

# Patch Based Super Resolution of Depth Maps

Sreelekshmi B

Department of Electronics and Communication  
Marian Engineering College  
Trivandrum, India

Sreena V G

Department of Electronics and Communication  
Marian Engineering College  
Trivandrum, India

**Abstract**—An issue of image technology is to improve the quality of an image. For improving the quality, Super Resolution technique is introduced. Super Resolution refers to the estimation of a high resolution (HR) image from one or more low resolution(LR)images.SR techniques handle the issues of alias removal, deblurring and denoising while interpolating the low resolution inputs. In 3D computer graphics, a depth map is an image or image channel that contains information relating to the distance of the surfaces of scene objects from a view point. Depth Maps have an advantage of limited bandwidth increase and offers flexibility and compatibility than color images. The depth map can vary depending on the depth sensors. One of the most popular depth sensors is Time of Flight (ToF) sensor. They are cheaper but provide depth maps of low resolution. So a patch based super resolution technique for improving the resolution of depth maps is proposed. Experimental results show better values for Peak Signal-to Noise Ratio (PSNR), StructuralSimilarityIndex Measurement (SSIM), Visual Information Fidelity (VIF) and entropy.

**Keywords**—Super Resolution, Depth Map,Patch, SSD, PSNR, SSIM

## I. INTRODUCTION

The most important technical term 'resolution' [1] is used in image manipulation which judges the quality of an image. It is defined as the smallest measurable detail in a visual presentation.In technical terms, there are three types of resolution.

- i. Low Resolution (LR)
- ii. High Resolution (HR)
- iii. Super Resolution (SR)

To increase the resolution of an image, mainly three methods are used. They are:

- Reducing Pixel Size of Image
- Increasing the Chip Size of the optical system
- Super Resolution

Low Resolution imaging systems utilize the technique of Super Resolution [12] and also it is cheap compared to other approaches. In simple words, Super Resolution is defined as the process of combining multiple low resolution images to form a high resolution image. The main aim of SR is to extract useful information or required image details. There are mainly four approaches of Super Resolution:

- Frequency domain based approach
- Interpolation based approach
- Regularization based approach
- Learning based approach

Learning based approach is commonly used approach and this approach is mostly used in this paper.

A depth map (fig 1) is defined as an image of values(integer or real) that represent distance from view point. In terms of definition, Depth Map is a kind of image which is composed of the gray pixels defined by 0~255 values. The "0" value of gray pixels stand for that "3D" pixels that are located at the most distant place in the 3D scene while the "255" value of gray pixels stand for that 3D pixels that are located at the most near place. Depth Sensors are widely used in many fields such as video-based rendering [4], robot manipulation [5], and gaming [6].Most popular active sensor is Time of Flight (ToF) sensor [2].ToF sensors can provide depth images from texture less scenes, which is not possible in the stereo vision techniques. However, a depth map captured by a ToF sensor has very low resolution 320x200 which is much lower than the resolution ofhigh definition color images. In this work, a patch based learning algorithmfor super resolution of Depth map is proposed.



Fig 1:Depth map for corresponding color images

Several methods can be used in the super resolution technique on color images. One of the most eminent methods is proposed by Freeman *et al.* [3].He proposed the single pass algorithmin which a high resolution color image is constructed from different low resolution color images.

On depth maps, many works were proposed. Yang *et al.* [2] proposed joint example based depth map super resolution in which high resolution test patches are constructed as sparse linear combinations from a learned dictionary and the registered color image is used as reference. This method is computationally complex.

Jing *et al.* [9] proposed the patchwork assembly algorithm on the depth image super resolution into an MRF multiple label image segmentation problems and an energy minimization function. This method introduces a huge set of training data base which includes the input low-resolution and a high resolution depth maps, self-similar scene structures, and an optional aligned intensity maps.

Cohen *et al.* [8] proposed the joint bilateral filtering by increasing the resolution of the depth map. The joint bilateral filter (JBF) is a modified version of the bilateral filter which takes the advantage of making low resolution image into high resolution and assumes that the edges between depth and color images are highly correlated.

Diebel *et al.* [7] uses the optimization schemes based on the Markov Random Field (MRF). A high resolution depth image is estimated by maximizing the posterior probability of each pixel value based on the MRF modeling of an image.

## II. PROPOSED METHOD

Ground truth depth map is obtained from the dataset which is mentioned in Section IV. Blurring and Subsampling them on even indices[10] creates a low resolution depth map and then an analytic image interpolation is performed, creating a depth map of the desired number of pixels that lacks high resolution details. It is then partitioned into overlapping low resolution patches. The low resolution patches are then processed in a raster scan order. For each low-resolution patch, a scaled mid band patch is generated. A search vector which is shown in fig 2 is constructed from pixels in the scaled mid band input patch and pixels in an overlap region of previously predicted high band patches. A nearest index vector to the search vector is located in a training database; the nearest index vector has an associated high band output patch which is shown in fig 3. The high band output patch is then combined with the interpolated low resolution patch to predict pixel values for the corresponding high resolution patch of the super-resolution depth map.

### a. Training Set Generation

Fig 2 shows the block diagram for generation of training database. The low resolution depth maps are first scaled up by a factor of two in each dimension by some conventional interpolation means, such as bilinear interpolation or bicubic spline interpolation, to form the interpolated low resolution depth map. In this paper, bicubic interpolation is used.

The interpolated low resolution depth maps are high-pass filtered, removing the lowest spatial frequency components, to obtain the mid band images. The interpolated low resolution depth maps are subtracted from the corresponding high resolution images, to obtain the high band depth maps, used as training data. The patches of the mid band depth

maps are contrast normalized, and the corresponding patches of the high band maps are contrast normalized by that same amount. The pixels in the low frequency patches and the overlapping section of high frequency patches are concatenated to form search vectors which is also used as training data. The overlapping section of the high frequency detail of depth map and the low resolution depth map are concatenated to form vectors. For finding the best match, sum of the squared differences (SSD) between each vector and the vectors in the training data is calculated. A threshold is set to compute the best match. Wherever, a minimum value of SSD is obtained, and then the corresponding high frequency detail from the training set is placed in that location. The training data carry assumptions about the structure of the visual world, and about image degradation when blurring and down-sampling. The high band depth map is then added to the interpolated low resolution depth map to form the high resolution estimated depth map, which is the output of the super-resolution algorithm.

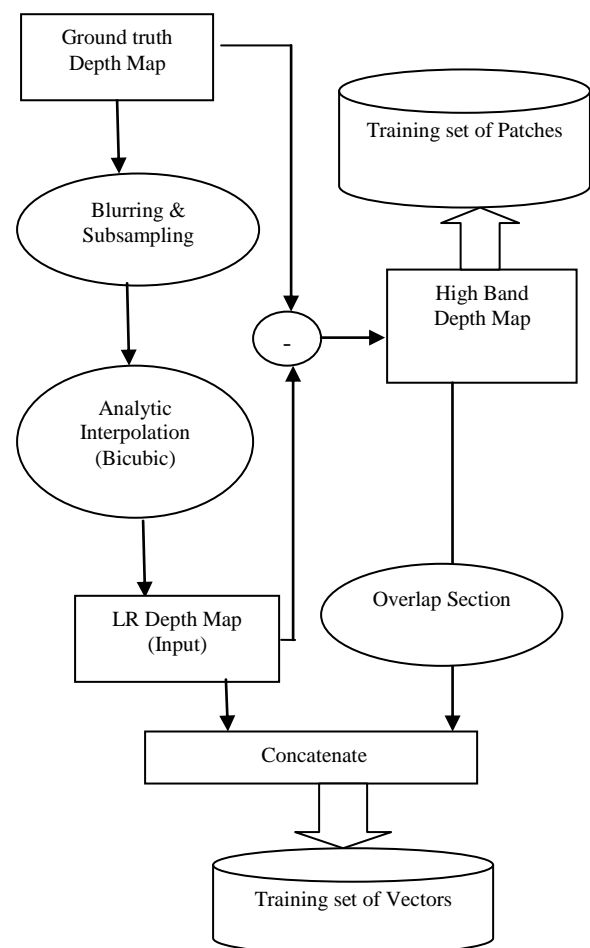


Fig 2: Training Set Generation

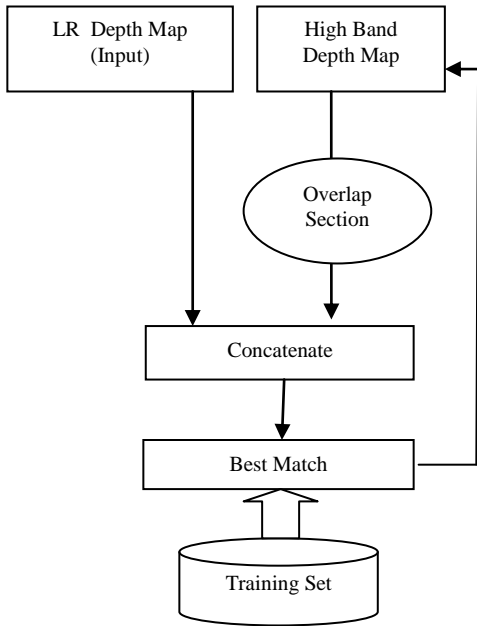


Fig 3: Proposed Method

### III. EXPERIMENTAL RESULTS

The proposed method is tested to various depth maps. The Middlebury Stereo Dataset [11], [9] provides the ground truth depth map. Also, different training ground truth depth maps are provided from this dataset. We have done patch based experiments on both color images and gray scale images. For color images, this is a popular method and in the case of depth maps, this is a rare area of work.

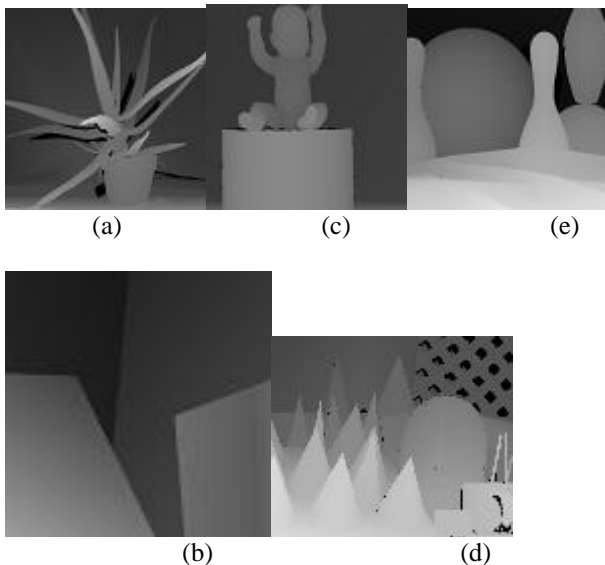
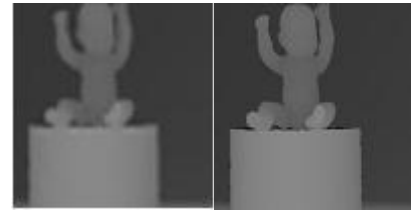
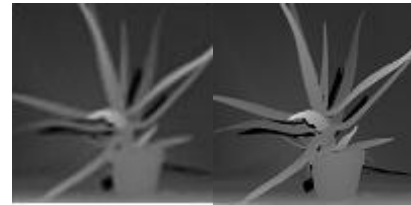


Fig4: (a),(b),(c) are used as test images and (c),(d),(e) used as training images in this method

Different training and test depth maps are used in this work which is shown in fig 4. Fig 5 shows the simulation results of test depth maps.



(a) (b)



(c) (d)



(e) (f)

Fig 5: Simulation results of test images :(a), (c), (e) shows the input low resolution depth maps and (b), (d ),(f) shows the super resolved output of the proposed method

### IV. PERFORMANCE EVALUATION

In super resolution technique, the quality of an image has been improved. This evaluation is mainly used to find the image quality and also gives the perceived image degradation. There are several metrics that can be used to measure. Quality assessment methods can be broadly classified into two: Full Reference Methods (FR) and No Reference Methods (NR). In FR methods, the quality of a test image is evaluated by comparing with a reference image. Some of the FR methods are PSNR, SSIM, and VIF. In NR method, there is no reference image. Only one NR method is evaluated which is entropy. The details of FR methods and NR method are described below and table 1 shows the PSNR values of bicubic interpolated and super resolved image and table 2 shows the values of SSIM, VIF and entropy respectively.

#### (a) Peak Signal- To- Noise Ratio

Mean squared error (MSE) is simply the squared error between a Super-Reconstructed image and the original high-resolution image. The Peak Signal-to-Noise Ratio (PSNR) is defined as a measuring of quality of reconstructed image and it is comparing with original image.

$$PSNR=20 * \log_{10}(MAX_I^2 / MSE) \quad (1)$$

where,  $MAX_I$  is the maximum possible pixel value of the image. When the pixels are expressed in 8 bits per sample, then  $MAX_I$  is 255. The PSNR expressed in decibels.

#### (b) Structural Similarity Index Measurement

The Structural Similarity Index is used to measure similarity between two images. SSIM was designed the better way for human visual system (HVS) processes structural information. SSIM measures the structure of an image, contrast and compare variance and covariance between the two images. It is expressed in ratios and its range is in between 0 and 1.

#### (c) Visual Information Fidelity

The concept of information fidelity measurement for image quality assessment is explained about an image information measure, and proposed the connections between image information and visual quality. The reference image is modelled as the output of a natural source that passes through the HVS channel. The information content of the reference image is considered as the mutual information between the input and output of the HVS channel. In the presence of an image distortion channel, the output of the natural source is distorted and thereby measuring this information as test image. These two information measures are combined to form a visual information fidelity measure which relates the visual quality relative to image information. It can be expressed as:

$$VIF = \text{Distorted Image Information} / \text{Reference Image Information}$$

Its value for a super resolved image is exactly unity.

#### (d) Entropy

It is defined as the maximum information quantity contained in an image. Larger values of entropy indicate that the super resolved image is perfect.

$$E = \sum p * \log_2 p \quad (2)$$

where, 'p' is the probability measurement in an image.

## V. CONCLUSION

In this work, a patch based super resolution algorithm is proposed for depth maps. This is a very simple and less time consuming compared to other methods and enhances the resolution of the depth image by using learning database. Simulation results show that the proposed method super resolves a depth image at high quality by using training database. Experimental results shows better values of PSNR, SSIM, VIF and entropy for super resolved depth map compared to the low resolution depth map.

TABLE 1: PSNR Values of different depth maps

Different DepthMaps	Bicubic Interpolation	Proposed Method
BABY	25.3356	25.3536
VENUS	25.1537	25.2190
ALOE	25.2893	25.4951

TABLE 2: SSIM, VIF and Entropy Values Of Different Super resolved Depth Maps

Different Depth Maps	SSIM	VIF	ENTROPY
BABY	0.9999	1.0000	5.3782
VENUS	0.9998	1.0000	6.4931
ALOE	0.9996	1.0000	6.1886

## REFERENCES

- [1] Amisha J. Shah, Suryakant B. Gupta, "Image Super Resolution-A Survey", IEEE International Conference on Emerging Technology Trends in Electronics, Communication and Networking, 2012.
- [2] Yanjie Li, TianfanXue, Lifeng Sun1, JianzhuangLiu, "Joint Example Based Depth Image Super Resolution", IEEE Conference on Pattern Recognition and Computer Vision (CVPR), 2012.
- [3] W.T. Freeman, T.R. Jones, and E.C. Pasztor, "Example-Based Super-Resolution", IEEE Computer Graphics and Applications, Vol. 22, pp. 56-65, 2002.
- [4] Kuster, C., Popa, T., Zach, C., Gotsman, C., Gross, M.: "Freecam: A hybrid camera system for interactive free-viewpoint video". In: Proceedings of Vision, Modeling, and Visualization (VMV). (2011).
- [5] Holz, D., Schnabel, R., Droschel, D., Stückler, J., Behnke, S.: "Towards semantic scene analysis with time-of-flight cameras". In: RoboCup International Symposium. (2010)
- [6] Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: "Real-time human pose recognition in parts from single depth images". In: CVPR. (2011).
- [7] Diebel, J. and Thrun, S., "An application of markov random fields to range sensing," NIPS, 291-298 (2005).
- [8] Kopf, J., Cohen, M. F. Lischinski, D., and Uyttendaele, M., "Joint bilateral upsampling," ACM Trans. On Graphics, vol. 26, no. 3, 96-96 (2007).

- [9] Jing Li Zhichao Lu Gang ZengRuiGanHongbinZha, " Similarity-Aware Patchwork Assembly for Depth Image Super-Resolution", IEEE Conference on Pattern Recognition and Computer Vision(CVPR),2014.
- [10] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low level Vision", Intl. J. Computer Vision, 40(1):25-47, 2000.
- [11] YongseokSoh, Jae-Young Sim, Chang-Su Kim, and Sang-Uk Lee, "Super pixel Based Depth Image Super Resolution" Proc. SPIE 8290, Three-Dimensional Image Processing (3DIP) and Applications II, 82900D (30 January 2012).
- [12] S.C. Park, M.K. Park, and M.G. KANG, "Super-Resolution Image Reconstruction: A Technical Overview", IEEE Signal Processing Magazine, Vol. 20, pp. 21-36, May 2003