Depth-Map Generation by Image Classification

Y V S HARISH - 20171402 , K S S VARMA - 20171406

***Abstract*—This project presents a technique to estimate depth information from a single input image. The proposed method is based on an image classification technique which is able to classify digital images (also in Bayer pattern format) as indoor, outdoor with geometric elements or outdoor without geometric elements. Using the information collected in the classification step, a suitable depth map is estimated. The implementation is fully unsupervised and is used to generate depth map from a single view of the scene, requiring low computational resources.**

I. Introduction :

3D images have become more and more popular in everyday life (3D games, PCs video technologies, 3D CAD systems, etc.) . Several techniques have been proposed to convert existing 2D images to 3D images. In many cases these techniques are semi-automatic : in a suitable operator identifies objects for depth placement in the 2D image while in a special effects artist guides the generation of depth maps using a Machine Learning Algorithm. The implemented method, based on the motion of the objects relative to the camera, have been proposed to calculate depth maps by estimating and analyzing optical flow. Another class of algorithms uses focus and defocus information. This project is based on a novel image classification technique able to classify digital images (also in Bayer pattern format) as indoor, outdoor with geometric elements or outdoor.

The method is well suited for real-time application **The input image is processed by the following steps:** :

1) Bayer to approximated-RGB color conversion.(In our case the input is already a RGB image.)
2) Color-based segmentation.
3) Rule-based regions detection to find specific areas (e.g. sky, land, mountain, etc.);
4) Image classifications to discriminate between outdoor with or without geometric elements and indoor images.
5) Approximated depth map estimation.
6) Experimental results are given in Section 5. Section 6 closes the paper tracking directions for future works.

II. Color based Segmentation

The color-based segmentation identifies chromatically homogeneous regions. We use the segmentation technique namely the mean shift algorithm able to group together pixels basing on their likeness. It generates a color segmented image in RGB format where the chromatic values of each identified region, are directly related to the original chromatic values. The resolution of the segmentation can be selected among: Under segmentation, Over segmentation and Quantization. For our purposes the Under segmentation has been chosen: the homogeneity is defined with wide range so that only the predominant colors are extracted from the original RGB image. Moreover, the borders of the regions in an image correctly under segmented correspond to the predominant edges in the image.

We have implemented Mean shift algorithm and there are threshold values in the code which can be modified to get different segmentes images .

Fig. 1. Our Image - Actual image

Fig. 2. Our Image - Segmented Image

Fig. 3. Other image(1) - Actual Image and Segmented Image

Fig. 4. Other image(2) - Actual Image and Segmented Image

III. Region Detection

The identification of semantic regions in a generic image is a crucial step needed to obtain a robust image classifier.The semantic region detection can be based on color-based rules aimed to characterize specific regions such as: Sky, Farthest Mountain, Far Mountain, Near Mountain, Land and Other. These semantic regions are typically present in landscape/outdoor images.

The regions detection is obtained as follows:

1) 5x5 median filter applied to the under segmented image
2) RGB to HSI color conversion
3) Image regions detection by color-based rules
4) 5x5 median filter applied to each detected region

The image regions are detected using a set of color-based rules taking into account typical chromatic correspondence between intensities values of R, G, B, H and I. For example given a pixel (x,y) it belongs to a Sky region if the following conditions are satisfied:

1) I(x,y) >0.65 AND B(x,y) >= 160 AND G(x,y) >= 70 AND B(x,y)+15 >= G(x,y) AND B(x,y)+15 >= R(x,y)

It belongs to a far mountain if the following conditions are satisfied:

1) I(x,y) >0.1 AND (B(x,y)>=20 AND B(x,y)<=160) AND (G(x,y)>=15 AND G(x,y)<=255) AND (B(x,y)>=G(x,y) AND B(x,y)>=R(x,y))
2) I(x,y) >0.4 AND I(x,y)<0.8 AND (R(x,y)>=80 AND R(x,y)<=160) AND (G(x,y)>=80 AND G(x,y)<=160) AND (B(x,y)>=80 AND B(x,y)<=160)

It belongs to a near mountain if the following conditions are satisfied:

1) I(x,y)>0.45 AND B(x,y)<=100 AND G(x,y)<=255 AND R(x,y)>=100 AND (R(x,y)>=B(x,y) AND R(x,y)>=G(x,y))
2) (I(x,y)>0.14 AND I(x,y)<0.65) AND B(x,y) <= 120 AND G(x,y)<=120 AND R(x,y)<=120 AND (G(x,y)+10>=B(x,y) AND G(x,y)+10>=R(x,y))

It belongs to a land if the following conditions are satisfied:

1) I(x,y)>0.5 AND (R(x,y)>100 AND R(x,y)<=200) AND (G(x,y)>100 AND G(x,y)<=190) AND (B(x,y)>135 AND B(x,y)<=180)
2) I(x,y)>0.4 AND B(x,y)<=100 AND R(x,y)<=200 AND G(x,y)>=B(x,y) AND G(x,y)>=R(x,y)

Starting from the under segmented image, once the regions detection are univocally located a grey level image is generated, where to each region is assigned a specific values according to the following rule:

Gray(Other)>Gray(Land)>Gray(Near-Mountain)>Gray(Far-Mountain)>Gray(Farthest-Mountain)>Gray(Sky).

The closest regions to the viewer are labelled with grey level bigger than farthest regions.

Please Note :: The optimal threshold values are designed by us and have not been provided in the original paper , therefore with more chopping and changing we might get better or worse image detections.

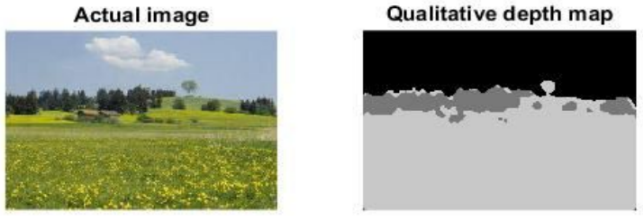

Fig. 5. Outputs

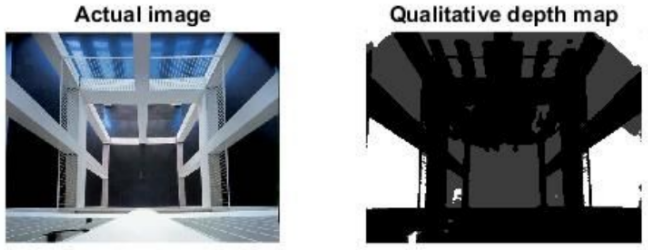

Fig. 6. Outputs

Fig. 7. Outputs

IV. Image Classification

To obtain a reliable depth map using a single view of the input scene taking into account only the semantic category of the input image, requires a robust image classifier. Our

TABLE I
LABELS CORRESPONDING TO REGIONS

Region	Labels
Sky	s
Farthest Mountain	m
Far Mountain	m
Near Mountain	m
Land	l
Other	x

preliminary results have been obtained focusing the classifier to the following categories: Outdoor/Landscape, Outdoor with geometric elements, and Indoor.

The proposed technique is based on the comparison of a group of N sample columns of the regions detections output with a set of typical sequences of a landscape. This approach is able to classify, based on statistically assumption, the real semantic category of a landscape image. A sequence is a string containing a collection of labels. Each label corresponds to a specific region as reported in Table , detected during the vertical scan of a column. To add a label in a sequence, a sufficient number of pixels have to be detected. Some typical sequences are: s, sm, sl, sml, m, ml, l, sms, smsl, msl, mls and ls.

The main steps of the algorithm are:

1) Sequences and jumps detection for each sample column. A jump is the number of regions encountered in the examined column.
2) Each sequence is compared to the set of typical sequences. If the sequence is recognized and the jumps number is smaller than a threshold `J_B`, then the value N_1 is increased, where N_1 represents the number of accepted sequences. If the sequence isnt a typical landscape sequence or if the jumps number is bigger than `J_B` then the sequence is rejected.
3) Final classification. The image is classified as Outdoor if the value of N_1 is bigger than R_1N , where N is the number of analyzed sequences and R_1 is a threshold in [0,1]. Otherwise if the number of sequences with the first region Sky is bigger than R_2N, where R_2 is another threshold in [0,1] the image is classified as Outdoor with geometric appearance else it is classified as Indoor.

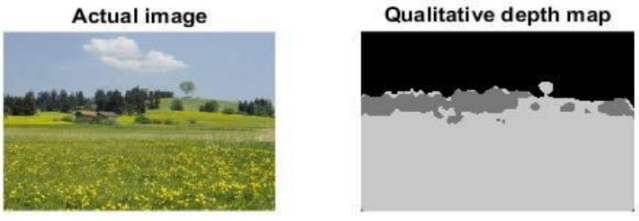

Fig. 8. This image is classified as outdoor without geometric elements or landscape.

Fig. 9. Above image is classified as outdoor with geometric elements .

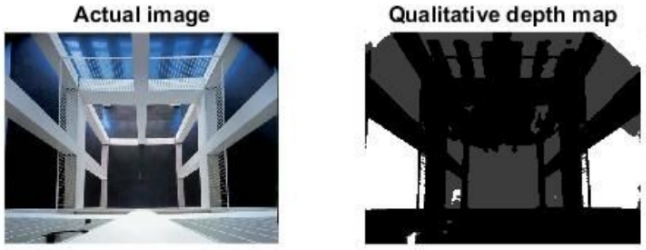

Fig. 10. Above image is classified as indoor .

Fig. 11. Above image is classified as outdoor without geometric elements .

V. GEOMETRIC DEPTH MAP GENERATION : VANISHING POINT DETECTION

If the input image is classified as Outdoor without geometric elements, the lowest point in the boundary between the region $A = Land \bigcup Other$ and the other regions is located. Using such boundary point (x_b ,y_b), the coordinates of the VP point are fixed to (W/2,y_b), where W is the images width.

When the image is classified as Outdoor with geometric appearance or Indoor, the algorithm is composed of the following steps:

1) Edge detection using a 3x3 Sobel masks. The resulting images, I_{S_x} and I_{S_y} , are then normalized and convert into a binary image I_E, eliminating redundant information.
2) Noise reduction of I_{S_x} and I_{S_y} using a standard low-pass filter 5x5.
3) Detection of the straight lines, using I_{S_x} and I_{S_y}, passing through each edge point of I_E
   $m(x,y) = I_{S_{y(x,y)}}/I_{S_{x(x,y)}}$
   $q(x,y) = y - m(x,y).x$
   where m is the slope and q is the intersection with the y-axis of the straight line.
4) Each pair of parameters (m,q) is properly sampled and stored in an accumulation matrix:
   $ACC[m,q] = ACC[m,q] + 1$

5) Higher values correspond to the main straight lines of the original image.
6) Computation of intersection between each pair of main straight lines.
7) The VP is chosen as the intersection point with the greatest number of intersections around it, while the vanishing lines detected are the main straight lines passing close to VP.

Fig. 12. Output - Lines meeting at the VP

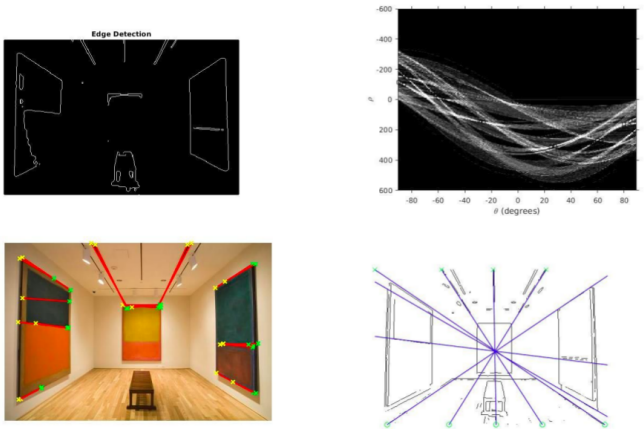

Fig. 13. Process for creation of Vanishing lines and Vanishing Points.

VI. Geometric Depth map generation : Gradient Plane generation

During this processing step, the position of vanishing point (relatively to the image) and the slopes of vanishing lines are analyzed. Five different cases can be distinguished:

1) $X_{vp} <= 0 AND (H-1/W-1) * X_{vp} < Y_{vp} < -(H-1/W-1) * X_{vp} + H - 1, LeftCase.$
2) $X_{vp} >= W-1 AND -(H-1/W-1) * X_{vp} + H - 1 < Y_{vp} < (H-1/W-1) * X_{vp}, RightCase.$
3) $Y_{vp} <= 0 AND (W-1/H-1) * Y_{vp} <= X_{vp} <= (W-1/H-1) * (H-1-Y_{vp}), UpCase.$
4) $Y_{vp} >= H-1 AND (W-1/H-1) * (H-1-Y_{vp}) <= X_{vp} <= (W-1/H-1) * Y_{vp}, DownCase.$
5) $0 < X_{vp} - 1 AND 0 < Y_{vp} - 1, InsideCase.$

where (X_{vp}, Y_{vp}) are the VP coordinates H, W are the image dimensions.

VII. Geometric Depth map generation : Depth gradient Assignment

A grey level (corresponding to a depth level) is assigned to every pixel belonging to depth gradient planes. Two main assumptions are used:

1) Higher depth level corresponds to lower grey values.
2) The vanishing point is the most distant point from the observer (this assumption is almost always true).

.

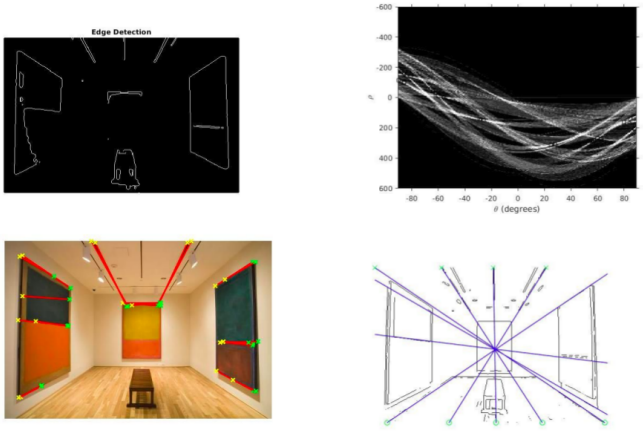

Fig. 14. Gray level assignments

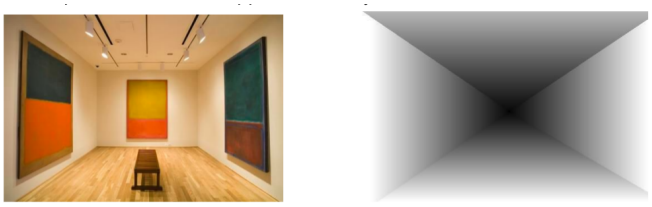

Fig. 15. Output of gradient assignment.

Fig. 16. Output

Fig. 17. Output

Fig. 18. Output

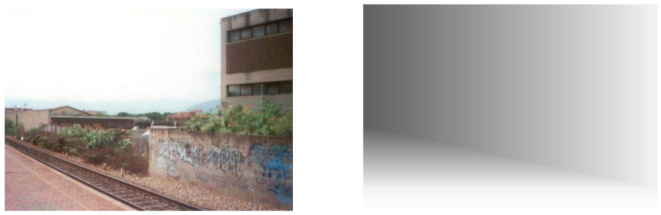

Fig. 19. Output

VIII. Geometric Depth map generation : Fusion of geometric and qualitative depth map

Now,we generate the final depth map using the above two depth maps according to the following conditions: Let $M_1(x,y)$ be the Geometric Depth Map, $M_2(x,y)$ be the Qualitative Depth Map, the Fusion depends on the image classification:

1) If the image belongs to the indoor category then M(x,y) coincides with $M_1(x,y)$:
   $M(x,y) = M_1(x,y) \ \forall(x,y)$:
   $0<=x<=W-1, 0<=y<=H-1$
2) If the image is classified as Outdoor with absence of meaningful geometric components (landscape) then the image M(x,y) is obtained as follows:
   a. $M(x,y) = M_1(x,y) \forall(x,y) \in Landand \forall(x,y) \in Other$
   b. $M(x,y) = M_2(x,y) \ \forall(x,y) \notin Landand \forall(x,y) \notin Other$

3) If the image is classified as Outdoor with geometric characteristics then the image M(x,y) is obtained as follows:
   a. $M(x,y) = M_2(x,y) \ \forall(x,y) \in Sky.$
   b. $M(x,y) = M_1(x,y) \ \forall(x,y) \notin Sky.$

IX. Results

Following images have results of the final pipeline onto images taken on our phones and other images .

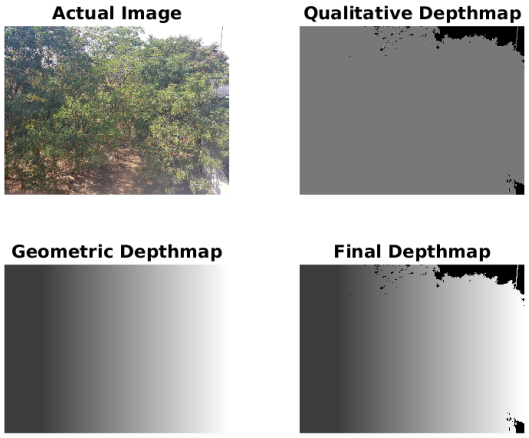

Fig. 20. Output

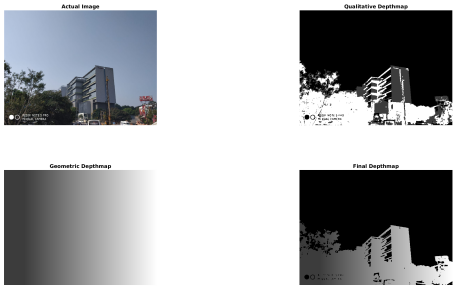

Fig. 21. Output

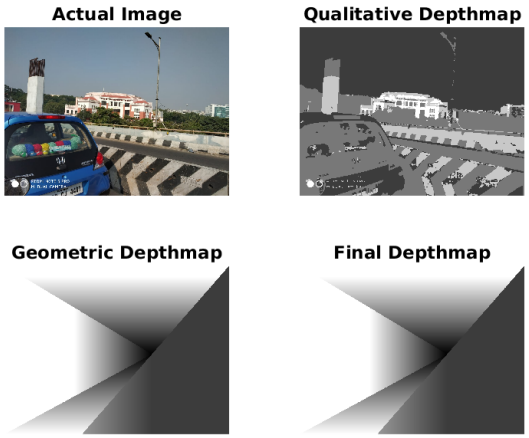

Fig. 22. Failed Output

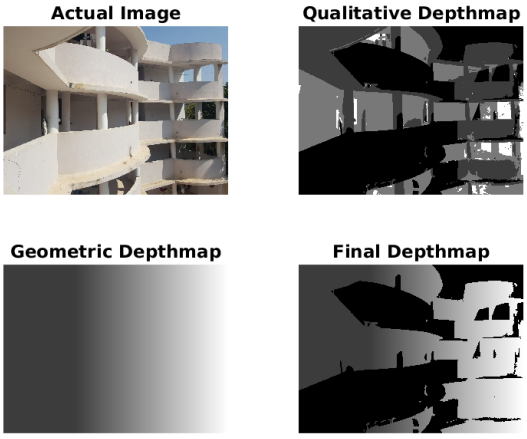

Fig. 23. Output

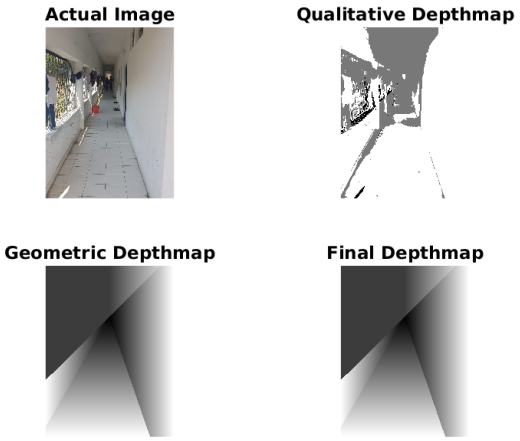

Fig. 24. Output

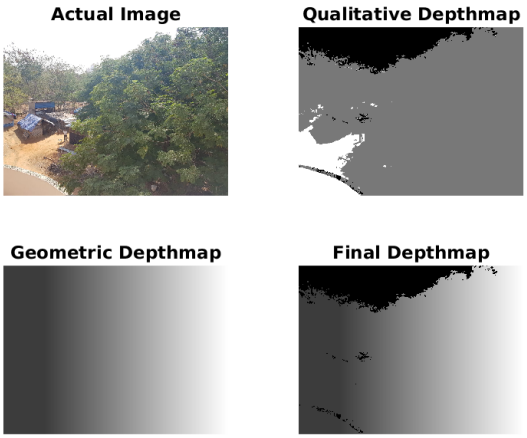

Fig. 25. Output

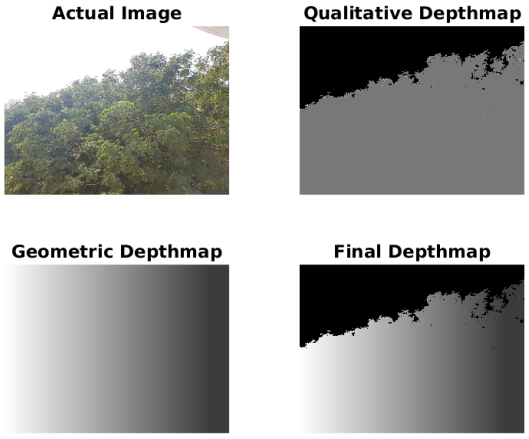

Fig. 26. Output

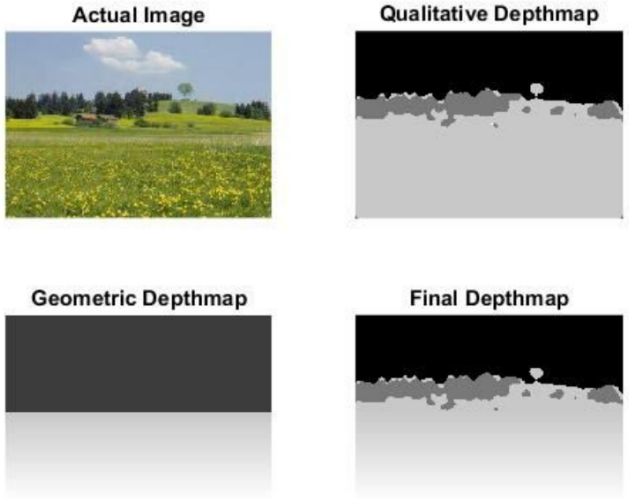

Fig. 27. Output

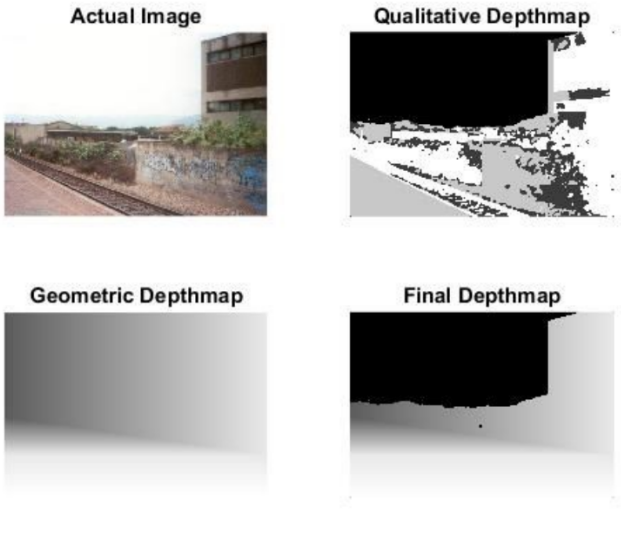

Fig. 28. Output

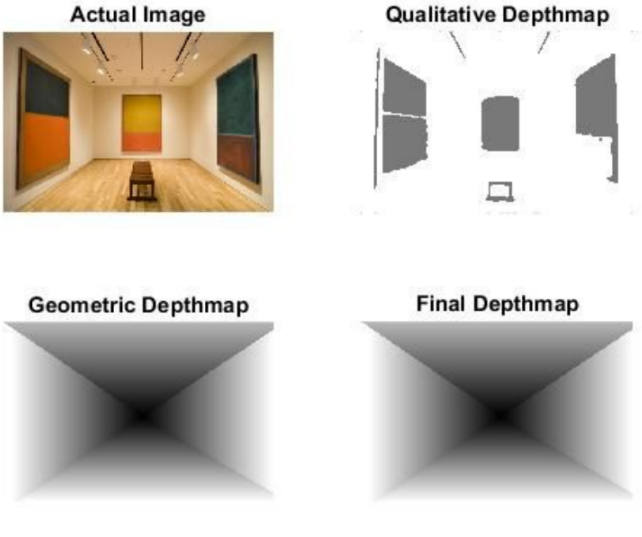

Fig. 29. output

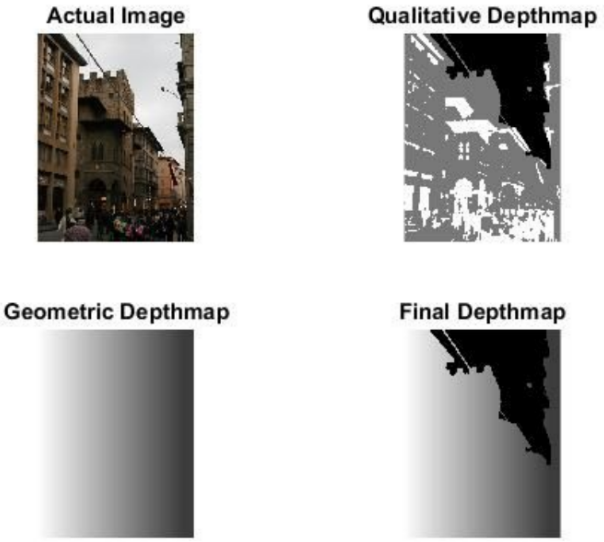

Fig. 30. Output

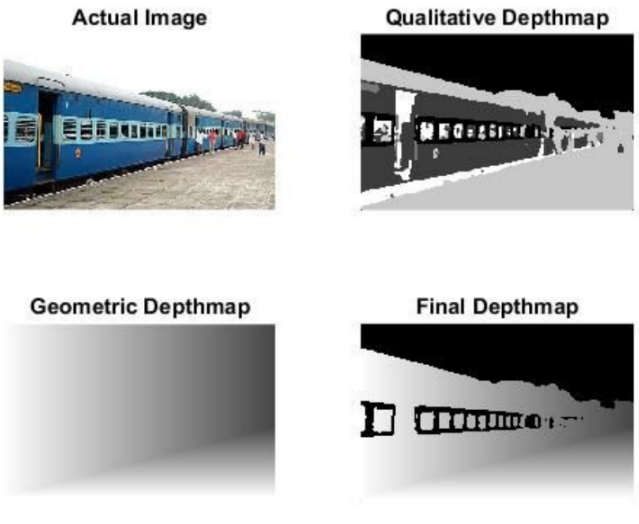

Fig. 31. output

X. LIMITATIONS :

1) We dont have the training data set so we cannot get the appropriate color values which can classify all the regions correctly for all types of images and Geometric Depth Map heuristics.
2) Similarly we cannot get the appropriate k1 and k2 values which could classify the image correctly for any image.

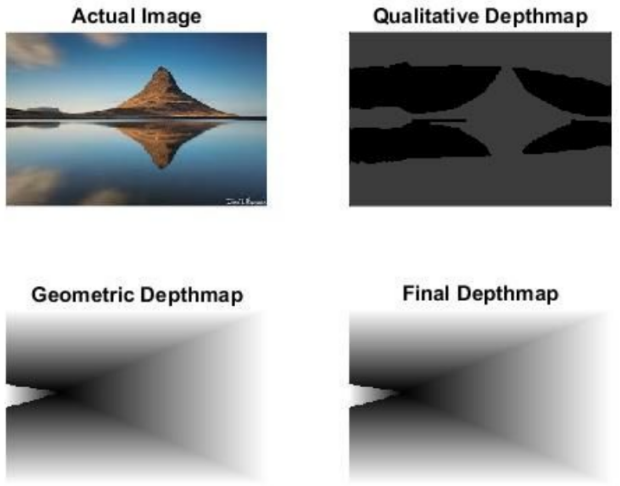

Fig. 32. For this image, we cant segment the water clearly and also some part of sky region is not detected as sky region. Also, the Geometric depth map is not correct because we dont have huge train data to estimate the proper heuristics.

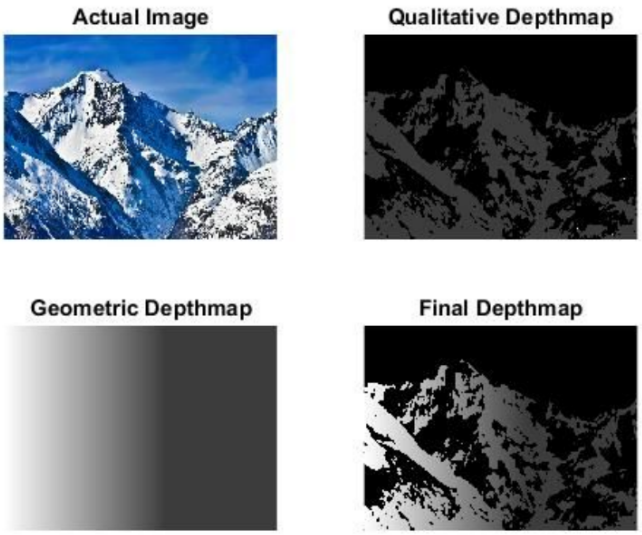

Fig. 33. In this image, regions arent properly segmented.