2, Baby3, and Bowling1, obtained by ScopeSelectionMethod1 lie on the left column and ScopeSelectionMethod2 on the right. Both of them adopt the exemplar-based inpainting. Red rectangles in Figure 5 point out the distinct advantage of ScopeSelectionMethod2 over ScopeSelectionMethod1. It’s obvious that the second method works better in structure-preserving and brings in fewer artifacts, especially along the edges.

Given the influence of these two selection methods on the final results, the same subsequent processing methods are adopted. The quantitative results with PSNR and SSIM are presented in Table 1, which obviously tells that ScopeSelectionMethod2 performs better than the other, with 0.3dB higher in PSNR and 0.02 higher in SSIM.

Table 1. Quantization results of average PSNR and SSIM on the testing images.

| ScopeSelection-Method1 | ScopeSelection-Method2 |
|------------------------|------------------------|
| PSNR                   | 27.05973               |
| SSIM                   | 0.92596                |
|                        | 27.35976               |
|                        | 0.92777                |

3.1.2 Picking out the target pixel. As described in Section 3.1.1, dealing with pixels in one connected holes region after another works better than taking all the holes into account at once.

Once one connected holes region, referred to as the candidate set, is picked out, there are three methods to determine the order of pixels to be inpainted. Firstly, the pixels can be selected one by one without priority to anyone, namely, all the pixels share the same priority. Secondly, each pixel maintains a priority, determining its subsequent inpainting order, computed by the confidence. Thirdly, the pixel priority is calculated by confidence together with a directional term, as defined in Equation (1).

The variable here is the way selecting some pixels from the specified scope, namely, other procedures are all the same. Comparison results are presented in Table2, which verifies that priority evaluated with confidence and directional term performs relatively better, about 0.09dB and 0.15dB higher in PSNR and 0.00078 and 0.00034 higher in SSIM than the first and second methods.

Table 2. Quantization results of average PSNR and SSIM on the testing images with different pixel priority evaluation methods.

| Calculation of priority | no priority | confidence only | confidence and directional term |
|-------------------------|-------------|-----------------|---------------------------------|
| PSNR                    | 27.26929    | 27.20078        | 27.35976                        |
| SSIM                    | 0.92699     | 0.92743         | 0.92777                         |

Taking pixels in connected holes regions into account one by one and defining pixel priority with confidence and directional term, adopted in our proposed method, demonstrate the advantages of themselves over their possible alternatives. And all the following experiments are based on the above-proposed pixel selection method combining confidence and directional term.

3.2 Comparison of Smoothing Strategies

As described in Section 2.3, smoothing is an essential procedure after holes filled in Kinect’s depth maps. Bilateral filter, referred to as BF, and joint bilateral filter, quoted with JBF, have achieved great popularity in intensity image noise-removing community. But depth enhancement has its own features, invalid pixels in depth maps lead to the BF’s inaccuracy and incomplete correspondence of color and depth values disables JBF too. Hence, both methods fail to be suitable perfectly, even though still work to some extent.

In our proposed method, depth similarity and its aligned image’s color similarity are both taken into account. Methods to combine the depth similarity, color similarity and geography distance
together could be the linear combination of BF and JBF or the joint trilateral filter, defined in Equation (4, 5). As discussed before, in order to suppress the error’s accumulation, higher priority is given to the color similarity.

### 3.2.1 Linear combination of BF and JBF

The method presented in Equation (4), uses the coefficients $\alpha$ and $\beta$ balancing the influence of BF and JBF on the final results. As is known to all, the weighting factor in BF is determined by intensity similarity of itself and the geography distance, while in JBF, it’s decided by its guidance’s intensity similarity and its own geography distance. Correspondingly, in depth enhancement, the image itself is the depth map and the guidance is generally its aligned color image. Therefore, it can be concluded that BF measures the depth similarity, while the geography distance and JBF measures the color similarity and the geography distance in depth enhancement community.

Therefore, if we want to enlarge the influence of color similarity on the final result, only need to make sure that $\alpha$ is bigger than $\beta$. Experiments are carried on to compare the results of different ratios, which is presented in Table 3.

Other parameters are set as below. Window size is $7 \times 7$, and sigma of distance is 3 while sigmas of color and depth similarity are both set 0.1.

**Table 3. Quantization results of average PSNR and SSIM on the testing images with different ratios of BF and JBF.**

| $\alpha : \beta$ | 1:1 | 2:1 | 3:1 | 4:1 | 5:1 | 6:1 | 7:1 | 8:1 |
|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|
| PSNR            | 27.4058 | 27.4221 | 27.4254 | 27.4244 | 27.4214 | 27.4179 | 27.4140 | 27.4100 |
| SSIM            | 0.9305 | 0.9305 | 0.9302 | 0.9299 | 0.9295 | 0.9292 | 0.9289 | 0.9285 |

More intuitive results are presented in Figure 6, from which we can clearly conclude that when $\alpha : \beta$ equals 3:1, the results gain the highest average PSNR, while SSIM achieves its maximum at 1 and 2, that’s to say, the value of SSIM decreases when color similarity’s proportion increases while PSNR enlarges when color similarity’s proportion increases.

![Figure 6](image)

(a)PSNR values with ratio of varies from 1:1 to 8:1 (b) SSIM values with ratio of varies from 1:1 to 8:1

**Figure 6.** Line Charts of PSNR and SSIM values obtained with different ratios of BF and JBF.

### 3.2.2 Joint trilateral filter

Another smoothing method is modeled as Equation (5), referred to as the joint trilateral filter. It combined the color similarity, depth similarity and geography distance with multiplication. All the weight factors are calculated with the Gaussian function, which is defined as Eq.9.

$$f(x) = ae^{-\frac{(x-b)^2}{2\sigma^2}}$$

Where, $f(x)$ defines the Gaussian value at $x$. And $a$ is its peak value, $b$ is the average value, namely, if $x = b$, then $f(x) = a$. All the Gaussian functions used in smoothing kernel share the same average value, i.e. $b = 0$ and the same peak value, i.e. $a = 1$. 
Figure 7. Gaussian curves with different sigmas while average value is 0 and peak value is 1.

As shown in Figure 7, Gaussian curves with the same peak value 1 and average value 0, vary with different $\sigma$ values. Along with Sigma varying from 0.5 to 4, the curves widen step by step. For example, the Gaussian function values become larger when sigma becomes bigger when mean value is 1.

Therefore, in order to enlarge the influence of color similarity, magnifying the sigma value may take effect. The sigma in color Gaussian function is set $n$ times to depth’s, namely if sigma of depth 0.1, then sigma of color $n\times0.1$. Besides, sigma of geography distance is set 3, the same as in Section 3.2.1. Experimental results are presented in Table 4. It can be told directly from the table 4 that when $n=2$, PSNR gains its highest value 27.4016, while 0.9319 is the best SSIM value at $n=3$.

Table 4. Quantization results of average PSNR and SSIM on the testing images with different times of color sigma and depth sigma.

| n   | 1    | 2    | 3    | 4    | 5    |
|-----|------|------|------|------|------|
| PSNR| 27.3568 | 27.4016 | 27.3920 | 27.3807 | 27.3721 |
| SSIM| 0.9236  | 0.9317 | 0.9319 | 0.9315 | 0.9312 |

With the same testing data and other former treatment, four smoothing methods, bilateral filter, joint bilateral filter, their linear combination ($\alpha: \beta = 3:1$ is adopted) and the joint trilateral filter ($n=3$ is adopted), obtain different results respectively. Their quantitative results with PSNR and SSIM are all presented in Table 5.

Table 5. Quantization results of average PSNR and SSIM on the testing images with different smoothing methods.

|                  | BF    | JBF   | Linear Combination(3:1) | Joint Trilateral Filter(3) |
|------------------|-------|-------|-------------------------|----------------------------|
| PSNR             | 27.3462 | 27.3189 | 27.4255                 | 27.3920                    |
| SSIM             | 0.9298  | 0.9209 | 0.9302                  | 0.9319                     |
| parameters       | $\sigma_r = 3$, $\sigma_r = 3$, $\sigma_v = 0.1$, $\sigma_v = 0.1$ | $\sigma_v = 3$, $\sigma_v = 3$, $\sigma_v = 0.1$, $\sigma_v = 0.3$ | $\sigma_v = 0.1$, $\sigma_v = 0.3$, $\sigma_v = 0.1$ |

Figure 8 shows the tendencies of the average PSNR and SSIM on the testing images of the four methods in a more visible way. The proposed method with the two proposed smoothing methods achieve close quality, both better than the conventional BF or JBF on PSNR and SSIM, while BF gains better results than JBF.
3.3 Comparative experiments with other enhancement algorithms

To better demonstrate the superiority of our proposed method, experiments are carried on to compare results with three other state-of-the-art depth enhancement methods, the GAD, GFMM and ImprovedInpainting proposed in [4, 12, 14].

All the three methods share a similar procedure with our proposed method, propagating information from known depth values to unknown regions. In GAD, this is completed with the heat diffusion framework, in which the known depth values are treated as the heat sources. GFMM fulfill this task based on the fast marching method, incorporated with its guidance image. And the ImprovedInpainting defines the priority with the confidence, data term and level set distance factor, and fills in holes with the exemplar-based inpainting method, followed by a certain filter to remove noises.

In our proposed method, similar processing flow is adopted, filling in the holes outside-in. Quantitative results on the testing images with the other three enhancement methods are presented in Table 6, which reveals that our proposed method holds an overwhelming superiority over the other three cutting-edge methods.

|                     | GAD     | GFMM   | ImprovedInpainting | Proposed Method (n=3) | Proposed Method (3:1) |
|---------------------|---------|--------|---------------------|-----------------------|-----------------------|
| PSNR                | 25.4332 | 25.6605| 26.6498             | 27.4255               | 27.3920               |
| SSIM                | 0.8423  | 0.6974 | 0.9084              | 0.9302                | 0.9319                |

Figure 9, quantization results of PSNR and SSIM on the testing images, further demonstrate the outstanding performance of our proposed method. Meanwhile, the two smoothing strategies in our proposed method share similar results, which are reflected in curves of them almost totally overlapping each other in the line charts. And Figure 10 presents the corresponding qualitative results.

(a)SSIM of the testing images on four methods. (b) PSNR of the testing images on four methods.

Figure 9. Quantitative results of PSNR and SSIM on the testing images of GAD, GFMM, ImprovedInpainting and our proposed method.
Figure 10. Qualitative Comparisons of different algorithms tested on the Aloe, Bowling2, Lampshade1, Monopoly and Teddy are presented above. From left to right, (a) ~ (g) represent the Original Depth images, Corresponding GroundTruth Maps, and results achieved by GAD, GFMM, ImprovedInpainting and our proposed method. (f) shows the results obtained by our proposed method with the linear combination of BF and JBF, while (g) is achieved by the proposed method with joint trilateral filter.

4. Conclusion
In this paper, a novel depth enhancement method for depth maps captured by Kinect is proposed. Our main contributions lie in the selection of the pixel to be filled in and the two smoothing method. Firstly, structural holes in the original depth maps are divided into separate connected holes regions, which are processed one by one. Otherwise, pixels in each connected holes region are filled in based on the priority, determined by the confidence level and directional term, while the directional term is defined with the inner product of isophote and the pseudo-normal. Secondly, two smoothing methods are proposed to enlarge the color similarity, the linear combination of BF and JBF and joint bilateral filter. To enlarge the color similarity, different coefficients are tested and the control variables are used to demonstrate their advantage.

The experimental results demonstrate that our proposed method performs much better than other three state-of-the-art methods, which share similar processing idea. Moreover, experiments on pixel selection present that the method used in our proposed method outperforms its alternatives. And smoothing methods, linear combination of BF and JBF and joint trilateral filter, obtains similar quantitative results with proper parameters, both better than the results of single BF or JBF.

Acknowledgments
This work is partially supported by National Natural Science Foundation of China (Grant No. 61305109, No. 61401324), 863 Program (No. 2013AA014601) and Shaanxi Scientific research plan (No. 2014K07-11).

 References
[1] Chen L, Wei H, Ferryman J 2013 Pattern Recognition Letters 34 1995-2006
[2] Chen L, Hui L, Shutao L 2012 21st International Conference on Pattern Recognition pp 3070-73
[3] Yang J, Xinchen Y, Kun L, Chunping H and Yao W 2014 IEEE transactions on image processing 23 3443-58
[4] Liu J, Gong X 2013 Advances in Multimedia Information Processing – PCM 2013 Springer International Publishing pp 408-417
[5] Le A V, Jung S W, Won C S 2014 Sensors 14 11362-78.
[6] Langmuir E 1985 Journal of the Royal Society of Arts 133 5351
[7] G. Emile-Male 1976 The Restorer’s Handbook of Easel Painting (New York: Van Nostrand Reinhold)
[8] Bertalmio, Marcelo, Guilleremo S, Vincent C and Coloma Br 2005 Siggraph 4 417-24
[9] Alexandru T 2004 Journal of Graphics Tools 9 23-34
[10] Bertalmio M, Bertozzi A L, Sapiro G 2001 Proceedings of the 2001 IEEE Computer Society Conference on, vol 1, pp 355-362
[11] Antonio C, Patrick P, and Kentaro T 2004 IEEE Transactions on image processing 13 1200-12
[12] Liu J, Gong X, Liu J 2012 In Pattern Recognition (ICPR), 21st International Conference on IEEE pp 2055-58
[13] Xiaojin G, Junyi, Wenhui Z and Jilin L 2013 Image and Vision Computing 31 695-703
[14] Zhang L, Shen P, Zhang S, Juan S, and Guangming Z. 2016 IEEE International Conference on Image Processing IEEE pp 4102-06.
[15] R Gonzalez, R Woods 2002 Digital Image Processing, 2nd Edition (Englewood:Prentice Hall)
[16] Medonell M J 1981 Computer Graphics & Image Processing 17 65-70
[17] Tomasi C, Roberto M 1998 In Computer Vision, 1998. Sixth International Conference on IEEE pp 839-46
[18] Petschnigg G, Szeliski R, Agrawala M, Cohen M, Hoppe H, Toyama K. 2004 ACM Transactions on Graphics (TOG) 23 664-72.
[19] He K, Sun J, Tang X 2013 IEEE transactions on pattern analysis and machine intelligence 35 1397-1409
[20] Khoshehlaam K, ISPRS 2011 ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 3812 133-38.
[21] Mallick T, Das P P, Majumdar A K 2014 Sensors Journal IEEE 14 1731-40.
[22] E. Bryan, B P. Velocity Estimation Using an Rgb-D Sensor[EB/OL]. http://www.et.byu.edu/~bmazzeo/ECEn_670_F11/mini_conference_files/Group_9_paper.pdf
[23] “Depth Camera”, Microsoft Robotics.User Guide.Robotics Common and Devices, Microsoft Corporation, 2012, https://msdn.microsoft.com/en-us/library/hh438997.aspx.
[24] Depth value of the Kinect and OpenNI, 2011, https://kheresy.wordpress.com/2011/12/27/depth_value_of_kinect_in_openni/.
[25] MiddleburyStereo Dataset: http://vision.middlebury.edu/stereo/data/
[26] Si Lu, Xiaofeng Ren, Feng Liu, Depth Enhancement via Low-rank Matrix Completion(2014), http://web.cecs.pdx.edu/~fliu/project/depth-enhance/.