3D Object Approximation from Multi-View Depth Information

Prof. Seema Idhate[1], Abhishek Kumar Pandey[2], Ainar Nanda[3], Tripti Pandey[4]

***Abstract*— Transmitting efficiently epitomized geometry of a dynamic 3D scene from a sender can facilitate a multitude of imaging functionalities at a receiver, such as amalgamation of virtual images at freely chosen viewpoints via depth-image-based translation. While depth maps—prognoses of 3D geometry onto 2D image planes at chosen camera viewpoints—can these days be readily captured by economical depth sensors, they are often despoiled by non-negligible procurement noise. Given depth maps need to be de-noised and compressed at the encoder for effective network transmission to the decoder, in this paper, we consider the de-noising and compression problems jointly, arguing that doing so will result in a better overall performance than the alternative of solving the two problems separately in two stages. In detail, we articulate a rate-constrained estimation problem, where given a set of observed noise-corrupted depth maps, the most probable 3D surface is sought within a search space of surfaces with illustration size no larger than a pre-specified rate restraint. Rate-controlled MAP solution diminishes to the conventional unrestrained MAP 3D surface reconstruction solution if the rate restraint is loose. To solve our posed rate-controlled approximation problem, we propose an iterative algorithm, where iteration, the structure and the texture of the depth maps are optimized alternately. We are using the MVC codec for compression of multi-view depth video and MPEG free viewpoint video sequences as input.**

***Index Terms*—Multiview Information, Depth Information Compression, 3D Image Recreation, Image DE noising .**

I. INTRODUCTION

With the beginning of consumer-level depth capturing sensors like Microsoft Kinect, depth images (per pixel distances between objects in the 3D scene and the capturing camera) can now be acquired cheaply from multiple viewpoints. Depth map constitutes a prognosis of the 3D geometry in the scene to a 2D image of fixed resolution. Thus, having acquired depth maps from multiple camera viewpoints, one can back task them to the 3D space to (partially) recover the original 3D geometry. If the multi-view depth maps—a illustration of the 3D geometry—are compressed and transmitted, then the receiver can accomplish a range of 3D imaging tasks, such as synthesis of virtual images from freely chosen viewpoints using texture and depth maps of neighboring camera views via depth-image-based interpretation (DIBR). To facilitate high quality communication of 3D geometry from sender to receiver, however, there are with two practical problems. The first problem is to approximate the actual 3D geometry of the scene from the depth maps attained from consumer-level depth sensors, which are typically despoiled by non-negligible acquisition noise.

The second problem is to find a compact illustration for the approximated 3D geometry— one that does not entail too many encoding bits—so that the communication cost will not be exorbitantly high. The conventional approach to these two problems— approximation of 3D surface from noisy interpretations and coding of chosen surface illustration—is to treat these problems as independent and solve them separately one after another. For example, one can use a 3D surface reconstruction solution from the computer vision literature, first to derive the most probable 3D surface from noisy depth observations, and then project this surface to chosen camera viewpoints as depth maps for compression.

There is disagreement that this is a sub-optimal approach; concerned problem of identifying a surface illustration that is both compact (require few encoding bits) and agrees with observations is innately a probabilistic one. If one first computes a most probable surface with no consideration for representation size, and then sends it as a deterministic input to a loss compression algorithm for depth map projection and coding, then all the probabilistic information available from observed data that could potentially be useful for compression is lost.

For example, a codec will not be able to compress lossily one part of the signal more destructively than another, even if they have very different local noise statistics. See Fig. 1 for an illustration.

Than a pre-specified rate restraint, e.g., r (ˆst) = ¯R. The most possible sequence of rate-constrained surfaces is then encoded using a MVC codec. Rate-constrained MAP solution reduces to an unconstrained MAP 3D mobilize surface reconstruction solution with no consideration for illustration size if the rate restraint is loose.

To formally formulate the problem, first defined is an error term -log p(ˆst |yt) that reflects the distance between the reconstructed 3D surface ˆst and the observed depth data yt. Then defined a rate term that approximates the coding bits of the re-projected depth maps dt (from the computed most probable rate constrained 3D surface ˆst) given a MVC codec is used for coding. To solve this optimization problem, proposed algorithm is efficient that finds a locally best solution by iterating between two steps:

i) Align edges in depth maps of consecutive views to match scene structure across views; and

Manuscript received Mar, 2015.

***Prof. Seema Idhate**, Department of Electronics and Telecommunication, JSPMs ICOER under UoP Pune, India,*
***Abhishek Kumar Pandey**, Department of Electronics and Telecommunication, JSPMs ICOER under UoP Pune, India,*
***Ainar Nanda,** Department of Electronics and Telecommunication, JSPMs ICOER under UoP Pune, India,*
***Tripti Pandey**, Department of Electronics and Telecommunication, JSPMs ICOER under UoP Pune, India,*

ISSN: 2278 – 1323

ii) Smooth surfaces within depth edges to match scene texture across views.

RATE-CONSTRAINED 3D SURFACE ESTIMATION

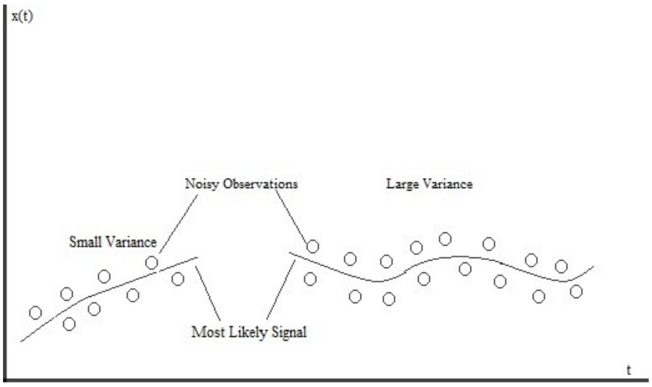

Fig. 1 Example of signal *x(t)* with different noise variances at different spatial regions *t*.

Using the MVC codec for compression of multi-view depth video and MPEG free lookout video test sequences as input, investigational results show that optimized 3D reconstructions computed by algorithm can reduce coding rate of depth maps by up to 32% compared to unconstrained estimated surfaces for the same quality of synthesized virtual views at the decoder than a pre-specified rate constraint, e.g., r (ˆst) ≤ .R .
Than a pre-specified rate restraint, e.g., r (ˆst) = ¯R. The most possible sequence of rate-constrained surfaces is then encoded using a MVC codec. Rate-constrained MAP solution reduces to an unconstrained MAP 3D surface reconstruction solution with no consideration for illustration size if the rate restraint is loose. To formally formulate the problem, first defined is an error term -log p(ˆst |yt) that reflects the distance between the reconstructed 3D surface ˆst and the observed depth data yt. Then defined a rate term that approximates the coding bits of the re-projected depth maps dt (from the computed most probable rate-constrained 3D surface ˆst) given a MVC codec is used for coding. Using the MVC codec for compression of multi-view depth video and MPEG free lookout video test sequences as input, investigational results show that optimized 3D reforms computed by algorithm can reduce coding rate of depth maps by up to 32% compared to unconstrained estimated surfaces for the same quality of synthesized virtual views at the decoder than a pre-specified rate constraint, e.g., r (ˆst) ≤ .R .

II. RELATED WORK

The difficulty of de-noising depth interpretations has been considered extensively in the literature, and can be broadly divided into two categories: i) de-noising of single depth images, and ii) reestablishment of 3D surfaces in space given noisy depth annotations from multiple viewpoints. The recent advance of depth sensing technologies such as time-of-flight cameras and structured light cameras has driven strong interest in the image processing researchers to study the depth image de-noising problematic. A common thread to these works is the manipulation of the known piecewise smooth signal prior in depth images for de-noising. A large portion of this work further assumes the availability of a perfectly affiliated color image along with the depth image as side evidence for depth de-noising. Some further assume the unique noise physiognomies of depth images captured by structured light cameras, which are very different from time of flight cameras. Though we also exploit the piecewise smooth characteristic of depth images in our algorithm, we differ in that we jointly solve the depth image de-noising and compression problem at the same time.

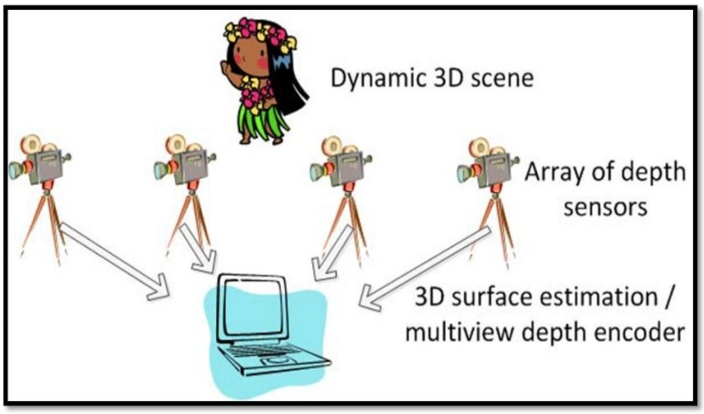

Fig. 2 Overview of multi-view depth capturing system for a dynamic 3D scene

To enable high quality communication of 3D geometry from sender to receiver, there are two practical problems. The first problem is to estimate the actual 3D geometry of the scene from the depth maps acquired from consumer-level depth sensors, which are typically corrupted by non-negligible acquisition noise. The second problem is to find a compact representation for the estimated 3D geometry that does not require too many encoding bits so that the communication cost will not be prohibitively high. So in this proposed approach focus is on above mentioned problems.

The problem of de-noising depth observations has been studied extensively in the literature, and can be broadly divided into two categories: i) De-noising of single depth images and ii) Reconstruction of 3D surfaces in space given noisy depth observations from multiple viewpoints. The recent advancement of depth sensing technologies is in field of image processing such as time-of-flight cameras. A large portion of this work further assumes the availability of a perfectly aligned colour image along with the depth image as side information for depth de-noising. As mentioned before, main focus is Depth Information De-noising. Most of the works in depth image enhancement can be grouped into two categories:

a) Super-Resolution
b) Segmentation Based Approach
c) Image in-painting

A. Super-Resolution Approach

It is an interpolation for depth super-resolution. A high resolution RGB camera was used to guide the up-sampling process on the depth image. To interpolate the missing depth

ISSN: 2278 – 1323

pixel, the scheme used neighbouring depth pixels mapped into the same colour segment as the target pixel. This method relied strongly on the extrinsic alignment between the colour and depth image. Drawbacks of Super-Resolution method includes he use of colour information for depth enhancement is based on the assumption that certain correlation exists between depth continuality and colour image consistency this assumption does not always hold as colour edges and depth edges do not necessarily coincide with each other. Accuracy of such implementation is very poor.

B. Segmentation Based Approach

Similar to previous method, a segmentation based method can also be used. In this method a non-local means filtering based approach was used to regularize depth maps and maintain fine detail and structure. Drawback of Segmentation Based Approach is that, it is Less Accurate.

C. Image In-Painting Based Approach

Wang et al. proposed a stereoscopic in-painting algorithm to jointly complete missing texture and depth by using two pairs of RGB and depth cameras. Regions occluded by foreground were completed by minimizing an energy function. The system required an additional pair of color and depth cameras to achieve the goal. Drawbacks of Image in-painting includes that it is highly complex. It's accuracy is completely dependent on in-painting and it is difficult to use in real time.

Second Objective of a project is Reconstruction of 3D object was: i) One naïve approach to finding a good rate constrained 3D surface is to separate the problem into two, ii) First estimate the underlying (ground truth) 3D surface from noise-corrupted observations regardless of representation size and iii) Perform conventional RD optimization as done in a standard video codec like H.264 given the estimated signal as input.

III. System Overview

We now provide an overview of our system model. We adopt a collection of V depth sensors capture depth images of the same dynamic 3D scene periodically from V different viewpoints, as shown in Fig.2. We undertake the cameras have the same spatial resolution and are synchronized in time. The caught depth observations are corrupted by non-negligible acquisition noise, modeled as multivariate Gaussian. Given the observed depth data, the encoder first estimates a rate-controlled 3D surface of the scene, for a given bit budget of $\bar{R}$ bits per frame. The chosen 3D surface is then re-projected back to the camera views, which are subsequently encoded as multi-view depth videos as a representation of the chosen 3D surface, using a known multi-view video coding scheme like MVC, for transmission over a communication channel of bandwidth $\bar{R}$. The task is to guesstimate the most prospective 3D surface given the observed depth data.

IV. Problem Formulation

We now present the formula used for image preprocessing the rate constraint 3D surface creation. We are using here the basics of noise calculative formulas such as means square error (MSE) and peak signal to noise ratio (PSNR).

These formulae are applied to the pre-processed images from average filter, Median Filter and Gaussian Filter. These filters are used to enhance the image quality. So after enhancement of images taken from various sources these formulae are applied so as to get the information of the images.

MSE is majorly used to measure the degree of distortion in image because they can represent the whole gray error values contain in the image. It is mathematically dutiful as well. It has virtues and is majorly used in image processing research.

$$MSE = \frac{1}{MN}\sum_{i=1}^{M} * \sum_{j=1}^{N} |I(i,j) - \hat{I}(i,j)|^2$$

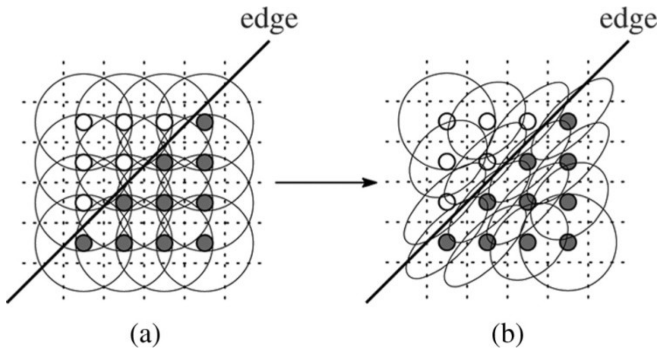

Fig. 3 Fixed versus adaptive kernels for image de-noising (a) fixed kernels employed by classical de-noising; (b) adaptive kernels change according to local structure.

PSNR is mostly used to compute the quality of renewal of lossy compression codec. For better image, PSNR should be greater. We evaluate 3D surface using camera image and depth information from multiple view-points of image. PSNR

$$PSNR = 10\log_{10}\frac{MAX_I^2}{MSE}$$

V. Algorithm/ Flowchart

ISSN: 2278 – 1323

Below Fig shows the flowchart of the subject proposed in the paper. It shows the flow by which the proposed paper is developed.

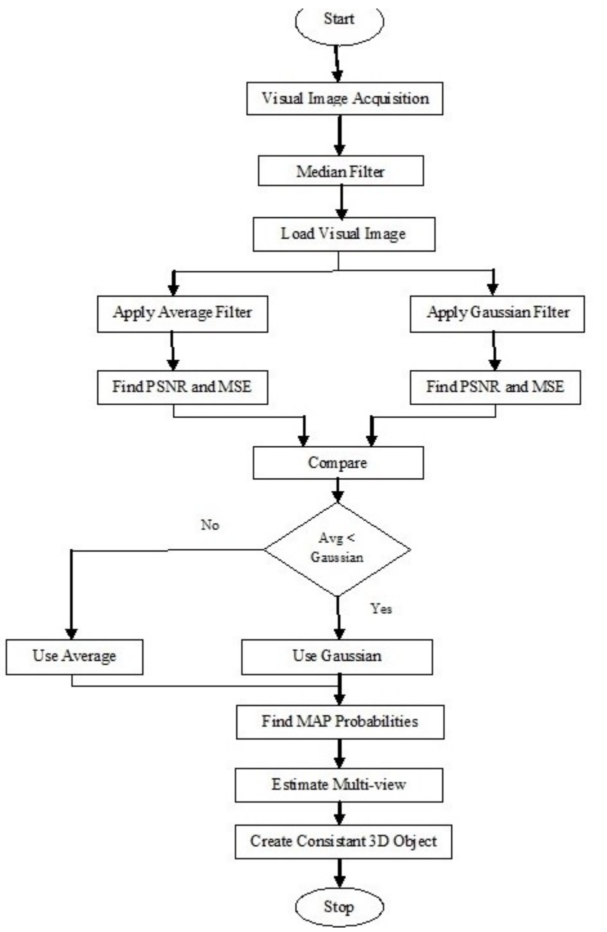

Fig. 4 Flowchart of 3D Object approximation

VI. Results

Result includes the Graphical User Interface screen of MATLAB. There we demonstrate the various parameters for

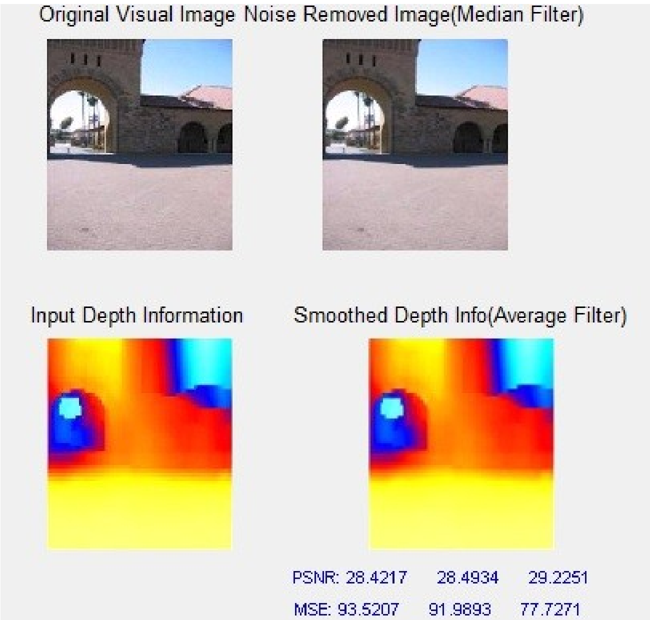

Fig. 5 PSNR and MSE values of Average Filter

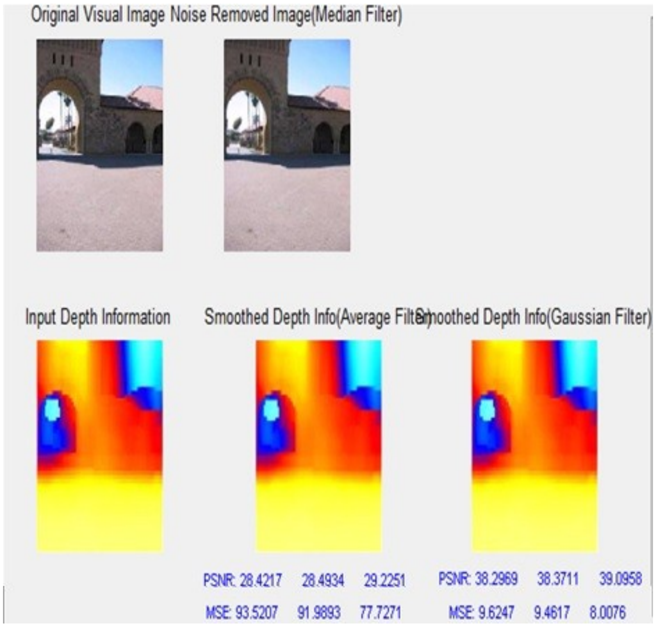

Fig. 6 PSNR and MSE values of Average Filter

Result shows the preprocessing of the image where we denoise the image taken from the various cameras. Preprocessing includes the filtering of image from various filters like Average, Median and Gaussian filter.

Table 1 : Average Filter Output

PSNR	28.4217	28.4934	29.2251
MSE	93.5207	91.9893	77.7271

Table 2 : Gaussian Filter Output

PSNR	38.2969	38.3711	39.0958
MSE	9.6247	9.4617	8.0076

Respective PSNR and MSE values are shown and demonstrated so as to indicate the accuracy achieved in filtering the image. The above tabular information shows that the Gaussian filter is much better in filtering as compared to average filter. So we prefer Gaussian filter for our proposed subject.

VII. Conclusion

Given noise despoiled depth observations from multiple view-points, in this paper we propose to construct a rate controlled 3D surface of a dynamic scene subject to a representation size restraint. Unlike previous work that finds the most likely 3D surface given noisy observations regardless of representation size, identified 3D surface optimally trades off the posterior probability with representation size. Method

ISSN: 2278 – 1323

proposed is an iterative algorithm that ultimately optimizes the scene arrangement (depth edges) and the scene texture (depth surfaces) until convergence. Experimental results show that using projections of rate constrained 3D rebuilding to multiple depth maps for multi-view depth video coding can reduce coding rate of depth ,aps by up to 32% compared to unconstrained estimated surfaces for the same quality of amalgamated virtual views at the decoder.

ACKNOWLEDGMENT

The authors wish to thank Prof. P Badadapude for his innovative input in improvising the subject and paper. This work was supported in part by a grant from JSPM's ICOER, Pune, Maharashtra, India

REFERENCES

[1] Z. Zhang and O. Faugeras,"A 3D world model builder with a mobile robot," *Int. J. Robot. Res., vol. 11, no. 4, pp. 269–285, Aug. 1992.*
[2] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in Proc. 23rd ACM Annu. Conf. Comput. Graph. Interact. Techn., New Orleans, LA, USA, Aug. 1996, pp. 303–312.
[3] R. Newcombe et al., "KinectFusion: Real-time dense surface mapping and tracking," *in Proc. 10th IEEE ISMAR, Basel, Switzerland, Oct. 2011, pp. 127–136.*
[4] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 11, pp. 1461–1473, Nov. 2007.*
[5] B. Huhle, T. Schairer, P. Jenke, and W. Straber, "Robust non-local denoising of colored depth data," *in Proc. IEEE CVPRW, Anchorage, AK, USA, Jun. 2008, pp. 1–7.*
[6] M. Tallón, S. D. Babacan, J. Mateos, M. N. Do, R. Molina, and A. K. Katsaggelos, "Upsampling and denoising of depth maps via joint-segmentation," *in Proc. 20th Eur. Signal Process. Conf., Bucharest, Romania, Aug. 2012, pp. 245–249.*
[7] D. Min, J. Lu, and M. N. Do, "Depth video enhancement based on weighted mode filtering," *IEEE Trans. Image Process., vol. 21, no. 3, pp. 1176–1190, Mar. 2012.*
[8] S.-W. Jung, "Enhancement of image and depth map using adaptive joint trilateral filter," *IEEE Trans. Circuits Syst. Video Technol., vol. 23, no. 2, pp. 258–269, Feb. 2013.*
[9] L. Chen, H. Lin, and S. Li, "Depth image enhancement for kinect using region growing and bilateral filter," *in Proc. 21st IEEE ICPR, Tsukuba, Japan, Nov. 2012, pp. 3070–3073.*
[10] J. Shen and S.-C. S. Cheung, "Layer depth denoising and completion for structured-light RGB-D cameras," *in Proc. IEEE Int. Conf. CVPR, Portland, OR, USA, Jun. 2013, pp. 1187–1194.*

ISSN: 2278 – 1323