ISSN : 0974 - 7435 Volume 10 Issue 13

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(13), 2014 [7081-7086]

Reconstruction of kinect sensor's depth maps based on inpainting and CS theory

Qian Zhang, Bin Cui, ShaoMin Li, Pei Wang, Jifeng Huang

College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, (CHINA)

E-mail : zq0120110@163.com

ABSTRACT

The Microsoft's Kinect sensor can generate a real-time dense depth map of a indoor scene and relatively commercial available. However, there are some artifacts like holes, instability of the raw input data which seriously affect the application. In this paper,by using inpainting algorithm and CS theory which enable the kinect generate a dense depth map, the large holes be inpainted without blurring between foreground and background and the edges of the objects are sharpened. Experiments on captured data including complex foreground objects and large areas of missing values demonstrate the effectiveness of the method.

KEYWORDS

Kinect; Depthmap; Compressive Sensing.

© Trade Science Inc.

INTRODUCTION

With the development of computer graphics and video coding technology, three-dimensional televisions has become the next logical development towards a more natural and live visual home entertainment experience.[1] The 3-D audio-visual (3DAV) work group Has been formed to investigate the standardization in this area. FVV (Free Viewpoint Video) is the main Application of 3D TV which need an array of arbitrary multiple camera set-up[2], thus leads to the generation of mass video streams which need storage and transmission[2]. Restricted to the limited bandwidth, we cannot transmit all the captured pictures. The common solution is to synthesize the intermediate views.

Depth map plays an important role in the view synthesis for 3D video applications, which is a gray image, the darker regions in the depth map represent far-off objects and lighter regions represent closer objects. With smooth regions and sharp edges in depth map, the most important information rely in the edge region, so we need to find efficient method to estimate and compress an accurate depth map of edge parts.

To generated depth map the usual method is computing the depth map algorithmically by stereo matching, or using depth-sensing devices or range-finding cameras, which is known for the contactless, non-intrusive sensing of the environment.

Among the optical sensors, Microsoft's kinect sensor has the highest influence on the robotics and mechatronics communities in the last three years[3-7], which is a RGB-D sensing device can generates a dense depth map and color image of the scene, the resolution is typically 640×480 at real-time rate of 30 frames/sec, which is comparable to a time-of-flight camera[8]

Compressive sensing theory is a hot research area in image processing, such as image denoising, image coding, etc. CS theory enable certain images can be reconstruct from far fewer measurements which breaks the limitation of Nyquist sampling criterion and has worldwide used[9-10].

Compressed sensing reconstruction algorithm can be divided into two categories including the convex optimization algorithm and greedy algorithm. The convex optimization algorithm by adding constraints to obtain the sparse representation, and the reconstruction method mainly including the basis pursuit method (Basis PurSuit, BP) and variation total algorithm (TV),and can get good reconstruction result. However, the convex optimization algorithm need large amount of calculation, the reconstruction speed is low. On the other side, greedy algorithm is based on the orthogonal matching pursuit method (Orthogonal Matching Pursuit, OMP), The reconstruction speed of greedy algorithm is fast, but the reconstruction precision is lower than convex optimization algorithm[11].

Traditionally, the image/video under coding should be transformed from pixel domain to frequency domain using Discrete Cosine Transform or wavelet transform, such as JPEG or H.264. However, if edge parts within depth map are projected on frequence domain, many large amplitude signals are produced [12], it is well known that the edge area in depth map play an import role in view rendering, if large amplitude signals quantized by quantization matrix, it leads to large quantization error[12]. The existing depth map compression method such as Conventional codec H.264 using prediction method to estimate the neighboring area, which only considering the horizontal, vertical or diagonal edges cannot efficiently predict edge blocks for depth map with arbitrary edge shapes and reduce the extra bits. On the other hand, CS method can be more efficient to compress depth map and preserve the edge area than traditional method.

In this paper, we propose an improved depth map coding method incorporated with the compressed sensing theory to improve the quality of intermedia view, which plays an important role in the 3D video applications[13],and the efficient compression of depth map is need to reduce the extra bits.

OPTIMIZATION OF DEPTH MAP GENERATED BY KINECT SENSOR

Despite the kinect sensor has apparent advantages, there are still some disadvantages in the Kinect generated depth maps. As we all known that the kinect sensor consisting of an infrared projector and an infrared camera at distance of 7.5cm. The projector using an IR laser with 830nm wavelength capturing the scene and calculate correspondence between projector and the camera. But in some special surfaces like glass wall which deflect infrared light result in losing depth value in such areas. Some light absorbing materials can also cause such problem.

The area that losing depth value is referred to as a “hole” with zero depth value. There are some reasons for the formation of holes[14], the first one is the objects in the scene beyond the range of the Kinect sensor,so that their depth cannot be captured. The other reasons include the shape of the objects and texture of the objects. Inpainting methods can reconstruct the damaged image which is usually used to fill the “holes”. Structural based inpainting method is widely used to find an intensity function of minimum complexity of the gap, Gradient based methods usually lead to a noticeable blurring in textured areas and large gaps[15,16] and textural based methods aim at reconstructing the thin stripe-shaped image regions regions. Such methods are however computationally costly.

In this paper, we adopt a fast inpainting algorithm based on color match for inpainting the holes in depth image to obtain a more smooth depth image[17].

We capturing the first frame of depth map and texture map as the initial maps and use frame difference method for the background update, after that we using background depth value to fill the hole in the current frame.

OVERVIEW OF COMPRESSIVE SENSING

Nyquist sampling criterion prove that signal can be exactly reconstructed from sampling signal at sampling rate greater than two times bandwidth. CS theory break the Nyquist sampling criterion, and predict that many signals can be represented or approximated with only a few coefficients in a suitable basis[18-20].

Suppose a spare signal X, is a K-sparse when it has at most K nonzero elements, and the number K $\ll$ N.

Consider X can be represented by a set of orthonormal basis $\psi = [\psi_1, \psi_2, \cdots, \psi_N]$, $X = \psi\theta$, θ is a coefficient vector.

We can express the sampling process as follows:

$$Y = \Phi X \tag{1}$$

Where Y is the measurement, and Φ is the M x N measurement matrix that takes M number of measurement from X.

According to the CS theory, we can recover the K-sparse signal X by estimating X from the measurement Y. So we have to find the sparsest solution from minimization problem as the following description:

$$\tilde{\theta} = \arg\min_{\theta} \|\theta\|_1, \quad s.t. \quad Y = \Phi X \tag{2}$$

by solving the minimization L1-norm problem, we can perfectly reconstruct X by

$$\tilde{X} = \psi\tilde{\theta}$$

DEPTH MAP RECONSTRUCTION AND VIEW RENDERING

Depth map plays an important role in the view rendering in 3D video applications[21]. We find that depth map usually has large smooth areas and very sharp edges, which result in sparse gradient, so We incorporate an additional total variation (TV) minimization for reconstruct the depth map for preserving the edges and avoid introducing noise to the results.

Sine depth map is the 2-D image, we should calculate the sum of the discrete gradient of the 2-D image, called TV norm

$$TV(U) = \sum \sqrt{|D_{UH}(i,j)|^2 + |D_{UV}(i,j)|^2} \quad \textbf{(3)}$$

$$D_{UH}(i,j) = U(i,j) - U(i-1,j) \quad \textbf{(4)}$$

$$D_{UV}(i,j) = U(i,j) - U(i,j-1) \quad \textbf{(5)}$$

where i and j represent the horizontal and vertical coordinates, respectively.

we reconstruct the data by Using the reconstructed disparity map $\hat{d}$ computed by performing above method, we can get the desired intermediate view I_s by linear interpolation, as follows:

$$I_s(x,y) = \begin{cases} I\ (x+\alpha*\hat{d}\ ,y) \\ \quad if\ (x+\alpha*\hat{d}\ <= ncols) and (x-\alpha*\hat{d}_L > 0) \\ 0 \\ \quad esle \end{cases} \quad \textbf{(6)}$$

Where I represent the texture image respectively, we think these regions are the remaining occluded regions, which should be repaired by using the inpainting method[22].

Taking into account both the real-time request and repaired quality for image restoration, we adopt the fast image inpainting method to complete repairing process.

RESULTS

To test the efficiency of our method, in this section, we present simulation results to show that our algorithm is good at staircase artifact, we use a Microsoft Kinect depth sensor connected to a computer running the Windows 7 operating system. The depth maps and the corresponding color images, the color image is a three channel RGB image and the depth map is a single channel gray image. The proposed method is implement in MATLAB, C, and OpenCV.

The results of the proposed algorithm can be seen in Figure 1 and it is clear that all the holes have been filled and the edges have been refined. In Figure 2, the view rendering quality of the proposed algorithm is compared with the method that not using the CS theory, we can see that the proposed method outperforms. The subjective quality difference can be clearly observed among the enlarged portions of above images, shown as Figure 2, from left to right, are the enlarged portions of the original images (Figure 2 (b), Figure 2(f)), the intermediate views reconstructed using our method (Figure 2 (d), Figure 2(h)), respectively. Note that, the edge and texture can be preserved well in our method, which indicates that the depths change smoothly and consistently, so that our method generates intermediate views with higher quality.

(a) (b)

Figure 1: the depth map obtained by our method

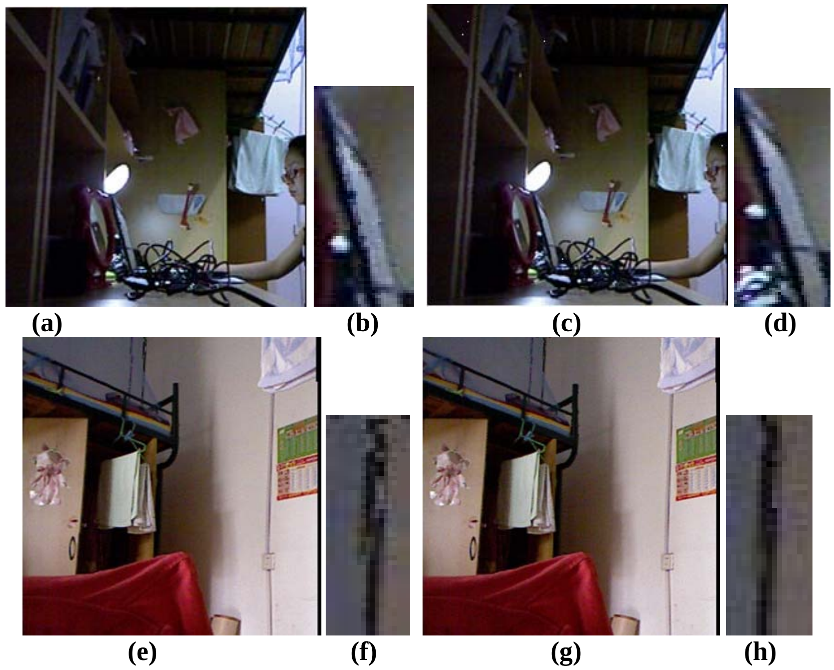

Figure 2 : Reconstructed view obtained by (a),(e)original method, (c),(g) our method,(b),(d),(f),(h) the enlarged portions

CONCLUSION

With further research about depth extraction and multi-view render, We investigated a depth extraction method by Kinect for the view rendering. By using inpainting algorithm and CS theory, large gaps could be inpainted without blurring between foreground and background and efficiently estimate the depth map with arbitrary edge shapes. Experiments on video data including complex foreground objects and large areas of missing values demonstrate the effectiveness of the method.

ACKNOWLEDGMENT

This research was jointly sponsored by Key Laboratory of Advanced Display and System Applications (Shanghai University), Ministry of Education, China (Project No. P200803), General research project of Shanghai Normal University (Grant No. SK201225), Shanghai municipal education commission of scientific research innovation projects (Grant No. 14ZZ125), Funding scheme for training young teachers in Shanghai Colleges.

REFERENCES

[1] Wojciech Motusik, et al; 3DTV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes, ACM, 23-24 **(2004)**.

[2] E.Cooke, O.Kauff, T.Sikora; Multi-view synthesis: A novel view creation approach for free viewpoint video, Signal Processing-Image Communication, **21(6)**, 476-492 **(2004)**.

[3] X.Ning, G.Guo; “Assessing spinal loading using the kinect depth sensor: A feasibility study,” IEEE Sensors J., **13(4)**, 1139–1140 **(Apr. 2013)**.

[4] M.Draelos, N.Deshpande, E.Grant; “The kinect up close: Adaptations for short-range imaging,” in Proc. IEEE Conf. MFI, 251–256 **(Sep. 2012)**.

[5] E.Machida, M.Cao, T.Murao, H.Hashimoto; “Human motion tracking of mobile robot with kinect 3D sensor,” in Proc. SICE, 2207–2211 **(2012)**.

[6] Z.Zhang, W.Liu, V.Metsis, V.Athitsos; “A viewpoint-independent statistical method for fall detection,” in Proc. 21st ICPR, 3626–3630 **(2012)**.

[7] C.Sun, T.Zhang, B.K.Bao, C.Xu, T.Mei; "Discriminative exemplar coding for sign language recognition with kinect," IEEE Trans. cybern., **43(5)**, 1418–1428 **(Oct. 2013)**.
[8] Krishna Rao Vijayanagar, Maziar Loghman, Joohee Kim; Real-Time Refinement of Kinect Depth Maps using Multi-Resolution Anisotropic Diffusion, Mobile Netw, **(Appl, 9, 2013)**.
[9] E.J.Candes; "Compressive sampling," in Int. Congress of Mathematics,Spain, **(2006)**.
[10] E.Candes, J.Romberg; "Sparsity and incoherence in compressive sampling,Inverse Problems, **23**, 969 **(2007)**.
[11] Zhao Yanmeng, Song Jiamxin; Method for Image Denoising Based on Compressed Sensing Total Variation Algorithm,video engineering, **38(5)**, 5-8 **(2014)**.
[12] TaeMin Cho, Youngwoo Lee, Jitae Shin; A Homogenizing Filter for Depth Map Compressive Sensing Using Edge-Awarded Method, ICTC, 591-595 **(2013)**.
[13] S.Tao, Y.Chen, M.M.Hannuksela, Y.K.Wang, M.Gabbouj, H.Li; Joint texture and depth map video coding based on the scalable extension of H.264/AVC," in 2009. ISCAS 2009. IEEE International Symposium on Circuits and Systems, 2353-2356 **(2009)**.
[14] M.Bertalmio, G.Sapiro, V.Caselles, C.Ballester; "Image inpainting," in Proc. ACM SIGGRAPH, 417–424 **(2000)**.
[15] Z.Tauber, Z.N.Li, M.Drew; "Review and preview: Disocclusion by inpainting for image-based rendering," IEEE Trans. Syst., Man, Cybern., Part C: Appl. Rev., **37(4)**, 527–540 **(Jul. 2007)**.
[16] Krishna Rao Vijayanagar, Maziar Loghman, Joohee Kim; Real-Time Refinement of Kinect Depth Maps using Multi-Resolution Anisotropic Diffusion, Mobile Netw, **(Appl, 9, 2013)**.
[17] Fast inpainting algorithm for kinect depth map, journal of shanghai university (natural science), **18(5)**, 454-458 **(2012)**.
[18] E.Cand`es, J.Romberg, T.Tao; Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory, **52**, 489–509 **(2006)**.
[19] D.Donoho; Compressive sensing, IEEE Trans. Inform. Theory, **52**, 1289–1306 **(2006)**.
[20] E.Cand`es, M.Wakin; An introduction to compressive sampling, IEEE Signal Processing Magazine, **25**, 21–30 **(2008)**.
[21] S.Tao, Y.Chen, M.M.Hannuksela, Y.K.Wang, M.Gabbouj, H.Li; Joint texture and depth map video coding based on the scalable extension of H.264/AVC, in 2009. ISCAS 2009. IEEE InternationalSymposium on Circuits and Systems, 2353-2356 **(2009)**.
[22] Q.Zhang, Ping An, Z.Y.Zhang, He Wang, Y.F.Wu, G.Y.Jiang; New reconstruction method for intermediate views from multiple Views, imaging science journal, **58(2)**, 89-95 **(2010)**.