

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/4091618>

# 3D stereoscopic image pairs by depth-map generation

Conference Paper · October 2004

DOI: 10.1109/TDPVT.2004.1335185 · Source: IEEE Xplore

CITATIONS

84

READS

1,587

4 authors:



**Sebastiano Battiato**  
University of Catania

371 PUBLICATIONS 3,312 CITATIONS

[SEE PROFILE](#)



**Alessandro Capra**  
STMicroelectronics

37 PUBLICATIONS 255 CITATIONS

[SEE PROFILE](#)



**Salvatore Curti**  
STMicroelectronics

13 PUBLICATIONS 277 CITATIONS

[SEE PROFILE](#)



**Marco La Cascia**  
Università degli Studi di Palermo

125 PUBLICATIONS 2,570 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Vision for Robotics [View project](#)



RECfusion [View project](#)

# 3D Stereoscopic Image Pairs by Depth-Map Generation

S. Battiato<sup>a</sup>, A. Capra<sup>b</sup>, S. Curti<sup>b</sup>, M. La Cascia<sup>c</sup>

<sup>a</sup>Università di Catania - Dipartimento di Matematica ed Informatica – 95125 Catania, Italy

<sup>b</sup>STMicroelectronics - AST Catania Lab, Stradale Primosole 50 - 95121 Catania, Italy

<sup>c</sup>Università di Palermo - DINFO, Viale delle scienze - 90128 Palermo, Italy

<sup>a</sup>battiato@dmi.unict.it

<sup>b</sup>{alessandro.capra, salvatore.curti}@st.com

<sup>c</sup>lacascia@unipa.it

## Abstract

*This paper presents a new unsupervised technique aimed to generate stereoscopic views estimating depth information from a single input image. Using a single input image, vanishing lines/points are extracted using a few heuristics to generate an approximated depth map. The depth map is then used to generate stereo pairs. The overall method is well suited for real time application and works also on CFA (Colour Filtering Array) data acquired by consumer imaging devices. Experimental results on a large dataset are reported.*

## 1. Introduction

Actually to generate a 3D view starting from a single view image different tools are required. These are not fully automatic application and require expensive computational resources. This paper presents a single framework aimed to obtain the stereoscopic view avoiding the user interaction and reducing the computation complexity. Moreover the depth map generation step is able to work directly onto a Bayer pattern image (CFA) further reducing band, memory and complexity requirements. Alternatively a sub sampled image can be used.

This proposed technique is based on a novel algorithm [1] able to generate a depth map from a single image. The strength of method [3] is then used to reconstruct the left and right view. The two steps are pipelined so obtaining a single automatic 3D generation procedure.

In order to obtain the depth map, an image pre-processing is required. It is composed of the following steps:

- 1) Image classification ;
- 2) Vanishing lines and vanishing point extraction.

The stereoscopic pair image is then generated by calculating the parallax value ([8], [9], [10], [12]) of each object in the image extracting information only from the grey level depth map. The final left and right eye images give to the user a 3D perspective entertainment. The highlighted process is fully automatic and well suited for real-time application. Effectiveness of the proposed processing pipeline has been validated by an exhaustive set of experiments.

This paper is organized as follows: Section 2 resumes the used image classification method describing also the techniques to extract the vanishing lines and the vanishing point. Section 3 explains how the depth map is generated. In section 4 an overview of the Stereoscopic pair image generation is given. Experimental results are reported in Section 5. Section 6 closes the paper tracking directions for future works.

## 2. Image pre-processing

To generate the stereoscopic pair image (left and right view) the depth information of the objects inside the scene has to be estimated.

In order to obtain the depth information, a preliminary image pre-processing is applied to extract the relevant information from the input image. The image is first classified as: Outdoor/Landscape, Outdoor with Geometric Elements, Indoor (Figure 10). According to each specific class, the relevant vanishing lines and the related vanishing point are then selected.

## 2.1. Image classification

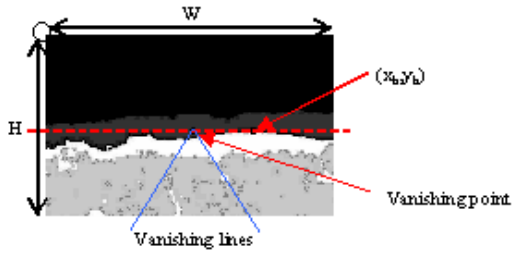
The main steps of the classification are summarised as follows:

- Semantic region detection (Sky, Farthest Mountain, Far Mountain, Near Mountain, Land and Other). A preliminary colour-based segmentation [7], which identifies chromatically homogeneous regions, helps to reduce incorrect region detection. To each detected region a fixed grey level is assigned.
- Comparison of  $N$  sampled columns of the semantic regions detection output with a set of typical strings containing allowed region sequences [1], [4], [5], [6].
- Final classification. The output of the previous step is used to classify the image according to heuristics [1].

## 2.2. Vanishing lines detection

The image classification result can also be used to properly detect some image features, like vanishing lines and related Vanishing Point (VP) [2].

If the input image is classified as *Outdoor without geometric elements*, the lowest point in the boundary between the region  $A = Land \cup Other$  and the other regions is located. Using such boundary point  $(x_b, y_b)$ , the coordinates of the VP are fixed to  $(W/2, y_b)$ , where  $W$  is the image's width. Moreover the method generates a set of standard vanishing lines (Figure 1).



**Figure 1- The boundary point and the resulting vanishing point for *Outdoor* images.**

When the image is classified as *Outdoor with geometric appearance* or *Indoor*, the VP detection is conducted as follows:

1. Edge detection using a 3x3 Sobel masks. The resulting images,  $I_{Sx}$  and  $I_{Sy}$ , are then normalized and converted into a binary image  $I_E$ , eliminating redundant information.

2. Noise reduction of  $I_{Sx}$  and  $I_{Sy}$  using a standard low-pass filter 5x5.
3. Detection of the main straight lines, using  $I_{Sx}$  and  $I_{Sy}$ , passing through each edge point of  $I_E$ , where  $m$  is the slope and  $q$  is the intersection with the  $y$ -axis of the straight line:

$$m(x, y) = \frac{I_{S_y}(x, y)}{I_{S_x}(x, y)} \quad (1)$$

$$q(x, y) = y - m(x, y) \cdot x \quad (2)$$

4. Each pair of parameters  $(m, q)$  is properly sampled and stored in an accumulation matrix (3) where higher values correspond to the main straight lines of the original image:

$$ACC[m, q] = Acc[m, q] + 1 \quad (3)$$

5. Computation of intersection between each pair of main straight lines.
6. 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 (Figure 8).

## 3. Depth-Map Generation

Taking into account the information collected in the pre-process analysis a series of intermediate steps are used to recover the final depth map. These steps can be resumed in:

- 1) gradient planes generation;
- 2) depth gradient assignment;
- 3) consistency verification of detected region;
- 4) final depth map generation.

In the following sections these steps are discussed.

### 3.1. Gradient planes generation

During this processing step, the position of the vanishing point in the image is analyzed. Five different cases can be distinguished, as shown in Figure 9, where  $X_{vp}$  and  $Y_{vp}$  are the vanishing point coordinates onto image plane and  $H$  and  $W$  the image size.

For each case a set of heuristics (Table 1), based on vanishing lines slope and origin of the vanishing lines onto image plane allow generating horizontal and/or vertical planes (gradient planes) used to gradually set the depth variation. Figure 9 shows how horizontal and/or vertical gradient planes can be located.

It is necessary that at least two vanishing lines are detected.

**Table 1: Rules to generate gradient planes**

<p><u>Left Case</u></p> <ul style="list-style-type: none"> <li>▪ <math>Xvp &lt;= 0</math></li> <li>▪ <math>(H/W) * Xvp &lt; Yvp &lt; -(H/W) * Xvp + H</math></li> </ul> <p><u>Right Case</u></p> <ul style="list-style-type: none"> <li>▪ <math>Xvp &gt;= W - 1</math></li> <li>▪ <math>-(H/W) * Xvp + H - 1 &lt; Yvp &lt; (H/W) * Xvp</math></li> </ul> <p><u>Up Case</u></p> <ul style="list-style-type: none"> <li>▪ <math>Yvp &lt;= 0</math></li> <li>▪ <math>(W/H) * Yvp &lt;= Xvp &lt;= (W/H) * (H - Yvp)</math></li> </ul> <p><u>Down Case</u></p> <ul style="list-style-type: none"> <li>▪ <math>Yvp &gt;= H - 1</math></li> <li>▪ <math>(W/H) * (H - 1 - Yvp) &lt;= Xvp &lt;= (W/H) * Yvp</math></li> </ul> <p><u>Inside Case</u></p> <ul style="list-style-type: none"> <li>▪ <math>0 &lt; Xvp &lt; W</math></li> <li>▪ <math>0 &lt; Yvp &lt; H</math></li> </ul> <p>where:</p> <p><math>(Xvp, Yvp)</math> = vanishing point coordinates;  <math>H, W</math> = image size</p>
---

### 3.2. 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:

- Higher depth level corresponds to lower grey values;
- The vanishing point is the most distant point from the observer (this assumption is usually true).

Figure 9 shows as in horizontal planes the grey level (depth level) is constant along the rows while in vertical planes it is constant along the columns. The depth level is approximated by a piece-wise linear function (Table 2) depending on slopes  $m1$  and  $m2$  of vanishing lines generating the depth gradient plane as shown in Figure 2.

**Table 2: Rules to generate the depth level**

<p>Horizontal gradient plane:</p> $D_{\min} = \min \left  \frac{y - q_1}{m_1} - \frac{y - q_2}{m_2} \right $ $D_{\max} = \max \left  \frac{y - q_1}{m_1} - \frac{y - q_2}{m_2} \right $ $0 \leq \frac{y - q_1}{m_1} \leq W - 1 \quad 0 \leq \frac{y - q_2}{m_2} \leq W - 1$ <p>Vertical gradient plane:</p> $D_{\min} = \min \left  (m_1 - m_2) \cdot x + q_1 - q_2 \right $ $D_{\max} = \max \left  (m_1 - m_2) \cdot x + q_1 - q_2 \right $ $0 \leq m_1 \cdot x + q_1 \leq H - 1$ $0 \leq m_2 \cdot x + q_2 \leq H - 1$ <p>where  <math>q_1</math> and <math>q_2</math> are the intersection with the boundaries image plane</p> <p>and</p> $D_1 = D_{\min} + \frac{D_{\max} - D_{\min}}{16}$ $D_2 = D_{\min} + \frac{D_{\max} - D_{\min}}{8}$ $D_3 = D_{\min} + \frac{D_{\max} - D_{\min}}{4}$
---

This choice is justified by the consideration that human vision is more sensible to deep variations of close objects than for far ones. Thus the deep levels have a slope that increases going from the closest position to the farthest one (VP).

Figure 3 shows an example of an output grey level image; it can be considered as a geometric depth map, because it is generated only by geometric information.

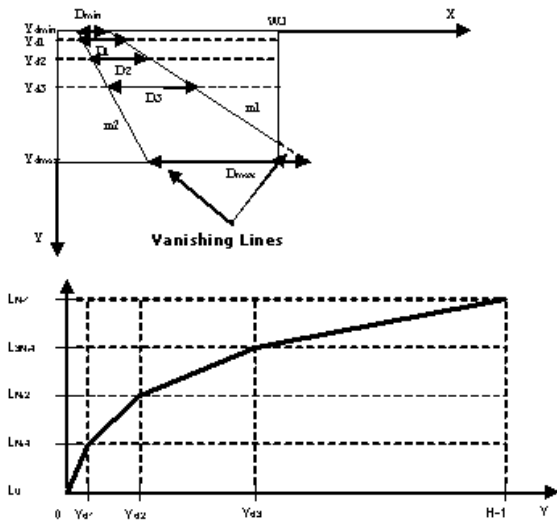


Figure 2 - Example of depth gradient assignment for a plane generated by two vanishing lines.

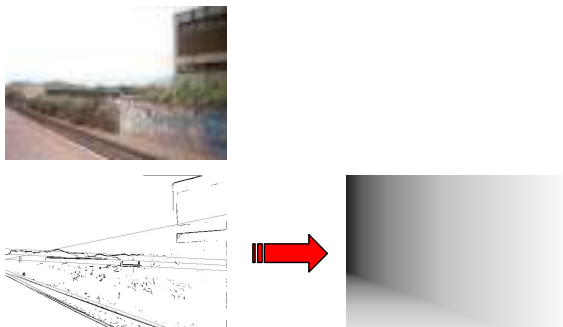


Figure 3- Example of *geometric depth map* generation.

### 3.3. Consistency verification of detected regions

The output image obtained by regions detection step (*qualitative depth map*) is analyzed to verify the consistency of the detected regions. In fact, the regions have been detected only by colour information. It is necessary, therefore, to analyze the positions, inside the image, of each region with respect to the others checking their dimensions. Using a set of heuristics, the columns of the image are properly scanned to produce some sequences of "regions" which are checked and, if necessary, modified for "consistency verification ". In

this way false regions are eliminated from the image (Figure 4).

For example, if between two regions of the image classified as Sky there is a different region (e.g. mountain, land or other), with a vertical size more than a fixed threshold, the second Sky region is recognised as false Sky region and is changed to the same type of the upper one.

Similar rules are used to detect the consistency for the others image regions.

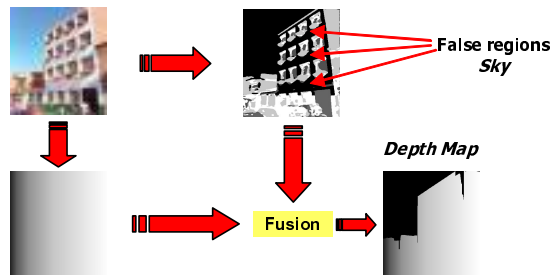


Figure 4- Examples of depth map generation by fusion.

### 3.4. Depth map generation by fusion

In this step *qualitative depth map* and *geometric depth map* are "fused" together to generate the final depth map  $M$ . Let  $M1(x,y)$  be the *geometric depth map* and  $M2(x,y)$  the *qualitative depth map* after the consistency verification analysis of the regions. The "fusion" between  $M1(x,y)$  and  $M2(x,y)$  depends on the image category:

1. If the image belongs to the *indoor* category then  $M(x,y)$  coincides with  $M1(x,y)$ :

$$M(x,y) = M1(x,y) \quad \forall (x,y) : 0 \leq x \leq W-1 \quad 0 \leq y \leq H-1.$$

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:

$$M(x,y) = M1(x,y) \quad \forall (x,y) \in Land \quad \text{and} \quad \forall (x,y) \in Other$$

$$M(x,y) = M2(x,y) \quad \forall (x,y) \notin Land \quad \text{and} \quad \forall (x,y) \notin Other$$

3. If the image is classified as *Outdoor* with *geometric characteristics* then the image  $M(x,y)$  is obtained as follows:

$$M(x,y) = M2(x,y) \quad \forall (x,y) \in Sky.$$

$$M(x,y) = M1(x,y) \quad \forall (x,y) \notin Sky.$$

Figure 4 shows an example of depth map fusion.

#### 4. Stereoscopic pair image generation

In [3] a method to reconstruct the binocular view of a source image has been proposed: the stereoscopic image pair is obtained by extracting the parallax values from the generated depth-map and applying them to the single view input image.

The parallax values can be thought as function of the inter-ocular distance, or baseline  $B$ . Considering that:

- an introduced big difference between consecutive depth layers, thus between foreground and background, affects the images with unresolved occlusions;
- the human eyes are commonly used to converge to a point focusing on it, while with a parallax close to  $B$  they have to stay parallel;

the max parallax should be less than  $B$  and, consequentially, the depth effect into the screen will be less visible. So fixing a max depth into the screen we can derive the max allowed parallax.

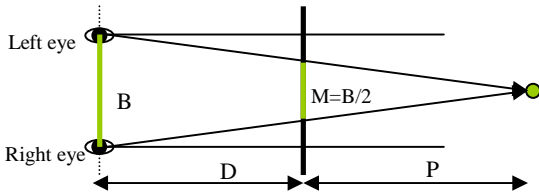


Figure 5- Parallax relations with respect to the screen.

Moreover, the viewer distance from the screen plays a fundamental role. The human vision system has 46 degrees as diagonal aperture; therefore the minimum distance allowing a comfortable vision of the screen can be evaluated. If the distance  $D$  (Figure 5), between the viewer and the screen, is equal to the max depth effect into the screen  $P$ , the achievable parallax is equal to  $B/2$ , thus more comfortable than  $B$ .

Considering the Figure 5, and exploiting the correlation between similar triangles, we have:

$$B : (D + P) = M : P \quad (4)$$

where  $M$  is the Max parallax,  $B$  is the inter-ocular distance,  $P$  is the depth into the screen and  $D$  is the user-to-screen distance.

Starting from the relationship (4), we propose:

$$M = F\{B, N, depth\_value\} \quad (5)$$

$$P = G\{D, B, M\} \quad (6)$$

where the *depth-value* is the depth-map pixel value and  $N$  is a reduction factor. Tuning different values of  $N$ , the max parallax value changes and the 3D entertainment can be improved.

In order to reconstruct the left and the right views, the equation (7) is used.

$$parallax = M * \left(1 - \frac{depth\_value}{255}\right) \quad (7)$$

For each pixels of the input image the value of the parallax is obtained from its *depth-value*. Considering the input image as a virtual central view, the left and the right views are then obtained, as shown in Figure 6, shifting the input image pixels by a value equal to  $parallax/2$  for each view. Further details can be found in [3].

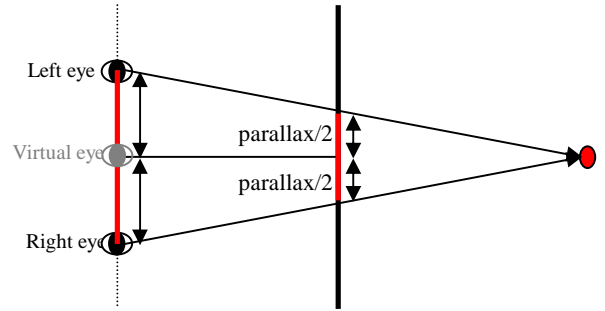


Figure 6- Stereoscopic pair generation.

#### 5. Experimental results

The overall methodology has been tested using a large dataset of images. A few images are part of a high quality database generally used to evaluate image processing algorithms.

The final perceived quality of the resulting anaglyph is good in almost all cases.

The experimental results confirm the robustness of the classification method (the percentage of the true positive results is about 93%). Some anaglyph has been generated starting from images acquired by CMOS digital sensor in Bayer pattern format [11]. In this case, as described in [1], a suitable heuristic allows to properly manage such data to obtain the final depth-map.

Figure 10 to Figure 13 show a series of depth-map images and related anaglyph obtained with our dataset.

The experimental results have confirmed how the computational complexity of the overall system is well suited even for real time application. For example,

starting from a VGA image the stereoscopic view is generated in about 1 second using a workstation equipped with a Xeon 2.8 GHz CPU and 1 GB RAM.

The proposed method presents many advantages as: automation, use of a single view of the scene, effectiveness especially in Outdoor and Panoramas images. Input images can also be acquired in Bayer Pattern format, reducing the overall complexity without sensibly affecting the final result.

## 6. Future works

Actually we are going to extend the image classification step in order to consider portrait images with landscape/indoor background and images with multiple VP. An image analysis based on structural components, and not only on image chromatic information, is necessary to avoid conditions such as sunset sky in a landscape image where the colour classifier could not identify the exact region category.

Moreover, the method can be improved in some aspects. For examples, regions detection could detect a greater number of regions (for example people and objects in foreground in which a gradient of depth should not be assigned). Also the possibility to use some advanced strategy as suggested in [12],[13] will be considered.

## 7. References

- [1] S. Battiato, S. Curti, M. La Cascia, E. Scordato, M. Tortora, "Depth Map Generation By Image Classification", SPIE IS&T/SPIE's 16th Annual Symposium on Electronic Imaging 2004.
- [2] V. Cantoni, L. Lombardi, M. Porta, N. Sicari, "Vanishing Point Detection: Representation Analysis

- and New Approaches", Dip. di Informatica e Sistemistica – Università di Pavia IEEE 2001.
- [3] S. Curti, D. Sirtori, F. Vella, "3D Effect Generation from Monocular View", In IEEE Proc. of - 3DPVT'02 International Symposium on 3D Data Processing Visualization and Transmission, pp. 550 –553, 2002.
- [4] J.R. Smith, C.S. Li, "Decoding Image Semantics Using Composite Region Templates", In Proc. of CVPR, Workshop on Content-Based Access of Image and Video Libraries, 1998.
- [5] J.R. Smith, C.S. Li, "Image Classification and Querying Using Composite Region Templates", Journal of Computer Vision and Image Understanding, 1999.
- [6] J.R. Smith, S.F. Chang, "Multi-stage Classification of Image from Futures and Related Text", In Proc. of Fourth DELOS workshop, Pisa, Italy, August, 1997.
- [7] D. Comaniciu, P. Meer, "Robust Analysis of Feature Spaces: Color Image Segmentation", In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 750-755, June 1997.
- [8] I.P. Howard, J.R. Brian, A.P. Howard, "Binocular Vision and Stereopsis", Oxford University Press, Chapter 7, 1995.
- [9] R. Klette, K. Schluns, A. Koschan, "Computer Vision", Springer, pp. 29 – 33, 1998.
- [10] E. Trucco, A. Verri, "Introductory Techniques for 3D Computer Vision", Prentice Hall, pp. 140 – 143, 1998.
- [11] S. Battiato, M. Mancuso, "An Introduction to the Digital Still Camera Technology", ST Journal of System Research, Special Issue on Image Processing for Digital Still Camera, Vol. 2, No.2, December 2001.
- [12] A. Criminisi, "Accurate Visual Metrology from Single and Multiple Uncalibrated Images", Springer-Verlag London Ltd, July 2001, ISBN 1852334681.
- [13] R. Cipolla and D.P. Robertson, "3D Models of Architectural Scenes from Uncalibrated Images and Vanishing Points", In Proc. IAPR 10th International Conference on Image Analysis and Processing, Venice, pages 824-829, (September) 1999.



(a)

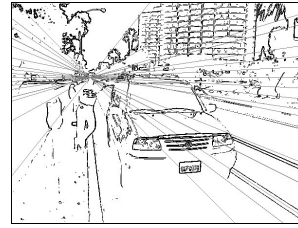


(b)

Figure 7- (a) Original Image; (b) segmentation image.

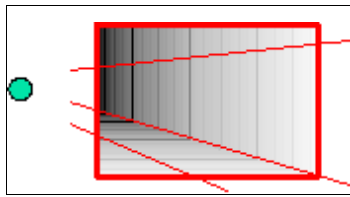


(a)

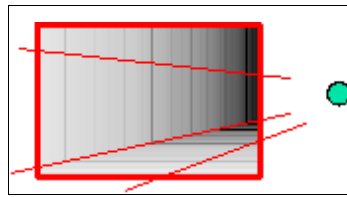


(b)

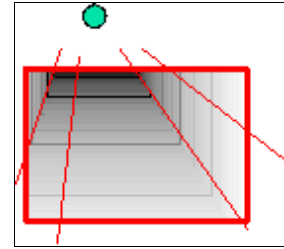
Figure 8 - (a) Input image classified as Outdoor with Geometric appearance; (b) Result of vanishing lines detection.



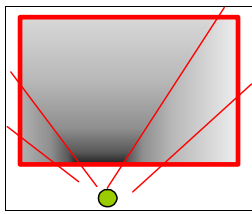
(a): Left case



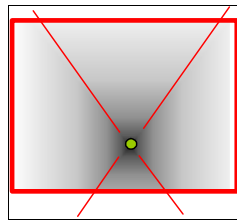
(b): Right case



(c): Up case



(d): Down Case



(e): Inside case

Figure 9- Examples of the heuristic rules to generate depth gradient planes: the green circle represents the vanishing point.



(a)



(b)



(c)

Figure 10 - Example of image categories. (a) Outdoor; (b) Outdoor with geometric appearances; (c) Indoor.



(a)



(b)



(c)

Figure 11- (a) Original indoor Image; (b) in indoor images the geometric depth map is also the final depth map (b); Anaglyph image (c).



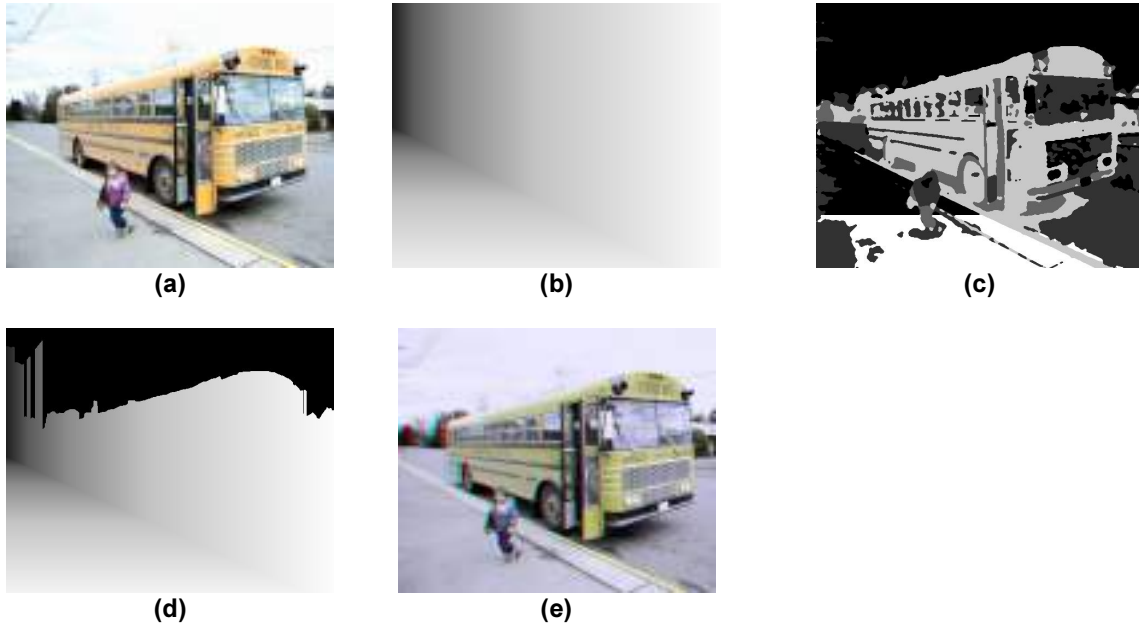


Figure 12 - (a) Original outdoor with geometric appearances Image; (b) Geometric depth map; (c) Qualitative depth map; (d) Final depth map; (e) Anaglyph image.

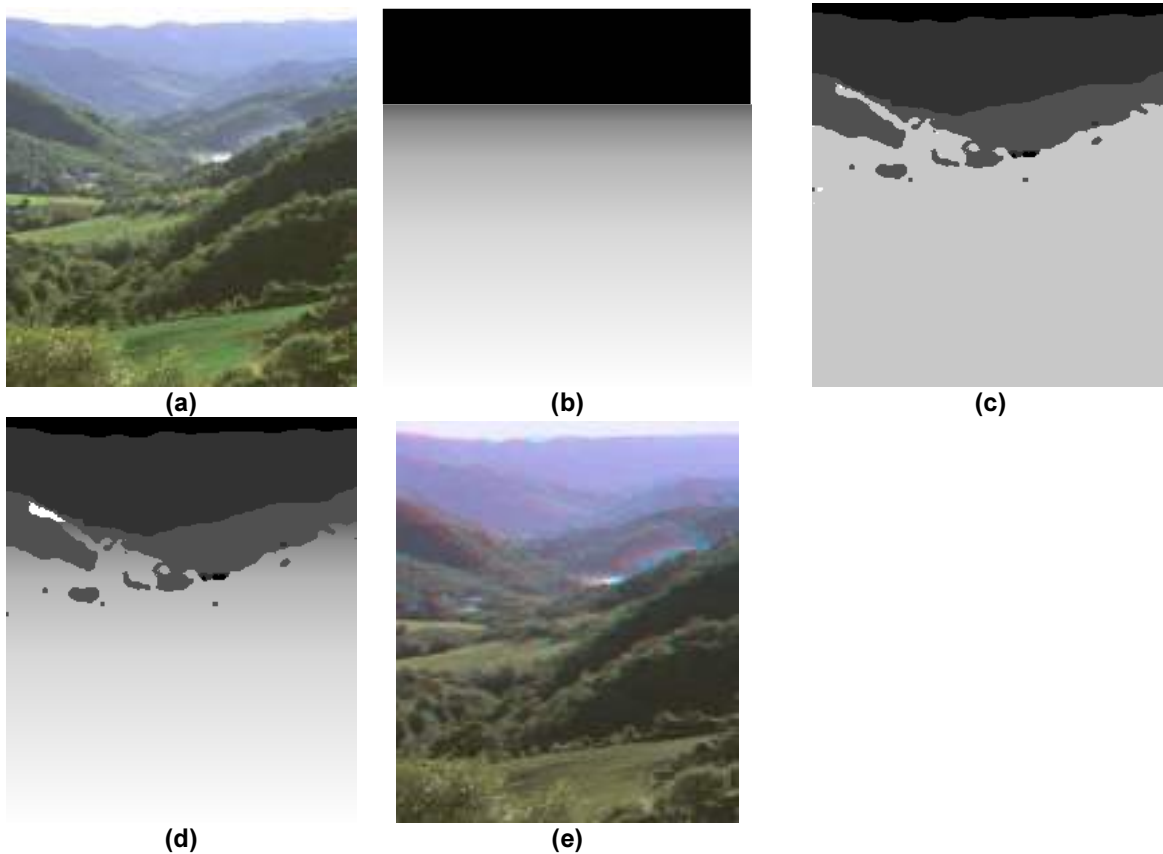


Figure 13: (a) Original panoramic Image; (b) Geometric depth map; (c) Qualitative depth map; (d) Final depth map; (e) Anaglyph image.