Digital Signal Processing 58 (2016) 93–101

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Depth map inpainting via sparse distortion model

Fen Chen, Tianyou Hu, Liwen Zuo, Zongju Peng *, Gangyi Jiang, Mei Yu

Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China

ARTICLE INFO

Article history:
Available online 1 August 2016

Keywords:
Depth map
Hole filling
Sparse representation
Kinect

ABSTRACT

The depth map captured from a real scene by the Kinect motion sensor is always influenced by noise and other environmental factors. As a result, some depth information is missing from the map. This distortion of the depth map directly deteriorates the quality of the virtual viewpoints rendered in 3D video systems. We propose a depth map inpainting algorithm based on a sparse distortion model. First, we train the sparse distortion model using the distortion and real depth maps to obtain two learning dictionaries: one for distortion and one for real depth maps. Second, the sparse coefficients of the distortion and the real depth maps are calculated by orthogonal matching pursuit. We obtain the approximate features of the distortion from the relationship between the learning dictionary and the sparse coefficients of the distortion map. The noisy images are filtered by the joint space structure filter, and the extraction factor is obtained from the resulting image by the extraction factor judgment method. Finally, we combine the learning dictionary and sparse coefficients from the real depth map with the extraction factor to repair the distortion in the depth map. A quality evaluation method is proposed for the original real depth maps with missing pixels. The proposed method achieves better results than comparable methods in terms of depth inpainting and the subjective quality of the rendered virtual viewpoints.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Currently, 3D video technologies have a wide range of commercial applications, including virtual reality simulations and entertainment. The multiview video plus depth (MVD) format presents users with a rich perception of depth [1,2]. Notably, the free viewpoint video (FVV) system of this format provides a broader visual perspective and richer 3D gradations. This enables users to immerse themselves in a multi-angle 3D visual experience [3]. Depth camera capturing and software estimation are the two primary methods of obtaining depth-related information. Recent technological advances have seen depth cameras such as Time of Flight (TOF) and Microsoft's Kinect become dynamic tools with a variety of applications. TOF cameras [4] are very effective in certain scenarios, but the image resolution of these fairly expensive cameras is relatively low. In comparison, images captured by the lower-cost Kinect [5] have a relatively high resolution. The main advantage of using the Kinect is the better alignment of color images and depth maps. As the Kinect uses infrared rays to measure distance and capture depth information, it is subjected to interference from many environmental factors. Multiple reflections from smooth objects, refractions from transparent objects, and additional scattering and occlusions could result in a loss of depth information and the occurrence of so-called depth holes. In addition, the noise introduced by the non-uniform sensitivities of the optoelectronic sensors often leads to inaccuracies in the depth map. Therefore, restoring is often required for depth maps.

So far, many methods have been proposed for depth video inpainting which can be roughly categorized into two types, reconstruction-based methods, and filtering-based methods. The reconstructed methods restore depth video by image inpainting technique [6]. Many image inpainting methods are proposed and applied in many application fields. In geophysical fields, splines approximation and finite element method can be used to fit the surface with surface patches, varying data, or curve set [7–11]. The spline approximation method can also be used for depth image inpainting [12–14]. Liu et al. [15] present a new energy minimization method to restore the missing regions and remove noise in a depth map. Their method using a TV 21 regularization could preserve sharp edges and achieve better inpainting result of depth map. In a study reported by Viacheslav et al. [16], the texture and structural characteristics of the sample were used to solve partial differential equations in order to perform inpainting.

Many depth image techniques belong to the filtering-based methods. Camplani et al. [17] proposed a hole filling strategy for the depth maps obtained with the Microsoft Kinect device.

* Corresponding author.
E-mail addresses: chenfen@nbu.edu.cn (F. Chen), hutianyou123@126.com (T. Hu), zuoliwen111@163.com (L. Zuo), pengzongju@126.com (Z. Peng), jianggangyi@nbu.edu.cn (G. Jiang), yumei@nbu.edu.cn (M. Yu).

http://dx.doi.org/10.1016/j.dsp.2016.07.019
1051-2004/© 2016 Elsevier Inc. All rights reserved.

The depth map is repaired by a joint-bilateral filtering which includes spatial and temporal information. Telea [18] developed a fast marching algorithm (FMM) for image restoration, and Shen et al. [19] utilized a joint bilateral filtering method to perform rapid inpainting of the depth map. However, their research did not consider the structural information of the corresponding color images. The missing edge of depth map could not be perfectively repaired. Jonna et al. [20] proposed an optical flow technique to remove fences/occlusions from depth maps, but their algorithm was not suitable for handling dynamic data associated with rapidly moving scenes. Bhattacharya et al. [21] used a guided method based on texture image edge information to repair distortions and reconstruct the depth map. However, their method did not account for the complexity of the color-texture information. Qi et al. [22] used a method of fusing depth and color structure information for depth map inpainting. Schmeing et al. [23] utilized color segmentation to repair depth map edges, and also introduced an edge evaluation model. Although this method utilized texture-color information, the algorithm for processing inaccurate depth information was too simplistic.

It is a trend that using learning-based method to restore depth map. Sparse representation theory has been successfully used in images inpainting. Aharon et al. [24] proposed an overcomplete dictionary for sparse representation which achieves better results. Xie et al. [25] proposed a robust coupled dictionary learning method with consistency constraints to reconstruct the corresponding depth map. In addition to the incorporation of an adaptively regularized shock filter to simultaneously reduce jagged noise and sharp edges, this method could effectively reduce data uncertainty and prevent the dictionary from being overfitted. Fan et al. [26] proposed a high-resolution depth map reconstruction method using a sparse linear model. Although this method improves the quality of high-resolution images reconstructed from low-resolution distorted depth maps, it requires early training of high-resolution image characteristics. In addition, the results obtained from training high-resolution noisy images were not ideal.

Although some learning-based approaches have been proposed, they are not effective to process high-resolution and noisy depth map. Therefore, we developed a depth map inpainting algorithm based on a sparse distortion model. The proposed method uses an overcomplete dictionary learning method to inpaint depth maps containing holes and noise. Although most conventional methods are based on color image characteristics, there are occasions where the corresponding color images of distorted depth maps are not available as a reference. In such cases, the proposed method is very effective for depth map inpainting. Our approach can be divided into several stages. First, a sparse distortion model is constructed. K-means singular value decomposition (K-SVD) is used to train the distorted depth map, as well as multiple undistorted depth maps. This procedure gives learning dictionaries for the distorted depth map and real depth maps. K-SVD is a high-precision training method that identifies the sample block based on a relatively small number of trained characteristics. Second, an orthogonal matching pursuit (OMP) algorithm is used. We use OMP to obtain the sparse coding coefficients of the distorted depth map and the undistorted depth map. Using the relationship between the learning dictionary and sparse coding coefficients of the distorted depth map, we then obtain the approximate depth values and depth map features. Subsequently, the joint space structure filter is employed for noise reduction, enabling the edge information of the image to be preserved. The results are applied to the threshold determination method to obtain the extraction factor. Finally, the relationship between the learning dictionary and sparse coding coefficients of the undistorted depth map is used to obtain the undistorted depth map features, which are combined with the extraction factor to perform the inpainting of the depth map. We also propose a pre-processing-based evaluation method for examining the quality of inpainting in depth maps. Experimental results demonstrate that the proposed method achieves improvements in both the subjective and objective quality of depth map inpainting, as compared to a variety of existing methods. Furthermore, the inpainted depth maps exhibit improved virtual-viewpoint subjective quality over those given by other techniques.

The main contributions of this paper are 1) a depth map inpainting algorithm which is based on a sparse distortion model, and 2) a joint space structure filter which can reduce noise of distorted depth map and improve accuracy of mask extraction.

The remainder of this paper is organized as follows. In Section 2, we describe the framework of our depth map inpainting algorithm. Section 3 presents a series of experimental results to evaluate the proposed method, and discusses its performance against that of other inpainting techniques. Our conclusions are summarized in Section 4.

2. Framework of the depth map inpainting algorithm

The framework of the proposed method is shown in Fig. 1 which E_1 and E_2 are undistorted and distorted depth maps, $\boldsymbol{D}_x$ and $\boldsymbol{D}_z$ are undistorted and distorted learning dictionaries, $\boldsymbol{a}_x$ and $\boldsymbol{a}_z$ are sparse coefficient vectors, and $\boldsymbol{P}$ is extractor factor. The method includes three parts, sparse distortion model building, training, and denoising and reconstruction. In training process, K-SVD algorithm is used to implement dictionary training on E_1 and E_2 to respectively obtain $\boldsymbol{D}_x$ and $\boldsymbol{D}_z$ firstly. Then, OMP sparse coding algorithm is applied to obtain $\boldsymbol{a}_x$ and $\boldsymbol{a}_z$. In denoising and reconstruction part, $\boldsymbol{D}_z$ and $\boldsymbol{a}_z$ are combined with the joint space structure filter to denoise the distorted depth map firstly. Then, the results are fed into the extraction factor judgment algorithm to obtain $\boldsymbol{P}$. Finally, $\boldsymbol{D}_x$, $\boldsymbol{a}_x$, and $\boldsymbol{P}$ are used to reconstruct the depth map.

2.1. Sparse distortion model building and depth map training

The distortion of the depth map can be described using an input–output model. A depth map distortion matrix can be expressed as the summation of the noise matrix and the product of the extraction factor matrix and the undistorted depth map matrix, i.e.,

$$\boldsymbol{Z} = \boldsymbol{P}\boldsymbol{X} + \boldsymbol{v} \tag{1}$$

where $\boldsymbol{Z}$ represents a distorted image, $\boldsymbol{X}$ is the original image, $\boldsymbol{P}$ is the extraction factor, i.e., hole-mask, and $\boldsymbol{v}$ represents Gaussian noise. If $\boldsymbol{Z}$ is known and $\boldsymbol{v}$ can be removed, $\boldsymbol{P}$ can be obtained from the specific characteristics of $\boldsymbol{Z}$ and $\boldsymbol{X}$. Based on the relationship between $\boldsymbol{Z}$ and $\boldsymbol{P}$, the model is then possible to obtain the repaired depth map, thereby realizing depth map hole filling and noise removal. Therefore, an adaptive dictionary based on the training stage is used to remove noise and inpaint depth holes. Let us assume that, in Eq. (1), $\boldsymbol{X} \in \mathbb{R}^n$ and $\boldsymbol{Z} \in \mathbb{R}^n$ are the original image and the distorted image, respectively, where n represents the number of pixels. $\boldsymbol{X}$ and $\boldsymbol{Z}$ can be expressed as l linear combinations of n-dimensional atoms. Therefore, the entire sparse representation can be expressed as:

$$\boldsymbol{X} = \boldsymbol{D}_x\boldsymbol{a}_x + \varepsilon_x \tag{2}$$

$$\boldsymbol{Z} = \boldsymbol{D}_z\boldsymbol{a}_z + \varepsilon_z \tag{3}$$

where $\boldsymbol{D}_x \in \mathbb{R}^{n\times l}$ and $\boldsymbol{D}_z \in \mathbb{R}^{n\times l}$ denote training dictionaries, in which column vectors represent dictionary atoms; $\boldsymbol{a}_x \in \mathbb{R}^l$ and $\boldsymbol{a}_z \in \mathbb{R}^l$ represent sparse coefficient vectors; and $\boldsymbol{\varepsilon}_x$, $\boldsymbol{\varepsilon}_z$ represent errors. When $\boldsymbol{\varepsilon}_x$ and $\boldsymbol{\varepsilon}_z$ are negligible, it is assumed that the estimated training samples can be represented by $\boldsymbol{X}_E$ and $\boldsymbol{Z}_E$, which can be expressed as:

Download English Version:

https://daneshyari.com/en/article/6951918

Download Persian Version:

https://daneshyari.com/article/6951918

Daneshyari.com