Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

# Depth-aided inpainting for disocclusion restoration of multi-view images using depth-image-based rendering*

Kai LUO†, Dong-xiao LI†‡, Ya-mei FENG, Ming ZHANG

(*Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China*)

†E-mail: luokai82@gmail.com; lidx@zju.edu.cn

Received Nov. 20, 2008; Revision accepted Jan. 4, 2009; Crosschecked Oct. 18, 2009

**Abstract:** A new algorithm is proposed for restoring disocclusion regions in depth-image-based rendering (DIBR) warped images. Current solutions include layered depth image (LDI), pre-filtering methods, and post-processing methods. The LDI is complicated, and pre-filtering of depth images causes noticeable geometrical distortions in cases of large baseline warping. This paper presents a depth-aided inpainting method which inherits merits from Criminisi's inpainting algorithm. The proposed method features incorporation of a depth cue into texture estimation. The algorithm efficiently handles depth ambiguity by penalizing larger Lagrange multipliers of filling points closer to the warping position compared with the surrounding existing points. We perform morphological operations on depth images to accelerate the algorithm convergence, and adopt a luma-first strategy to adapt to various color sampling formats. Experiments on test multi-view sequence showed that our method has superiority in depth differentiation and geometrical loyalty in the restoration of warped images. Also, peak signal-to-noise ratio (PSNR) statistics on non-hole regions and whole image comparisons both compare favorably to those obtained by state of the art techniques.

**Key words:** Depth-aided inpainting, Disocclusion restoration, Depth-image-based rendering (DIBR), Image warping, Stereoscopic image, Multi-view image, 3D-TV

**doi:**10.1631/jzus.A0820806 **Document code:** A **CLC number:** TP391.41

## INTRODUCTION

The next emerging revolution in the high definition video world will be multi-view applications such as 3D-TV (Fehn *et al*., 2006; Akar *et al*., 2007; Smolic *et al*., 2007). Techniques for multi-view compression and transmission can be classified into two categories. The first is based on the MPEG (moving picture experts group) families. The current multi-view video coding (MVC) standard (Flierl *et al*., 2007; Merkle *et al*., 2007) exploits the ability of H.264 codec in the spatial correlations of neighboring views and transmits an equal number of multi-channel views to the end terminal. The main drawback of this approach is its bandwidth requirement. The second category is based on depth-image-based rendering (DIBR) (Fehn *et al*., 2002; Fehn, 2004; Kauff *et al*., 2007). Given one view with its depth information, the technique theoretically has the ability to render views (termed 'warping' in computer graphics) in any other spatial positions. Thus in the recent years DIBR has received much attention in the video communication research community.

McMillan (1997) derived a warping equation that maps the visible points in the reference image to their correct positions in any desired view. Fehn (2004) developed this idea and built up an experimental 3D-TV system based on DIBR. It codes one center view with its depth map and symmetrically warps two neighboring views on the decoder side. However, DIBR technology has intrinsic limitations. New areas would appear in the warped view other than the reference view. Since newly exposed areas (termed 'disocclusions' or 'holes') do not have correspondence to the reference view, their texture and

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (No. 60802013) and the Natural Science Foundation of Zhejiang Province, China (No. Y106574)

depth attributes are uncertain. As the distance between the warped view and the reference view increases, the disocclusion becomes bigger.

Disocclusion is caused by insufficient points information in the reference image and is manifest as a discontinuity in depth image. There are a number of methods for dealing with the disocclusion problem. We classify them into three categories.

The first, termed 'layered-depth-image' (LDI) (Shade *et al*., 1998; Yoon and Ho, 2007), aims to provide sufficient points information during rendering. Therefore, it stores scenario information not only of the visible surface points, but also of the invisible points. The method can generate high quality warped images but the acquisition and transmission of the layered images can be complicated.

The second category, we term 'pre-processing approaches'. Since discontinuity in depth image gives rise to disocclusion in the warped image, smoothing these discontinuities can reduce the disocclusion.

Fehn (2004) used a 2D Gaussian filter to smooth the center view's depth image. Through subjective evaluation, Zhang and Tam (2005) proposed the idea of adopting an asymmetrical 2D Gaussian filter instead of a symmetrical one. By differentiating the standard deviation and window size of the Gaussian filter in horizontal and vertical directions, warped images with fewer holes could be achieved. Chen *et al*.(2005) found that after whole depth image filtering, the warped image showed visual distortion. They ameliorated the filtering method by employing a specifically designed edge-dependent Gaussian filter. Inspired by this idea, Daribo *et al*.(2007) proposed a distance-dependent filter method to smooth sharp depth changes near object boundaries. In contrast to the Gaussian-filter-based method above, Wang *et al*.(2007) segmented the depth image into several layers and changed depth values point by point by limiting the object's displacement to 1 or 2 pixels.

By changing the point intensity of the depth image, a pre-processing approach reduces or even eliminates disocclusion regions. However, changes of depth values bring the problem of object geometrical distortion in the warped image.

The third approach, we term 'post-processing'. This approach does not alter the depth image before warping. It processes the warped image to restore the disocclusion through image processing techniques.

Vázquez *et al*.(2006) assessed the subjective quality of warped images using different post-processing techniques. The simplest method is average filtering. An unmapped point's intensity is assigned the averaged intensity of its surrounding points. Considering that the disocclusion usually lies in object boundaries in the horizontal, interpolation or extrapolation in the horizontal is also feasible. Due to their simplicity, the textures and structures in a disocclusion cannot be retained and they blur the processed region, a fault perceivable to human eyes.

Tauber *et al*.(2007) argued that image based rendering (IBR) should be combined with inpainting methodology. Inpainting is a family of post-processing techniques which is used to restore the specks and scratches in damaged paintings. As there is a natural similarity between damaged holes in paintings and disocclusions in DIBR warped images, inpainting has the potential to exhibit its power in disocclusion restoration. Also, satisfactory results have been observed by Lee and Kim (2007) and Wang *et al*.(2008).

This paper proposes a depth-aided inpainting algorithm for disocclusion restoration based on Criminisi's algorithm (Criminisi *et al*., 2004). According to the characteristics of DIBR process, we perform morphological operations on depth images before warping. During the disocclusion restoration, we adopt the luma-first strategy and handle the depth ambiguities on the object boundary by combining depth information into a cost function. We also retain a visually better structural connectivity in the disocclusion interior by revising the point priority computation. The proposed algorithm effectively restores the disocclusion by filling in photo-realistic points with depth levels similar to surrounding points. This paper gives a detailed description of the method and includes comparisons with state of the art methods.

## RELATED BACKGROUND

This section introduces the DIBR equation and Criminisi's algorithm. The first subsection explains the warping equation, and then presents the warped image with necessary re-sampling and visibility processing. The second subsection provides the notation diagram and introduces the procedure of Criminisi's algorithm.

**Depth-based-image rendering**

DIBR is the process of synthesizing 'virtual' views of a scene from reference images and associated per-pixel depth. The warping process can be conceptually decomposed into two steps including first projecting the reference image into the 3D world, then projecting the spatial scene onto the desired image plane (McMillan, 1997).

The perspective projection equation (Fehn, 2004) is

$$\boldsymbol{m}\cong\boldsymbol{A}\boldsymbol{P}_n\boldsymbol{D}\boldsymbol{M}. \tag{1}$$

The equation indicates that the point's coordinate $\boldsymbol{m}$ in the camera plane is equal to the vector product of camera intrinsic matrix $\boldsymbol{A}$, identity matrix $\boldsymbol{P}_n$ (Xu and Zhang, 1996), extrinsic matrix $\boldsymbol{D}$ and the point's coordinate $\boldsymbol{M}$ in the 3D world. The symbol '≅' denotes equality up to a non-zero scale factor (Hartley and Zisserman, 2000).

For illustration purposes, we write a 4×4 projection matrix

$$\boldsymbol{P}=\boldsymbol{A}\boldsymbol{P}_n\boldsymbol{D}, \tag{2}$$

which is computed from camera parameters.

Substituting $\boldsymbol{P}$ into Eq.(1), we have

$$\begin{pmatrix} w\cdot u_0 \\ w\cdot v_0 \\ w \\ 1 \end{pmatrix} = \begin{pmatrix} p_{00} & p_{01} & p_{02} & p_{03} \\ p_{10} & p_{11} & p_{12} & p_{13} \\ p_{20} & p_{21} & p_{22} & p_{23} \\ p_{30} & p_{31} & p_{32} & p_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \tag{3}$$

where $(u_0, v_0)^{\mathrm{T}}$ are the coordinates of a point in the image plane, $(w\cdot u_0, w\cdot v_0, w)^{\mathrm{T}}$ represent the point's camera coordinates, and $(X, Y, Z)^{\mathrm{T}}$ indicate the point's world coordinates. Thus, given the image plane coordinates $(u_{\mathrm{ref}}, v_{\mathrm{ref}})^{\mathrm{T}}$ and depth $Z$ of a reference point, its position $(X, Y)^{\mathrm{T}}$ in the 3D world can be obtained through solving Eq.(3). Once the point's world coordinates $\boldsymbol{M}=(X, Y, Z)^{\mathrm{T}}$ are obtained, performing the matrix multiplication, the target point in the warped image can be found as

$$\boldsymbol{m}_{\mathrm{target}}=\boldsymbol{P}_{\mathrm{target}}\boldsymbol{M}, \tag{4}$$

where the variable $\boldsymbol{P}_{\mathrm{target}}$ indicates the projection matrix (as defined in Eq.(2)) of the target camera and $\boldsymbol{m}_{\mathrm{target}}$ indicates the point's coordinates in the target image plane.

Fig.1 shows the first image with depth on the center camera position (camera 4) of the test sequence 'breakdancers'. Fig.2 shows the result of the first image warped to its neighboring camera position. White areas indicate disocclusion. The depth image is not pre-processed. We employ nearest neighbor analysis to handle re-sampling and $z$-buffering to handle visibility. The reference points' positions of target points are registered for our later depth-aided disocclusion restoration.

(a) (b)

**Fig.1 The first image with depth from camera 4 of a test sequence 'breakdancers'**
(a) Color image; (b) Depth image

**Fig.2 Warping result**
The reference image is from camera 4 and the target image is from camera 5. White areas indicate disocclusion

**Criminisi's inpainting algorithm**

Both natural and computer-synthesized images consist of texture and structure. Basically, holes in a DIBR rendered image could be filled with ambient texture. Furthermore, the structure that breaks off at the hole boundary should also be continued. Criminisi *et al*.(2004) proposed a patch exemplar based algorithm. Their insight is that the 'filling front' should first propagate along the linear structure (termed 'isophote' in the inpainting literature), thus making the structure extended and connected regardless of the shape of the hole boundary.

They devised a priority function to determine the filling order. For ease of description, we adopt the identical notation (Fig.3).

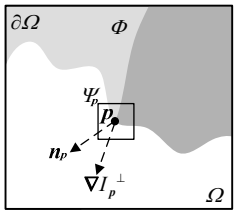



**Fig.3 Notation diagram (Criminisi *et al*., 2004)**
Given the block $\Psi_p$, $\boldsymbol{n}_p$ is the normal to the contour $\partial\Omega$ of the hole region $\Omega$. $\Phi$ is the non-hole region. $\nabla I_p^{\perp}$ is the isophote at point $\boldsymbol{p}$

Given a block $\Psi_p$ centered at point $\boldsymbol{p}$ for some $\boldsymbol{p}\in\partial\Omega$, the priority is computed as

$$P(\boldsymbol{p})=C(\boldsymbol{p})D(\boldsymbol{p}), \tag{5}$$

where the confidence term $C(\boldsymbol{p})$ and data term $D(\boldsymbol{p})$ are defined as

$$C(\boldsymbol{p})=\frac{\sum_{\boldsymbol{q}\in\Psi_p\cap\Phi}C(\boldsymbol{q})}{\text{area}(\Psi_p)}, \tag{6}$$

$$D(\boldsymbol{p})=\frac{|\nabla I_p^{\perp}\cdot\boldsymbol{n}_p|}{\alpha}. \tag{7}$$

In brief, $C(\boldsymbol{p})$ is the percentage of non-hole pixels in $\Psi_p$, and $D(\boldsymbol{p})$ is the dot product of gradient term $\nabla I_p$ and normal term $\boldsymbol{n}_p$. $\perp$ is the orthogonal operator and $\alpha$ is a normalizing factor.

The algorithm first computes the priorities of the points on the hole contour. It then selects block $\Psi_{p\max}$ with the highest priority and performs a block matching on the same image with the distance measure

$$\Psi_{\hat{q}}=\arg\min_{\Psi_q\in\Phi}d(\Psi_{p\max},\Psi_q), \tag{8}$$

where $\Psi_q$ is a block in the non-hole area $\Phi$ of the same size as $\Psi_p$. Next the hole points of $\Psi_{p\max}$ are filled with extracted samples from the best matching block $\Psi_{\hat{q}}$. After the block filling, the shape of the contour changes and the above three steps are repeated until all the hole points are restored.

Criminisi's algorithm is capable of propagating both linear structure and 2D texture into the target region. The approach inspired us. In the next section, we incorporate the depth cue into the cost function and make a revision to the priority computation. Also we develop warping-specific processing strategies including depth image morphological processing and luma-first processing together with depth-aided inpainting to achieve exciting disocclusion restoration results in the multi-view warped images.

Besides its capability to consider both texture and structure, there are two additional reasons for us to adopt the inpainting algorithm. First, in the reference image, there are no mapping positions of the hole points in the warped image. Thus, the disocclusion areas are implicitly marked without a requirement for explicit manual labeling, which eases the auto-execution of the algorithm. Second, the depth image used by DIBR can be reused as a depth cue by the inpainting to accurately handle depth ambiguities.

## PROPOSED METHODS

In this section we describe our proposed depth-aided inpainting algorithm in detail. As DIBR warped images have characteristics that are different from those of ordinary damaged paintings, we perform morphological operations on the depth image to reduce needle-hole in the warped image, and restore the luminance component first, to adapt to various image sampling formats. The key improvement is the incorporation of a depth cue into the texture estimation. Also, the revision of the priority calculation plays an important role. We describe the four steps in order of their function.

### Depth image morphological processing

Before DIBR, we perform morphological processing on the depth image.

White areas inside the warped image indicate holes (Fig.2). The more notable lump-holes are along object and image boundaries. The other needle-holes are typically one pixel wide and spread over the warped image.

Holes are caused by missing information exhibited as depth discontinuities (Fig.1b). To make the depth value transit smoothly, pre-filtering methods (Chen *et al*., 2005; Zhang and Tam, 2005; Daribo *et al*., 2007) employ a Gaussian filter. But with large baseline warping (the baseline is the distance between the reference and warping positions), the approach would introduce visible geometrical distortions (Section 4).

Besides depth discontinuity, we also noticed that there are imperfections in the original depth image generated using current technology. Fig.4a shows the horizontal arm of the front man cropped from the depth image (Fig.1b). It is noticeable that the shape is fairly coarse. The coarseness to some extent brings needle-hole and uneven surface into the warped image (Fig.4b).

Aimed at reducing needle-holes, but not altering other points' depth values, we employ first opening then closing operations on the depth image. Fig.4c shows the same parts of Fig.4a after morphological operations. Fig.4d shows parts of the image warped from the processed depth image.

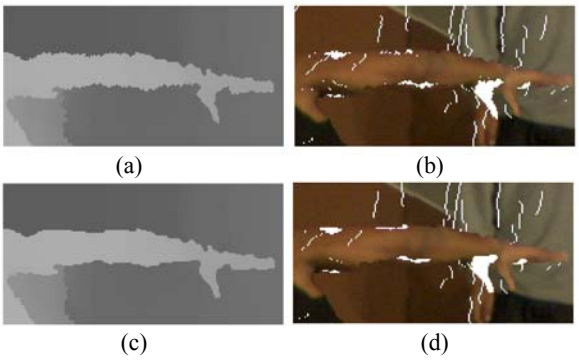



**Fig.4 The impact of morphological operation**
(a) Segment of horizontal arm from Fig.1b; (b) Segment warped from Fig.4a; (c) Result of opening and closing operations on Fig.4a; (d) Segment warped from Fig.4c. Note that the arm contour in Fig.4c is smoother than that in Fig.4a. Correspondingly, Fig.4d is smoother and has fewer holes than Fig.4b

The opening and closing reduces the needle-holes, while keeping gray intensities unchanged. Readers can refer to Digital Image Processing (Gonzalez and Woods, 2002; Gonzalez *et al*., 2005) for more details. From experiments, it was shown that there is about a 10% time saving from restoring holes on images warped from morphological processed depth images, compared with that from the unprocessed depth image case (details shown in the subsection of 'Subjective evaluation').

**Priority calculation**

The processed depth images are used in the DIBR warping process, as described in Section 2. Next we assign priority to points on the holes' contour to select a point with the highest priority to be processed first.

Criminisi *et al*.(2004) defined the priority as the product of confidence and data terms. Referring to Eq.(5), the confidence term assigns high priority to blocks surrounded by non-hole regions, while the data term favors blocks that contain strong isophotes. As filling proceeds in the hole interior, because $C(\boldsymbol{p})$ (Eq.(6)) is in the range [0, 1), once updated, it becomes smaller, indicating that we are less sure of the color value of filled points in the hole interior. Under a multiplicative case (Eq.(5)): $P(\boldsymbol{p})=C(\boldsymbol{p})D(\boldsymbol{p})$, the priority $P(\boldsymbol{p})$ may decay to zero. Thus, the isophote would stop propagating.

The impact of priority is subtle and its negligible power in restoring holes in DIBR-warped images can be seen in Section 4. We revise the priority as

$$P(\boldsymbol{p})=C(\boldsymbol{p})+D(\boldsymbol{p}). \tag{9}$$

Another revision we made was to the normalization method. We employ the unit normalization instead of intensity normalization ($\alpha$=255 for ordinary gray level luminance image) on $\nabla I_p$. The result of $\alpha(\nabla I_p)$ is a unit vector. This revision clips $D(\boldsymbol{p})$ into the identical dynamic range as $C(\boldsymbol{p})$, thus achieving meaningful addition in Eq.(9).

**Depth-aided texture estimation**

Once the priorities of contour points are calculated, the block containing the highest priority point is selected to be inpainted.

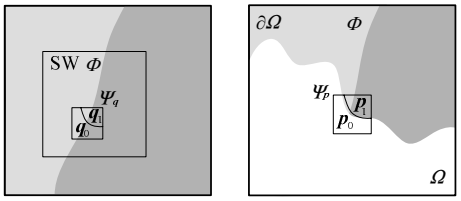
In the real image, points of similar depth are usually aggregated. Thus, one inpainting principle would be to select points with similar depth as existing surrounding points to remedy the missing area.

Considering the depth information, we devise the cost function $d(\ ,\ )$ in Eq.(8) as

$$d(\Psi_p,\Psi_q)=\mathrm{SAD}(\Psi_p,\Psi_q)+\lambda\left|\mathrm{average}\left(\sum\mathrm{depth}(\boldsymbol{p}_1)\right)-\mathrm{arerage}\left(\sum\mathrm{depth}(\boldsymbol{q}_0)\right)\right|, \tag{10}$$

where SAD is sum of absolute difference of mapped points. The SAD term favors blocks with similar texture as the surrounding region.

Fig.5 indicates the points' definition in Eq.(10). Mapped points in $\Psi_p$ are defined as $\boldsymbol{p}_1|\boldsymbol{p}_1\in\Psi_p\cap\Phi$, while unmapped points are defined as $\boldsymbol{p}_0|\boldsymbol{p}_0\in\Psi_p\cap\Omega$. $\boldsymbol{q}_0$ in $\Psi_q$ is of corresponding positions as $\boldsymbol{p}_0$ in $\Psi_p$.




(a) (b)

**Fig.5 Block match process diagram**

(a) Reference image (the block match process is performed within SW in the reference image); (b) Current warped image, in which $\boldsymbol{p}_0$ in $\Psi_p$ are to be inpainted. SW indicates search window. $\Psi_p$ and $\Psi_q$ are odd sized

$\lambda$ is a Lagrange multiplier, which is defined as

$$\lambda=\begin{cases}\lambda_{\mathrm{b}}, \operatorname{average}\left(\sum \operatorname{depth}(\boldsymbol{p}_1)\right) \le \operatorname{average}\left(\sum \operatorname{depth}(\boldsymbol{q}_0)\right),\\ \lambda_{\mathrm{f}}, \operatorname{average}\left(\sum \operatorname{depth}(\boldsymbol{p}_1)\right) > \operatorname{average}\left(\sum \operatorname{depth}(\boldsymbol{q}_0)\right).\end{cases} \tag{11}$$

Note that in Eq.(11) $\lambda_{\mathrm{f}}$ is larger than $\lambda_{\mathrm{b}}$, for filling unmapped points in $\Psi_p$ with textures from objects of smaller depth values is penalized. The depth term $\lambda|\cdot|$ of Eq.(10) favors candidate textures from farther background.

Based on two considerations, we perform the block searching in the reference image instead of the current warped image. First, there are many holes in the warped image, which would reduce the reliability of cost computation of candidate blocks and increase the difficulty of hole inpainting. Second, since every point's depth of reference image is preserved in the depth image, we could more conveniently use depth cues to aid the search process.

We locate the center of the search window on the point $\boldsymbol{p}_{\mathrm{c}}$, which is the projection point of the point with maximum gradient in block $\Psi_p$. There is a higher probability of finding the best matching block near $\boldsymbol{p}_{\mathrm{c}}$. Also, it is easy to locate $\boldsymbol{p}_{\mathrm{c}}$ as we register the projection positions during DIBR.

Since we process multi-view warped image in YUV space, the image is usually in 4:2:0 format. The position of the hole region between luma space and chroma space is always inconsistent. Thus, we simplify the computation of the SAD term by considering only the *Y* component. Also, we choose the odd-sized block $\Psi_p$, for its symmetry to place the contour point $\boldsymbol{p}$ at the center position.

**Chroma restoration**

Having found the best matching block $\Psi_{\hat{q}}$ in the reference image, the luma values of $\boldsymbol{p}_0$ are copied from $\boldsymbol{q}_0$ in $\Psi_{\hat{q}}$. The coordinates of $\boldsymbol{q}_0$ are also registered for the succeeding texture estimation.

After one block luma restoration, the affected points and their priorities on the hole contour are updated. The above described priority calculation and texture estimation steps repeat until all holes with their luma values are filled in.

Once all the luma holes are restored with their mapping positions registered, the chroma restoration can be performed. We copy the value from the reference image chroma space to the warped image chroma holes according to the registered mapping position of the specific luma hole point.

The luma-first restoration strategy adapts to various image sampling formats. Under a high fidelity 4:4:4 sampling format, by slight modification of the SAD computation, the estimation can more accurately find the best matching block.

In the next section, we will demonstrate the advantages of the depth-aided inpainting algorithm through experimental results. Also, we will discuss the mechanisms underlying the superiority.

## EXPERIMENTAL RESULTS AND DISCUSSION

To evaluate performance, we compared the proposed algorithm with that of Criminisi *et al*.(2004) and the typical Gaussian filter based asymmetrical filtering algorithm of Zhang and Tam (2005).

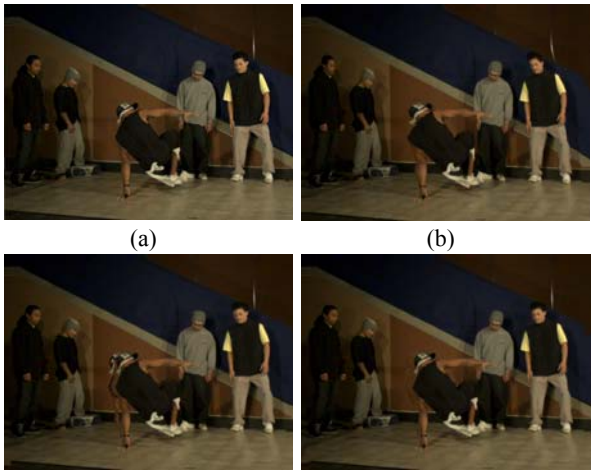
Experiments were based on test sequence 'breakdancers' with camera parameters and depth images provided by Microsoft Research Asia (Zitnick *et al*., 2004). The sequence has 8 views with 20 cm horizontal spacing between each. The video images have a resolution of 1024×768. The depth images are point based. For the DIBR warping, without explicit indication, the reference image corresponds to camera 4, while the target image corresponds to camera 5. All the experiments were run on a PC with 3.0 GHz Pentium IV CPU and 1 GB of RAM.

For the proposed algorithm, the block size was set to 9×9 and the search window size set to 15×15. There was a maximum of 225 candidate blocks. The

Lagrange multipliers $\lambda_f$ and $\lambda_b$ were set to 10 and 0, respectively. The mask for the depth image morphological processing was a 3×3 unit square. The specific value was derived from our experiments. Compared with other values, it showed relatively good subjective results.

We set identical parameter values for the corresponding parameters in Criminisi's implementation. For Zhang's algorithm, the resolution of the video image was 1024×768, while the baselines between cameras were 20 cm, which is about 70% of the image width, and thus we applied strong smoothing parameters. The standard deviation of the Gaussian core was set to 20 horizontally and 60 vertically. The filter window size was set to 3 times the corresponding standard deviation, i.e., 61 points horizontally and 193 points vertically.

Fig.6 presents the real image and the warped images of the above three algorithms.



(a) (b)

(c) (d)

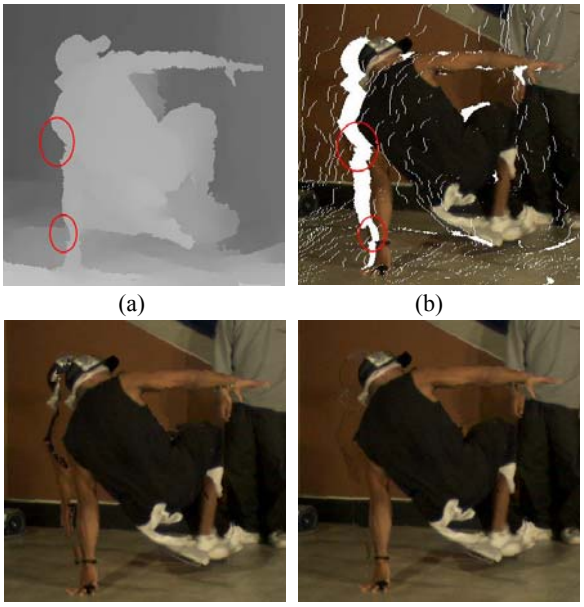
**Fig.6 The first image on the position of camera 5**

(a) Real image; (b) Zhang's algorithm: the depth image is asymmetrically filtered before DIBR; (c) Criminisi's algorithm: the warped image is inpainted; (d) Proposed algorithm: the depth image is opened and closed, and the warped image is depth-aided inpainted. (b), (c) and (d) are warped images

## Subjective evaluation

This subsection presents details and evaluates the subjective visual quality of the disocclusion restored region.

The first detail shows the power of depth information in aiding disocclusion restoration. We crop parts of Figs.1b, 2, 6c, and 6d, and show them in Fig.7.



(a) (b)

(c) (d)

**Fig.7 Segments of images**

(a) From Fig.1b, which is the unprocessed depth image (the ellipses indicate defects in the depth value); (b) From Fig.2, which is warped without any processing; (c) From Fig.6c, which is inpainted by Criminisi's method; (d) From Fig.6d, which is inpainted by our method

Note the hole boundary between the man's body and background (Fig.7b). Fig.7c shows that parts of the man's left body are duplicated as the result of the SAD term only in Eq.(10). Fig.7d shows that when taking the filling points' depth into consideration, the 'ghost' artifact is much reduced.

The depth term $\lambda|\cdot|$ in Eq.(10) plays an important role during texture estimation. The multiplier $\lambda$ varies according to the difference in depth between existing points and filling points. If the average depth of filling points is smaller than that of existing points, which means that relative to the existing points, filling points are closer to the observer position, these filling points are penalized to larger $\lambda$ values. Otherwise, the filling points are bestowed smaller $\lambda$ values. In general, the function favors points of equal or larger depth than that of existing points. This mechanism explains why much fewer foreground points are assigned to the left hole boundary in Fig.7d.

However, there is also artifact in Fig.7d. In the depth image in Fig.7a, the regions indicated by ellipses on the arm sink in, which makes the lumps away from the arm in Fig.7b. This is the cause of the 'ghost' artifact. Our algorithm detects that there is

notable depth difference between a lump and its surrounding points (actually they are both on the arm, but the depths are different) and therefore selects blocks from the background instead and leaves the lump alone. Without the depth term, the artifact would become much worse. Though not perfect, the efficacy of our depth-aided inpainting is proved.

The second detail illustrates the impact of our revision to priority computation.

Fig.8b shows that when filling proceeds into the hole interior, textures indicated inside the square have a tendency to spread. Our algorithm forces the line structures to grow until they intersect (Fig.8c). This structure preservation feature comes from the addition of Eq.(9).

(a) (b) (c)

**Fig.8 Segments of images**

(a) From Fig.2, which is warped without any processing (note that it has a hole area on the right side); (b) From Fig.6c, which is inpainted by Criminisi's method; (c) From Fig.6d, which is inpainted by our method

The addition of confidence and data enforces the function of the data term. Under the multiplication case of Eq.(1), while confidence decays in the hole interior, though structure exists, the point priority inevitably becomes small, which makes it uncertain to propagate first. Instead, the addition weakens the confidence influence. A strong isophote takes precedence in propagation.

The third detail demonstrates the discrepancy between an image warped from a Gaussian filtered depth image and the real image.

The main advantage of the depth image pre-filtering method is the smoothness presented in the warped image. It seems that there are fewer artifacts in Fig.6b than in Fig.6d. However, note the two ends of each vertical line when the real image, the inpainted warped image by our method and the Gaussian filtered image are aligned (Fig.9). It can be seen that there are geometrical distortions in foreground objects in the Gaussian filtered image, while there are almost no distortions in our inpainted image. The arm and shoulder of the man in the center of the bottom image are broadened, while the faces of two men on the left are narrowed. The alignment also reveals that the men's physical positions are left-biased compared with the real image.

**Fig.9 Aligned images for comparison**

The top image is of Fig.6a, which is the original real image. The middle image is of Fig.6d, which is the warped image inpainted by our method. The bottom image is of Fig.6b, which is warped from asymmetrical Gaussian filtered depth image. The ends of each vertical line indicate positions of geometrical distortions

We explain the cause of the above observations as follows. Fig.10 shows the comparison between the original depth image and the Gaussian filtered image. In our experiment, the positions of warped objects target to the right of the original positions in the reference image. Thus, under the normal horizontal left to right scan of the depth image, the transition from dark to light indicates disocclusion position in the warped image, while the transition from light to dark does not.

(a) (b)

**Fig.10 Depth image comparison**
(a) Original image; (b) Asymmetrical filtered image

The Gaussian filter smoothes these transitions on the depth image. In positions neighboring original dark to light transition, the light color is darkened, which results in expansion of a foreground object in the warped image. In positions neighboring original light to dark transition, the dark color is lightened, which results in expansion of background texture in the rendered image. Thus, the foreground object is narrowed, even concealed, by the expanded background texture.

In brief, the positions of foreground objects change from what they should be because the value of the depth image alters. Pre-filtering approaches such as edge-dependent filtering (Chen *et al*., 2005; Daribo *et al*., 2007) may to some extent alleviate these distortions. Under cases of short baseline warping, the approach could produce satisfactory results.

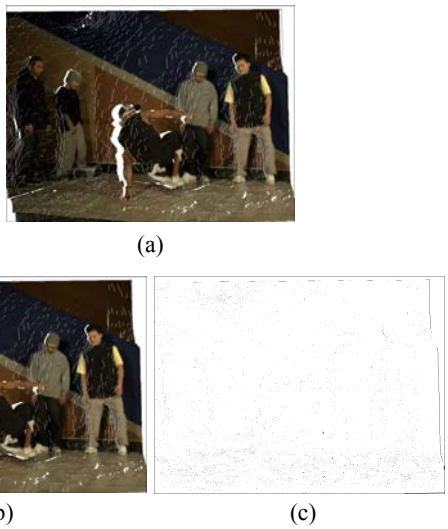
The fourth detail demonstrates the effect of the morphological operation on the depth image in reducing needle-hole in the warped image.

Fig.11 demonstrates the effect of needle-hole reduction in the warped image when employing morphological processing on the depth image.

Comparing Fig.11a with Fig.11b, if the position in Fig.11a is a hole (in white), but the same position in Fig.11b is not a hole, then we mark this position in Fig.11c in black. The non-hole areas in both Figs.11a and 11b are marked in white in Fig.11c. Thus, Fig.11c shows the hole reduction in Fig.11b compared with Fig.11a.

The morphological operation reduces holes in the warped image before it is inpainted. Our software (without code optimization) shows that to finish one image with or without morphological operation takes about 8 and 9 s, respectively.

For the final image quality, we compared the whole sequence, and found few differences between the inpainted images with and without depth image



(a)

(b) (c)

**Fig.11 Hole reduction illustration**
(a) Warped image without morphological processing (of the target position on camera 5, the same as Fig.2), where white areas indicate holes. (b) Warped image with morphological processing. (c) The black point indicates that the position in Fig.11a is a hole, while in Fig.11b it is not a hole; the white background indicates the same non-hole area in both Fig.11a and Fig.11b

morphological processing. The peak signal-to-noise ratio (PSNR) statistics in Fig.15 ('No-Morph' vs. 'Proposed') confirm this. Thus, we may say that the main function of morphological operation lies in reducing needle-hole in the warped image, thereby accelerating the overall process. Its effect in improving the image quality is limited.

The above comparative details could be found throughout the test sequence. Our proposed method inherits features from the original inpainting algorithm, and demonstrates its power in depth texture discrimination, structure preservation, and geometrical loyalty. Nevertheless, the algorithm also has limitations. It is best suited to restoring small hole areas. Fig.12a shows the warping result from the position of camera 7. The reference position is camera 4. There is about 60 cm between the reference and the target positions. We can see there are several lump holes in Fig.12a indicated inside the squares, and the inpainted result in Fig.12b seems not so perfect. The edge of the foreground man indicated in the two squares to the left in Fig.12b is not clear and the restored area shows the wrong textures. Also, the diagonal structure in the square to the right in Fig.12b does not intersect with the horizontal structure.

(a) (b)

**Fig.12 The first image from the position of camera 7**

(a) The warped image containing lump holes inside the squares; (b) The inpainted result of Fig.12a (the restored area inside the squares is not perfect)

## PSNR comparison

A metric for objective comparison is difficult to apply to DIBR rendered images. The concern is that, besides disocclusion, there are several factors that influence the quality of the warped image, such as the accuracy of the depth image and re-sampling of fractional points on the projected plane. To some extent, under current technology, DIBR rendered images are analogous to real images on projected position, but not identical.

Chen *et al*.(2005) used a model to evaluate the PSNR quality by comparing the non-hole region rendered from an unprocessed depth image with the same region rendered from a pre-processed depth image. For these methods consisting only of post-processing, the non-hole region is the same as the region warped from the unprocessed depth image. Our method employs morphological operation on the depth image. However, the impact on the non-hole region in the warped image is minor. To demonstrate fully the PSNR quality of the warped image, we employ the comparison system shown in Fig.13. We compare the non-hole region warped from the unprocessed depth image with the same region warped from the processed depth image. Also, we present the whole image PSNR comparison between the warped image and the real image.

Fig.14 shows the non-hole region comparison results. The PSNR value of the proposed method is high since the morphological operation alters only the depth value of a limited number of points. While Gaussian filtering alters depth values of many more points, the PSNR of Zhang and Tam (2005)'s method is not high.

Fig.15 shows the whole image comparison results. The curve of our depth-aided inpainting method is always superior to those of other methods. Note that the PSNR value is not high enough to keep the fidelity of DIBR warped images to match real images. Nevertheless, the comparison results show the advantage of our algorithm in disocclusion restoration.

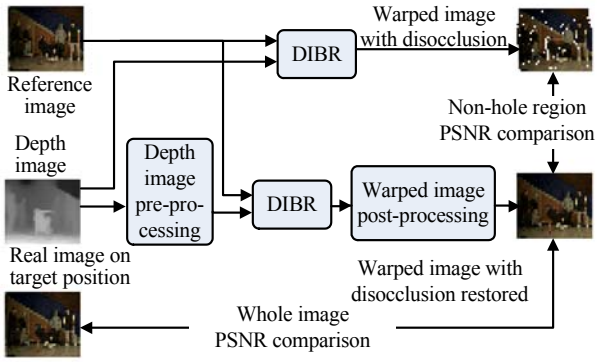



**Fig.13 PSNR comparison system**

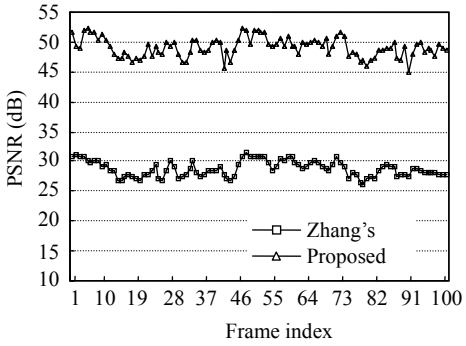



**Fig.14 Non-hole region PSNR comparison**

Non-hole region warped from the unprocessed depth image works as reference. The same region warped from the pre-processed depth image is compared with the reference region

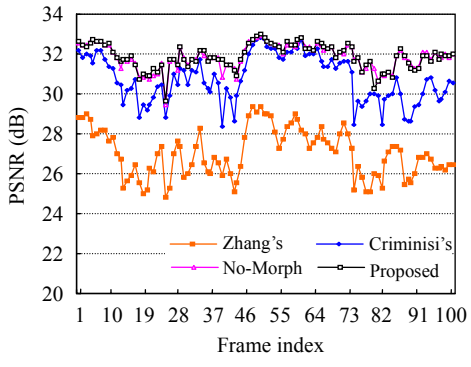



**Fig.15 Whole image PSNR comparison**

'No-Morph' indicates using our proposed method with no morphological operation on the depth image. Disocclusion restored warped images are compared with the real images

## CONCLUSION

This paper has presented a depth-aided texture inpainting algorithm for disocclusion restoration of multi-view images in a DIBR system. The target disocclusion inside a warped image is restored with points from the reference image considering the maximum similarity of their textures and depth.

The proposed algorithm inherits the features of Criminisi's algorithm. It is capable of propagating both structure and texture into the target region. By taking depth information into account, we efficiently handle the depth ambiguities in deciding the best matching block. The revision to point priority calculation ensures the structure connectivity in the disocclusion interior. The morphological operations on the depth image reduce needle-hole, and luma-first restoration makes the algorithm adapt to various color sampling formats.

Experimental results showed that our method has an advantage in artifact reduction on the object boundary. Compared with depth image pre-filtering methods, our method has no geometrical distortions. The inpainting method keeps the fidelity of the warped image close to the real image. Statistical results also show that our method is superior to a pre-filtering method on the whole warped image.

The proposed algorithm achieves (1) depth differentiation, (2) structure preservation, (3) object fidelity, and (4) flexible adaptation. The method has beneficial implications for multi-view applications based on DIBR technologies such as 3D-TV. While conventional post-processing methods such as surrounding points interpolation could provide only a quality-limited image, and a pre-filtering method such as asymmetrical filtering may bring distortion, the depth-aided inpainting method provides a worthwhile approach to warped image restoration.

## ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers for valuable guidelines, and Interactive Visual Media Group at Microsoft Research for multi-view video with fine quality depth data and camera parameters.

## References

Akar, G.B., Tekalp, A.M., Fehn, C., Civanlar, M.R., 2007. Transport methods in 3DTV: a survey. *IEEE Trans. Circ. Syst. Video Technol.*, **17**(11):1622-1630. [doi:10.1109/TCSVT.2007.905365]

Chen, W.Y., Chang, Y.L., Lin, S.F., Ding, L.F., Chen, L.G., 2005. Efficient Depth Image Based Rendering with Edge Dependent Depth Filter and Interpolation. IEEE Int. Conf. on Multimedia and Expo, p.1314-1317. [doi:10.1109/ICME.2005.1521671]

Criminisi, A., Perez, P., Toyama, K., 2004. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Process.*, **13**(9):1200-1212. [doi:10.1109/TIP.2004.833105]

Daribo, I., Tillier, C., Pesquet-Popescu, B., 2007. Distance Dependent Depth Filtering in 3D Warping for 3DTV. IEEE 9th Workshop on Multimedia Signal Processing, p.312-315. [doi:10.1109/MMSP.2007.4412880]

Fehn, C., 2004. Depth-Image-Based Rendering (DIBR), Compression, and Transmission for a New Approach on 3D-TV. SPIE, **5291**:93-104. [doi:10.1117/12.524762]

Fehn, C., Cooke, E., Schreer, O., Kauff, P., 2002. 3D analysis and image-based rendering for immersive TV applications. *Signal Process. Image Commun.*, **17**(9):705-715. [doi:10.1016/S0923-5965(02)00079-6]

Fehn, C., de la Barre, R., Pastoor, S., 2006. Interactive 3-DTV: concepts and key technologies. *Proc. IEEE*, **94**(3):524-538. [doi:10.1109/JPROC.2006.870688]

Flierl, M., Mavlankar, A., Girod, B., 2007. Motion and disparity compensated coding, for multiview video. *IEEE Trans. Circ. Syst. Video Technol.*, **17**(11):1474-1484. [doi:10.1109/TCSVT.2007.903780]

Gonzalez, R.C., Woods, R.E., 2002. Digital Image Processing (2nd Ed.). Prentice Hall, Upper Saddle River, New Jersey, USA.

Gonzalez, R.C., Woods, R.E., Eddins, S.L., 2005. Digital Image Processing Using Matlab. Prentice Hall, Upper Saddle River, New Jersey, USA.

Hartley, R.I., Zisserman, A., 2000. Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, UK.

Kauff, P., Atzpadin, N., Fehn, C., Muller, M., Schreer, O., Smolic, A., Tanger, R., 2007. Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability. *Signal Process. Image Commun.*, **22**(2):217-234. [doi:10.1016/j.image.2006.11.013]

Lee, J., Kim, C., 2007. Removing Foreground Objects by Using Depth Information from Multi-view Images. SPIE, **6696**:669627-9. [doi:10.1117/12.735824]

McMillan, L., 1997. An Image-based Approach to Three-dimensional Computer Graphics. PhD Thesis, University of North Carolina at Chapel Hill, USA.

Merkle, P., Smolic, A., Muller, K., Wiegand, T., 2007. Efficient prediction structures for multiview video coding. *IEEE Trans. Circ. Syst. Video Technol.*, **17**(11):1461-1473. [doi:10.1109/TCSVT.2007.903665]

Shade, J., Gortler, S., He, L.W., Szeliski, R., 1998. Layered Depth Images. 25th Annual Conf. on Computer Graphics and Interactive Techniques, p.231-242. [doi:10.1145/280814.280882]

Smolic, A., Mueller, K., Stefanoski, N., Ostermann, J., Gotchev, A., Akar, G.B., Triantafyllidis, G., Koz, A., 2007. Coding algorithms for 3DTV: a survey. *IEEE Trans. Circ. Syst. Video Technol.*, **17**(11):1606-1621. [doi:10.1109/TCSVT.2007.909972]

Tauber, Z., Li, Z.N., Drew, M.S., 2007. Review and preview: disocclusion by inpainting for image-based rendering. *IEEE Trans. Syst. Man Cybern. Part C-Appl. Rev.*, **37**(4):527-540. [doi:10.1109/TSMCC.2006.886967]

Vázquez, C., Tam, W.J., Speranza, F., 2006. Stereoscopic Imaging: Filling Disoccluded Areas in Depth Image-based Rendering. SPIE, **6392**:63920D-12. [doi:10.1117/12.685047]

Wang, L., Jin, H., Yang, R., Gong, M., 2008. Stereoscopic Inpainting: Joint Color and Depth Completion from Stereo Images. IEEE Conf. on Computer Vision and Pattern Recognition, p.1-8. [doi:10.1109/CVPR.2008.4587704]

Wang, W., Huo, L., Zeng, W., Huang, Q., Gao, W., 2007. Depth Image Segmentation for Improved Virtual View Image Quality in 3-DTV. Int. Symp. on Intelligent Signal Processing and Communication Systems, p.300-303. [doi:10.1109/ISPACS.2007.4445883]

Xu, G., Zhang, Z., 1996. Epipolar Geometry in Stereo Motion and Object Recognition. Kluver Academic Publishers, Dordrecht, The Netherlands.

Yoon, S.U., Ho, Y.S., 2007. Multiple color and depth video coding using a hierarchical representation. *IEEE Trans. Circ. Syst. Video Technol.*, **17**(11):1450-1460. [doi:10.1109/TCSVT.2007.905363]

Zhang, L., Tam, W.J., 2005. Stereoscopic image generation based on depth images for 3D TV. *IEEE Trans. Broadcast.*, **51**(2):191-199. [doi:10.1109/TBC.2005.846190]

Zitnick, C.L., Kang, S.B., Uyttendaele, M., Winder, S., Szeliski, R., 2004. High-quality video view interpolation using a layered representation. *ACM Trans. Graph.*, **23**(3):600-608. [doi:10.1145/1015706.1015766]