Developing 3D Viewing Model from 2D Stereo Pair with its Occlusion Ratio

Himanshu Johari himanshujohari@gmail.com
Jaypee Institute of Information Technology
A-10,Sector-62 Noida-201 307
Uttar Pradesh, India.

Vishal Kaushik contact.vishal007@gmail.com
Jaypee Institute of Information Technology
A-10,Sector-62 Noida-201 307
Uttar Pradesh, India.

Pawan Kumar Upadhyay pawan.upadhyay@jiit.ac.in
Jaypee Institute of Information Technology
A-10,Sector-62 Noida-201 307
Uttar Pradesh, India

Abstract

We intend to make a 3D model using a stereo pair of images by using a novel method of local matching in pixel domain for calculating horizontal disparities. We also find the occlusion ratio using the stereo pair followed by the use of The Edge Detection and Image SegmentatiON (EDISON) system, on one the images, which provides a complete toolbox for discontinuity preserving filtering, segmentation and edge detection. Instead of assigning a disparity value to each pixel, a disparity plane is assigned to each segment. We then warp the segment disparities to the original image to get our final 3D viewing Model.

Keywords: 3D model, Stereo Pair, Depth Perception, Parallax Method, Occlusion, Disparity Map.

1. INTRODUCTION

3D models and 3D viewing is catching great pace in the field of computer vision due to its applicability in diverse fields of heath, aerospace, textile etc. We in our paper intend to propose a simplistic and a robust method of generating a 3D model given a pair of stereo images. We start by segmenting our image in color space by using the adaptive mean shift segmentation and edge detection. The segmented image hence reproduced has a unique label assigned to every segment. We then calculate the occluded regions for our stereo set, colour them black and let remaining ones be white and go on to calculate the occluded pixel ratio.

Next we try to calculate the pixel disparity by using the using the method of local matching in pixel domain. The recovery of an accurate disparity map still remains challenging, mainly due to the following reasons:
(i) Pixels of half occluded regions do not have correspondences in the other image, leading to incorrect matches if
not taken into account.
(ii) Images are disturbed because of sensor noise. This is especially problematic in poorly textured regions due to the
low signal-to-noise-ratio (SNR).
(iii) The constant brightness or color constraint is only satisfied under ideal conditions that can only roughly be met in practice.

We then assign a disparity plane to each segment by associating a segment with the median value of the disparity values of the pixels associated with it. The disparity that we get at this step filters out most of the noise that might hamper the performance of our final output i.e. the 3D model.

The disparity plot on a 3D mesh gives a pretty fair idea of the relative positions of the various objects in the images; But to improve the user understandability we try to regain the lost characteristics of the image by warping the intensity color values of the image on the disparity and plotting it on a 3D view. All the steps will be explained separately in the course of the paper.

The output we present in our paper should however, not be compared to outputs generated by more than two stereo images because of mainly two reasons :

i. A large portion of the 3D model remains occluded as we cannot estimate the shape or characteristics of the occluded portions.
ii. A complete 3D model cannot be generated without having covered all faces of any object, which requires a minimum of three cameras.

The idea can however be modified and improvised further to generate a complete 3D model provided we have the complete data set. We in this paper try to analyze the feasibility of our proposed method of generating a 3D model from a stereo pair.

2. LOCAL MATCHING IN PIXEL DOMAIN

2.1 Cost Estimation

Local matching requires to define a matching score and an aggregation window. The most common dissimilarity measures are squared intensity differences (SD) and absolute intensity differences (AD) that are strictly assuming the constant color constraint. Other matching scores such as gradient-based and non-parametric measures are more robust to changes in camera gain. In our approach[1] we are using a self-adapting dissimilarity measure that combines sum of absolute intensity differences (SAD) and a gradient based measure that are defined as follows:

$$C_{SAD}(x,y,d) = \sum_{(i,j)\in N(x,y)} I_1(i,j) - I_2(i+d,j) \qquad \text{-------(1)}$$

And

$$C_{GRAD}(x,y,d) = \sum_{(i,j)\in N_x(x,y)} |\nabla_x I_1(i,j) - \nabla_x I_2(i+d,j)| + \sum_{(i,j)\in N_y(x,y)} |\nabla_y I_1(i,j) - \nabla_y I_2(i+d,j)| \qquad \text{-----(2)}$$

where N(x, y) is a 3 × 3 surrounding window at position (x, y), N_x(x, y) a surrounding window without the rightmost column, N_y(x, y) a surrounding window without the lowest row, r_x the forward gradient to the right and r_y the forward gradient to the bottom. Color images are taken into account by summing up the dissimilarity measures for all channels.

An optimal weighting between C_{SAD} and C_{GRAD} is determined by maximizing the number of reliable correspondences that are filtered out by applying a cross-checking test (comparing left-to-right and right-to-left disparity maps) in conjunction with a winner-take-all optimization (choosing the disparity with the lowest matching cost). The resulting dissimilarity measure is given by:

$$C(x,y,d) = (1-\omega) * C_{SAD}(x,y,d) + \omega * C_{GRAD}(x,y,d) \qquad \text{--------(3)}$$

Though this is a simple and robust method there are now better enhanced methods of matching like MACO.[8]. A comparison between various feature extraction and recognition can be studied and appropriately used [9].

2.2 Horizontal Disparity Calculation

Using the method of cost estimation as explained above we calculate the disparity values for every pixel by subtracting the positions of correspondingly matched pixels in the two images. We assume a case of zero horizontal disparity throughout the course of our paper. To improvise on the results obtained we repeat the above step twice, firstly keeping the right image at the base and sliding the left image over it and vice versa the second time. The minimum among the two disparity values for every pixel is considered as the final disparity value for that corresponding pixel.

Figure 1, precisely shows our result of horizontal disparity calculation for a given test case of images.

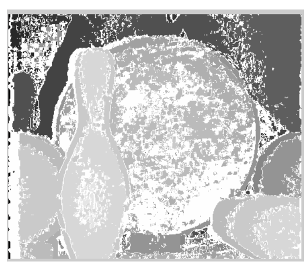

Figure1. Disparity Map

3. OCCLUSION ESTIMATION

As discussed earlier one of the major challenges in generating a 3D model lies in handling occlusion. We in our paper have used a simplistic method to show the half occluded regions i.e. those regions which do not have a corresponding match in the other image.

This method uses the disparity maps generated by the two iterations discussed in the Chapter 2.2. The disparity maps are subjected to scaling (to optimize our results) and appropriate thresh holding is done to uniquely identify the occluded protions.

The extent of half occlusion can be estimated by the absolute occlusion ratio, which is given as :

Aocc. = Nocc / (x * y)

Here, Nocc is the total number occluded pixels and ‘x’ and ‘y’ represent the dimension of the image matrix. The multiplication of the x and y dimensions of the matrix give us the total number of pixels in the image. Figure 3, shows the identified half occluded regions. The portions in black are the identified occluded pixels.

Figure 2. Occluded Regions

4. COLOUR SEGEMENTATION

a. Adaptive Mean Shift Segmentation and Edge detection

There are many ways of segmenting an image in an like Colour Histograms[7], JND histograms[6] etc. In our color image segmentation algorithm a five-dimensional feature space was used [2]. The color space was employed since its metric is a satisfactory approximation to Euclidean, thus allowing the use of spherical windows. The remaining two dimensions were the lattice coordinates. A cluster in this 5D feature space thus contains pixels which are not only similar in color but also contiguous in the image.

The quality of segmentation is controlled by the spatial hs, and the color hr, resolution parameters defining the radii of the (3D/2D) windows in the respective domains. The segmentation algorithm has two major steps. First, the image is *filtered* using mean shift in 5D, replacing the value of each pixel with the 3D (color) component of the

5D mode it is associated to. Note that the filtering is discontinuity preserving. In the second step, the basins of attraction of the modes, located within -'9;:_< in the color space are recursively *fused* until convergence. The resulting large basins of attraction are the delineated regions, and the value of all the pixels within are set to their average. See [3] and [4] for a complete description and numerous examples of the segmentation algorithm. It is important to emphasize that the segmenter processes gray level and color images in the same way. The only difference is that in the former case the feature space has three dimensions, the gray value and the lattice coordinates.

The mean shift based color image segmentation is already popular in the computer vision community and several

implementations exist. Along with image segmentation the open source EDISON we have used also does gradient based edge detection in the image. However, using the gradient magnitude for decisions causes a well known deficiency, sharp edges with small magnitudes can be detected only at the expense of allowing a large amount of edge clutter. A recently proposed generalization of the gradient based edge detection procedure eliminates this trade-off [5].

The result for the above mentioned process is shown in Figure 3. The code was written in MATLAB and the functions of C++ EDISON code were called using a MEX file.

Figure 3. Segmented Image

5. SNR CALCULATION:

Signal-to-noise ratio (often abbreviated SNR or S/N) is a measure used in to quantify how much a signal has been corrupted by noise. It is defined as the ratio of signal power to the noise power corrupting the signal. A ratio higher than 1:1 indicates more signal than noise. While SNR is commonly quoted for electrical signals, it can be applied to images also.

An alternative definition of SNR is as the reciprocal of the coefficient of variation, i.e., the ratio of mean to standard deviation of a signal or measurement.

SNR = μ/σ --------(4)

where μ is the signal mean or expected value and σ is the standard deviation of the noise, or an estimate thereof. Notice that such an alternative definition is only useful for variables that are always positive (such as photon counts and luminance). Thus it is commonly used in image processing, where the SNR of an image is usually calculated as the ratio of the mean pixel value to the standard deviation of the pixel values over a given neighborhood. Sometimes SNR is defined as the square of the alternative definition above.

We have accordingly calculated the SNR for all segments before filtering; some of them are shown in Table1.

Table 1: SNR for various Segments

1.2903	2.6692	3.3551	4.0759	5.3434	7.7233
8.1880	9.3920	9.7224	10.3752	11.2215	12.1938
13.5314	14.5834	15.8698	16.4618	20.2655	21.1895
22.9675	24.9660	25.3624	26.3882	27.8676	28.9518
29.5893	30.6746	31.1829	32.6312	33.1157	34.4666
35.5848	36.0835	37.5370	38.9565	39.7654	40.7948
41.9086	42.3569	43.9115	44.3283	45.5586	46.4970

48.0103	50.2326	55.8950	61.4123	63.9514	65.2723
67.6221	68.3737	70.7674	72.1460	92.3039	93.4421
109.4394	112.9547	114.5421	123.0981	127.1221	130.4643
135.1333	139.1510	144.1111	156.1265	199.2787	254.2282

After this we try to reduce the noise by applying a median filter on each corresponding segment. Hence we assign each segment with its corresponding median disparity value. By the end of filtering we try to achieve a SNR ratio of infinity for each segment. This process is better explained in 5.

6. DISPARITY SEGMENT ASSIGNMENT

The purpose of this step is to increase the accuracy of the disparity plane set by repeating the plane fitting for grouped regions that are dedicated to the same disparity plane. This can be done in two simple steps.

Firstly, we find out the total pixels associated with a segment and their corresponding disparity values. Secondly, we find the median value of all the disparities in that segment and assign that disparity to the segment. This process makes the disparity map neater and also helps in reducing the SNR. The method gets rid of sudden impulses or large unwanted variations in the value of the disparity. Though the method may trade-off with the actual disparity values but it helps the final result of generating a 3D viewing model. The result is shown in Figure 4.

Figure 4. Filtered Disparity

7. DEPTH CALCULATION

In this module we try to calculate the depth of individual segments assuming a parallel camera case. We use the filtered disparity values to calculate the depth using the formula and figure shown below :

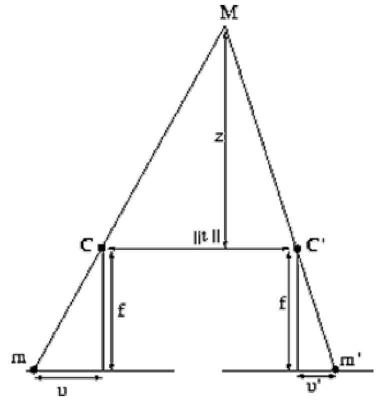

Figure 5

Given a scene point *M* and its two projection points *m* of coordinates (*u*,*v*) and *m*' of coordinates (*u*',*v*'), the *disparity value d* is defined as

$$d = u' - u \qquad \text{--------(5)}$$

Note that $v = v'$ as there is no vertical parallax between the two cameras. The *depth measure z* of *M* is related to the disparity value *d* as follows:

$$z = \frac{f\|\mathbf{t}\|}{d} \qquad \text{--------(6)}$$

8. GENERATNG A 3D VIEWING MODEL

a. 3D Plot of filtered disparity

This section deals with generating the final 3D view using the given pair of the images. The disparity calculated in the above step can be used directly to plot on a 3D mesh to get an estimate of the relative distances of various objects in the image. But, there still exists a major problem, i.e. the loss of original 3D intrinsic characteristics of the image in the 3D model

.

b. Warping image characteristics on disparity

Here we make an attempt to regain the original characteristics of the image. We warp the image intensity values from one of the input images onto the filtered disparity value matrix we got in Chapter 5. This method allows us to simultaneously plot the disparity and the intensity values on a 3D space, hence giving the user a fair idea of the relative depths of various objects identified in the images. The result is shown in Figure 6.

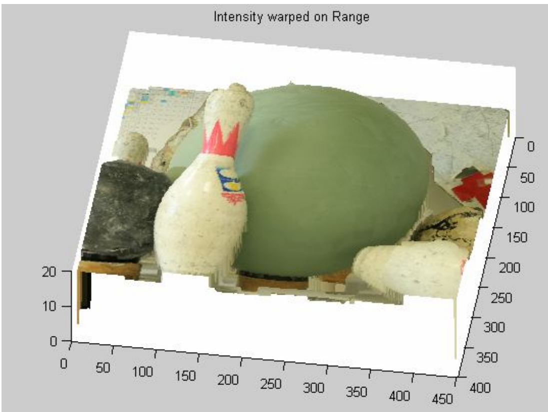

Fig 6(a)

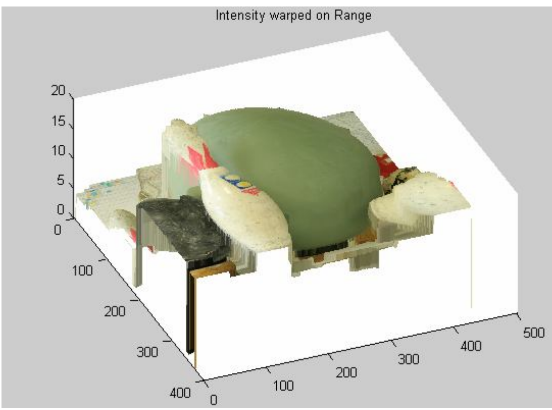

Fig 6(b)

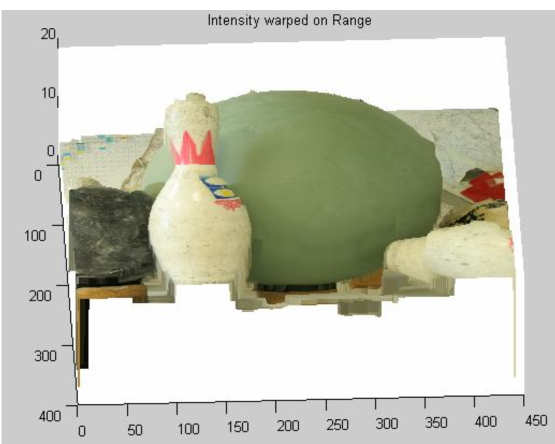

Fig 6(c)

Figure 6. Different Views of 3D Model

9. RESULTS

We have tested our proposed algorithm on a large set of stereo pairs. We have taken stereo pairs from the website of middlebury.edu. The occlusion ratios for different images are shown in Table2 and some other sample outputs are given in Figures 7(b),8(b),9(b) and 10(b).

Table 2 : Occlusion Ratios of test cases

Sample Pair	**Occlusion Ratio**
Bowling	0.338
Aloe Plant	0.1801
Baby	0.2223
Pots	0.3939

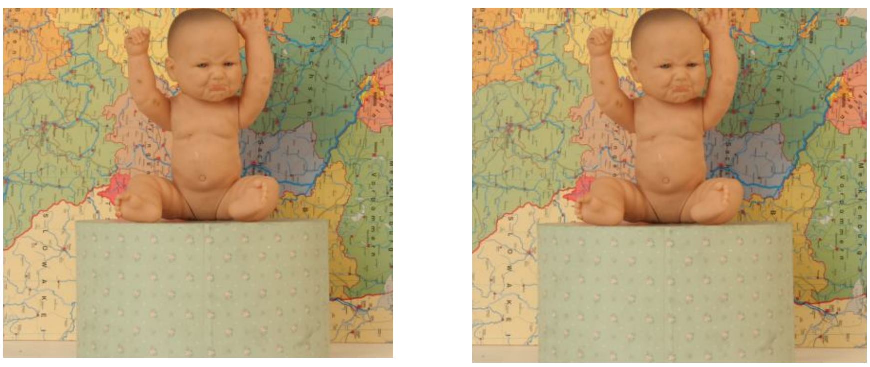

Fig 7(a) “Baby” stereo pairs (right and left)

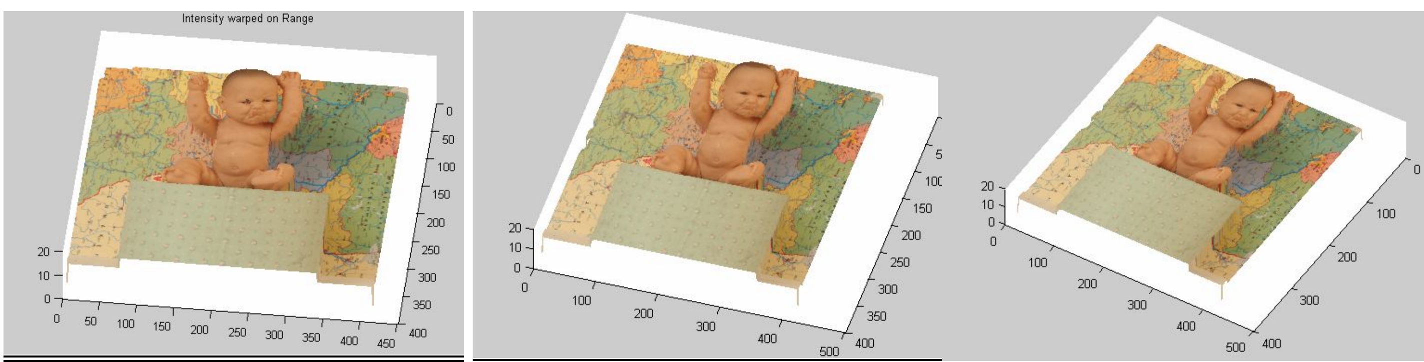

Fig 7(b) Some 3D model views of Stereo Images in 7(a)

Fig 8(a) “Pots” stereo pairs (right and left)

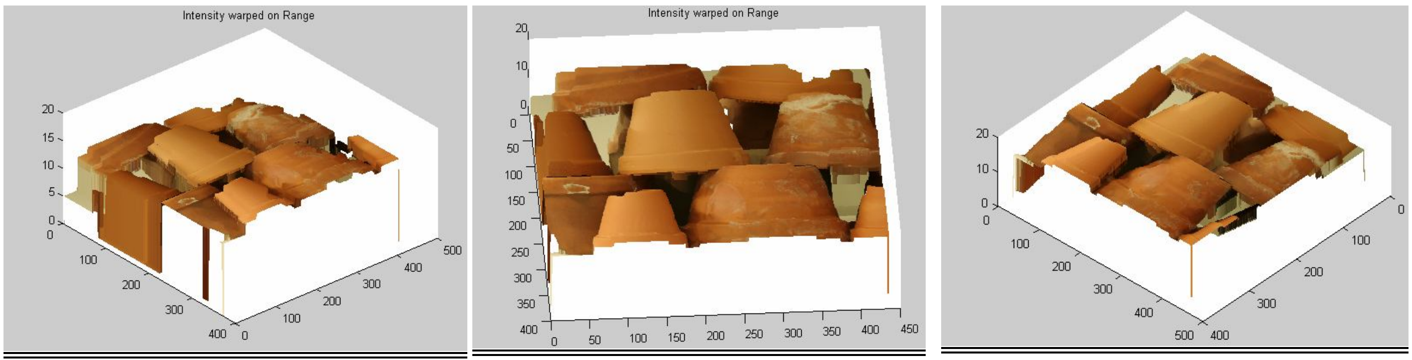

Fig 8(b) Some 3D model views of Stereo Images in 8(a)

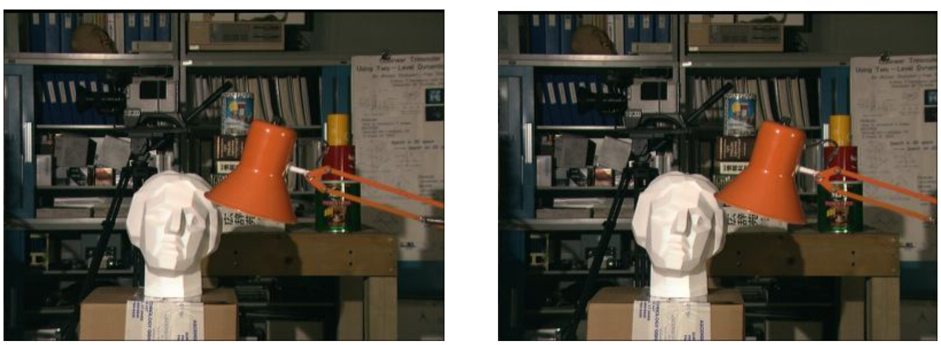

Fig 9(a) “Room” stereo pairs (right and left)

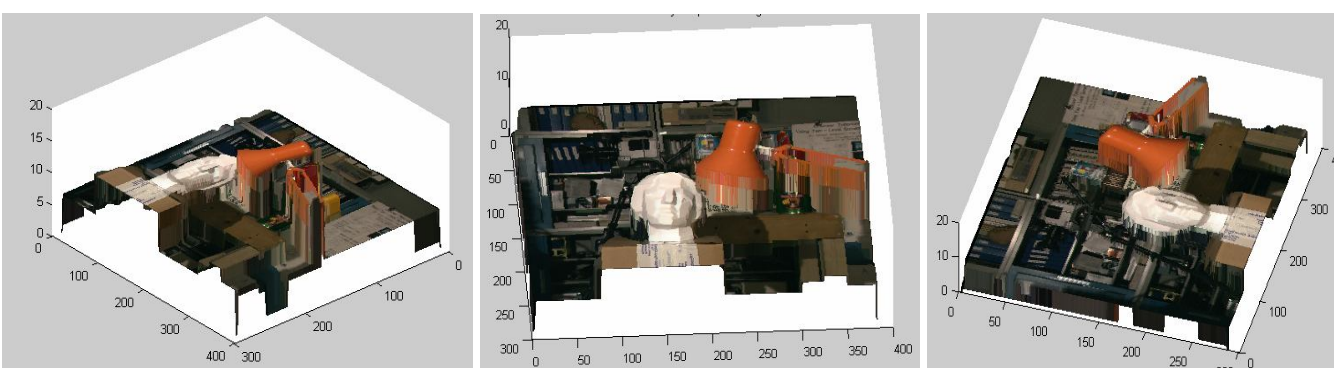

Fig 9(b) Some 3D model views of Stereo Images in 9(a)

Fig 10(a) “Aloe plant” stereo pairs (right and left)

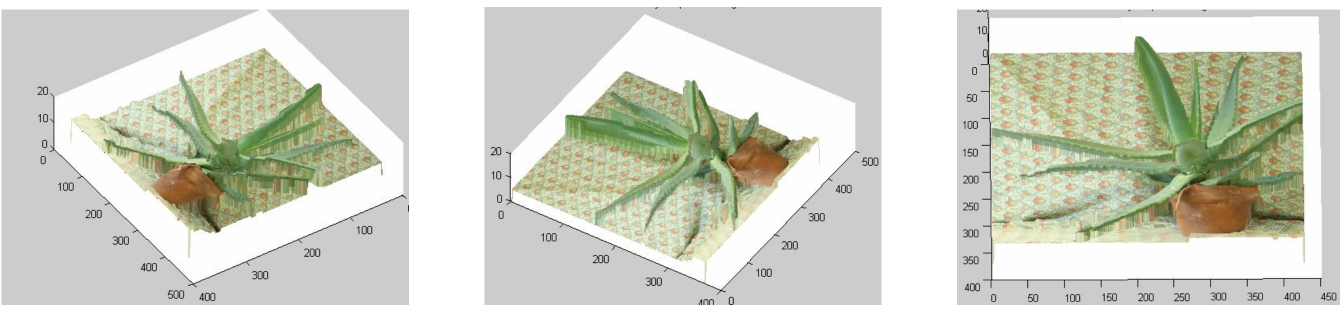

Fig 10(b) Some 3D model views of Stereo Images in 10(a)

10. CONCLUSION AND FUTURE WORK :

As it is clear from our result set that our proposed method works well for all set of stereo pairs. Our output set does not depend on the type of image and works equally well for grayscale and colored images. The number of objects is also not a constraint, just the occlusion ratio increases as the number of objects in the image increases. Our approach can be further used in various applications like :

- ‘Virtual catwalk’ to allow customers to visualize themselves in clothing prior to purchasing such goods on-line via the Internet.
- Potential to revolutionize autonomous vehicles and the capabilities of robot vision systems

11. REFERENCES

FOR JOURNALS:

[1] Andreas Klaus, Mario Sormann and Konrad Karner : Segment Based Stereo Matching using Belief Propagation and Self Adapting Dissimilarity Measure, VRVis Research Center, 8010 Graz, Austria, 2006.

[2] Christopher M. Christoudias, Bogdan Georgescu and Peter Meer : Synergism in Low Level Vision**,** Rutgers University, 2002.

[3] D. Comaniciu and P. Meer. Mean shift analysis and applications. In 7th international Conference on computer vision, pages 1197-1203, Kerkyra, Greece, September 1999.

[4] D. Comaniciu and P. Meer. Mean shift:A robust approach toward feature space analysis. IEEE Trans, Pattern Anal, Machine Intell,24,May 2002.

[5] T. Pavlidis and Y. T. Liow. Integrating region growing and edge detection. *IEEE Trans. Pattern Anal. Machine Intell.*, 12:225–233, 1990.

[6] Kishore Bhoyar and Omprakash Kadke, Color Image **Segmentation** Based On JND Color Histogram, International Journal of Image Processing(IJIP), Volume(3), Issue(6), 2010

[7] M.Swain and D. Ballard,"*Color indexing*", International Journal of Computer Vision, Vol.7, no.1,1991.

[8] XiaoNian Wang and Ping Jiang, “Multiple Ant Colony Optimizations for Stereo Matching”, International Journal of Image Processing(IJIP), Volume(3), Issue(5), 2010

[9] Luo Juan and Oubong Gwun, “A Comparison of SIFT, PCA-SIFT and SURF” International Journal of Image Processing(IJIP), Volume(3), Issue(5), 2010

FOR CONFERENCES:

[10] D. Comaniciu and P. Meer. Mean shift analysis and applications. In 7th international Conference on computer vision, Kerkyra, Greece, September 1999.

FOR BOOKS:

Gonzaleez and Woods, Digital Image Processing