Depth-Map Generation by Image Classification

S. Battiato[a], S. Curti[b], M. La Cascia[c], M. Tortora[c], E. Scordato[c]
[a]Università di Catania, Dipartimento di Matematica ed Informatica - Catania, Italy
[b]STMicroelectronics- AST Catania Lab - Catania, Italy
[c]Università di Palermo, DINFO - Palermo, Italy

ABSTRACT

This paper presents a novel and fully automatic technique to estimate depth information from a single input image. The proposed method is based on a new image classification technique 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 proposed technique is fully unsupervised and is able to generate depth map from a single view of the scene, requiring low computational resources.

Keywords: Image classification, Depth-Map

1. INTRODUCTION

3D images have become more and more popular in everyday life (3D games, PC's video technologies, 3D CAD systems, etc.) [1]. Several techniques have been proposed to convert existing 2D images to 3D images. In many cases these techniques are semi-automatic: in [2] a suitable operator identifies objects for depth placement in the 2D image while in [3] a special effects artist guides the generation of depth maps using a Machine Learning Algorithm. Other methods, based on the motion of the objects relative to the camera, have been proposed to calculate depth maps by estimating and analyzing optical flow [4]. Another class of algorithms uses focus and defocus information [5], [6].
This paper proposes a method based on a novel image classification technique able to classify digital images (also in Bayer pattern format [7]) as indoor, outdoor with geometric elements or outdoor. In particular, the input image is processed by the following steps:
1) Bayer to approximated-RGB color conversion;
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.

The method is well suited for real-time application. Effectiveness of the proposed processing pipeline has been validated by an exhaustive set of experiments.
The paper is organized as follows. Section 2 describes the techniques used to extract some specific features from the input image. The image classifier is described in Section 3 while in Section 4 the depth map estimation is discussed. Experimental results are given in Section 5. Section 6 closes the paper tracking directions for future works.

2. IMAGE PRE-PROCESSING

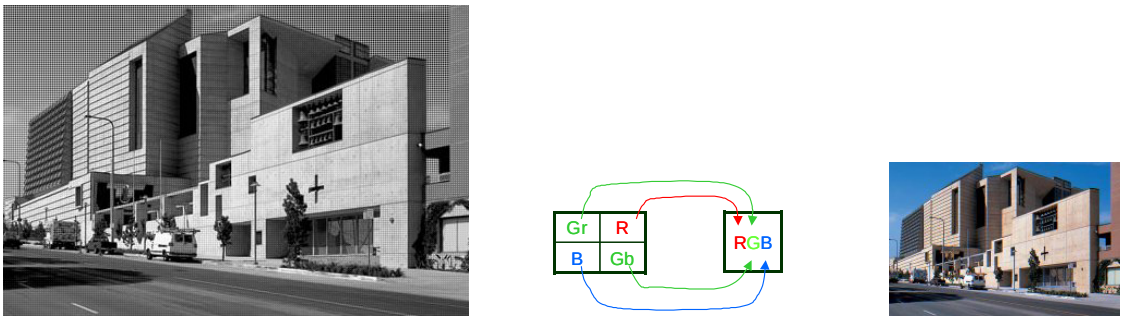
In this phase some features of the input image used by classification and by depth map estimation are extracted. The main processing consists of three steps: Macropixel Bayer to RGB color conversion, Color-based segmentation and Regions detection.

2.1. Macropixel Bayer to RGB color conversion

Acquiring images by digital CCD/CMOS sensor, in Bayer Pattern format [7], the final chromatic components of the image have to be reconstructed by some color reconstruction technique [8]. For our purposes, using low computational resources, an RGB images is generated converting 2x2 blocks of the input bayer data into a RGB pixel, in the following way (Figure 1). The green value is obtained as

$$G_i = \frac{G_{R_i} + G_{B_i}}{2} \qquad (1)$$

where Gr and Gb are the green values in the *i-th* 2x2 block. The red and the blues values are simply retained. This color conversion technique leads to a RGB image with reduced dimensions respect to the original image. The chromatic information are however enough to proceed with the successive steps.



(a) (b) (c)

Figure 1 - Bayer Pattern image (a); Macropixel interpolation (b); RGB image after color recovery (c).

2.2. Color-based segmentation

The color-based segmentation identifies chromatically homogeneous regions. We use the segmentation technique described in [9]: 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.

As the figure 2 shows, the output image have a reduced range of colors. Generally, the output of the under segmentation is an image with about ten different colors.

(a) (b)

Figure 2 - Original Image (a); Under segmentation image (b).

2.3. Regions detection

The identification of semantic regions in a generic image is a crucial step needed to obtain a robust image classifier. As suggested by Smith in ([10],[11],[12]), 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 (see Figure 3.a):

- B(*x*,*y*)>=G(*x*,*y*)+30 AND G(*x*,*y*)>=R(*x*,*y*)
- B(*x*,*y*)>=G(*x*,*y*) AND G(*x*,*y*)>=R(*x*,*y*)
- B(*x*,*y*)>=R(*x*,*y*) AND R(*x*,*y*)>=G(*x*,*y*)
- I(*x*,*y*)>=200 AND |B(*x*,*y*)-G(*x*,*y*)|<=30 AND |R(*x*,*y*)-G(*x*,*y*)|<=30

Similar conditions have been used in order to detect the *Farthest Mountain*, *Far Mountain*, *Near Mountain*, *Land* and *Other* regions. These color-based rules have been learned heuristically, using a large number of landscape images.
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. The output image can be thought as a *qualitative depth map* (Figure 3.b, Figure 4).

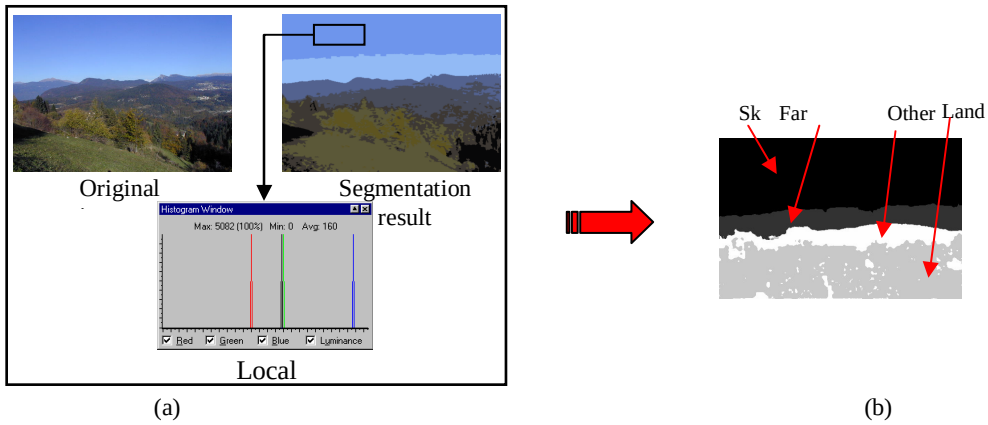


Figure 3 - Example of the heuristic rule to classify the *sky* region (a); result of the detected regions (b)

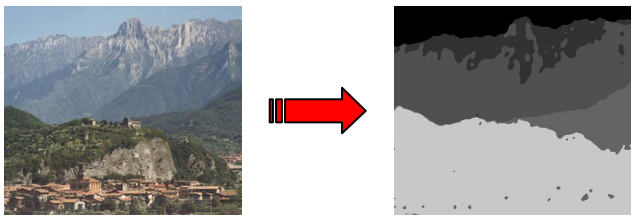


Figure 4 - Example of qualitative depth map generation.

3. 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 preliminary results have been obtained focusing the classifier to the following categories: *Outdoor/Landscape*, *Outdoor with geometric elements*, and *Indoor* (Figure 5).

(a) (b) (c)

Figure 5 - Example of image categories. (a) Outdoor, (b) Outdoor with geometric appearances, (c) Indoor.

The proposed technique is based on the comparison of a group of *N* sample columns of the regions detection's output with a set of typical "sequences" of a landscape. This approach as also described in ([10],[11],[12]) 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 1, 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".

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

Table 1 - Regions and corresponding labels.

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 (Figure 6).
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 isn't a typical landscape sequence or if the jumps number is bigger than *J_B* then the sequence is rejected. Such analysis can be summarized by the schematic description reported in Figure 7.
3. Final classification. The image is classified as *Outdoor* if the value of N_1 is bigger than $R_1 \cdot N$, 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 $R_2 \cdot N$, where R_2 is another threshold in]0,1[the image is classified as *Outdoor with geometric appearance* else it is classified as *Indoor*

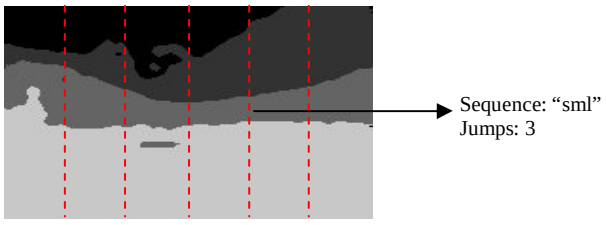


Figure 6 - Example of sequence and jumps for a sample column.

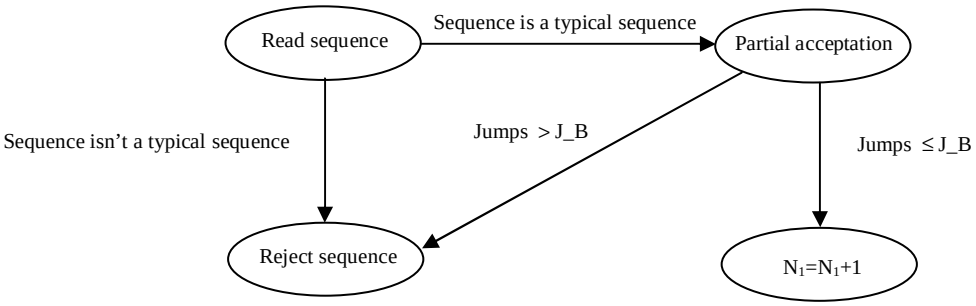
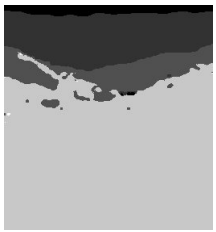


Figure 7- Schematic description of the sequences accept/reject procedure.



	1ª colunm	2ª colunm	3ª colunm	4ª column	5ª colunm
Sequence	**ml**	**ml**	**ml**	**ml**	**ml**
Jumps	5	5	3	3	3

N_1	N^*R_1	N_2	N^*R_2
5	4	0	1

Outdoor

(a) (b) (c)

Figure 8 - Example of image classification: (a) Original image, (b) Regions detection, (c) Final results with N=5, J_B= 10, R_1=0.8, R_2=0.2.

4. DEPTH-MAP GENERATION

An approximated depth map can be generated taking into account the information collected by the previous classification stage. In the following, a series of intermediate steps, used to recover the final depth are properly presented and discussed.

4.1. Vanishing lines detection

Initially some image features, like Vanishing Point (VP) and vanishing lines [13], have to be detected. The proposed method analyzes the result of the previous image classification.

If the input image is classified as *Outdoor without geometric elements*, the lowest point in the boundary between the region A = *Land* ∪ *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 image's width. Moreover the method generates a set of standard vanishing lines (Figure 9).

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_{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 straight lines, using I_{Sx} and I_{Sy}, passing through each edge point of I_E:

$$m(x,y) = \frac{I_{S_y}(x,y)}{I_{S_x}(x,y)} \quad (2)$$

$$q(x,y) = y - m(x,y) \cdot x \quad (3)$$

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:

$$ACC[m,q] = Acc[m,q] + 1 \quad (4)$$

Higher values correspond to the main straight lines of the original image.

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 10).

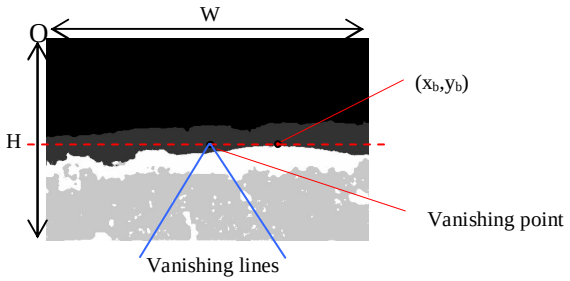
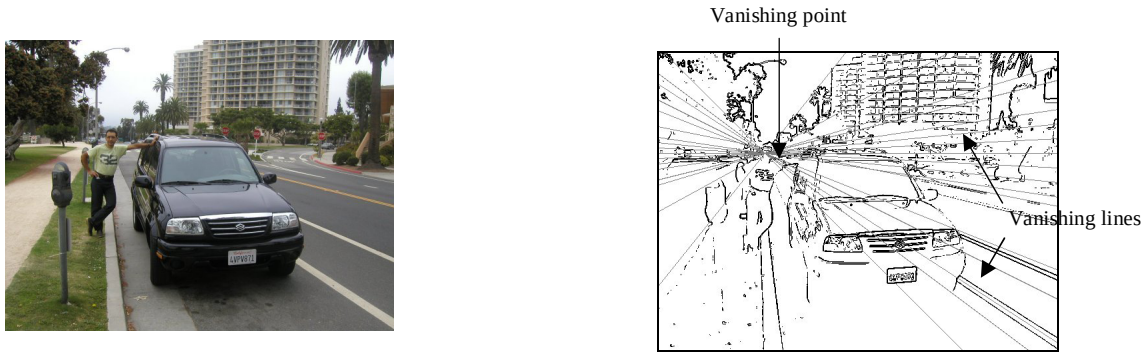


Figure 9 - The boundary point and the resulting vanishing point for *Outdoor* images.



(a) (b)

Figure 10 - (a) Input image classified as Outdoor with Geometric appearance, (b) Result of vanishing lines detection.

4.2.Gradient planes 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:

Xvp<=0 AND (**H**-1/**W**-1) * **Xvp**<**Yvp**<-(**H**-1/**W**-1) * **Xvp**+**H**-1	**Left Case**	(5)
Xvp>=**W**-1 AND -(**H**-1/**W**-1) * **Xvp**+**H**-1<**Yvp**<(**H**-1/**W**-1) * **Xvp**	**Right Case**	(6)
Yvp<=0 AND (**W**-1/**H**-1) * **Yvp** <= **Xvp** <=(**W**-1/**H**-1) * (**H**-1-**Yvp**)	**Up Case**	(7)
Yvp>=**H**-1 AND (**W**-1/**H**-1) * (**H**-1-**Yvp**) <=**Xvp**<=(**W**-1/**H**-1) * **Yvp**	**Down Case**	(8)
0<**Xvp**<**W**-1 AND 0<**Yvp**<**H**-1	**Inside Case**	(9)

where (**Xvp, Yvp)** are the VP coordinates **H**, **W** are the image dimensions.
For each case, a set of heuristic rules, based on vanishing lines slopes, allows generating planes (gradient planes) used to gradually set the depth variation. Figure 11 shows how horizontal and/or vertical gradient planes are located.

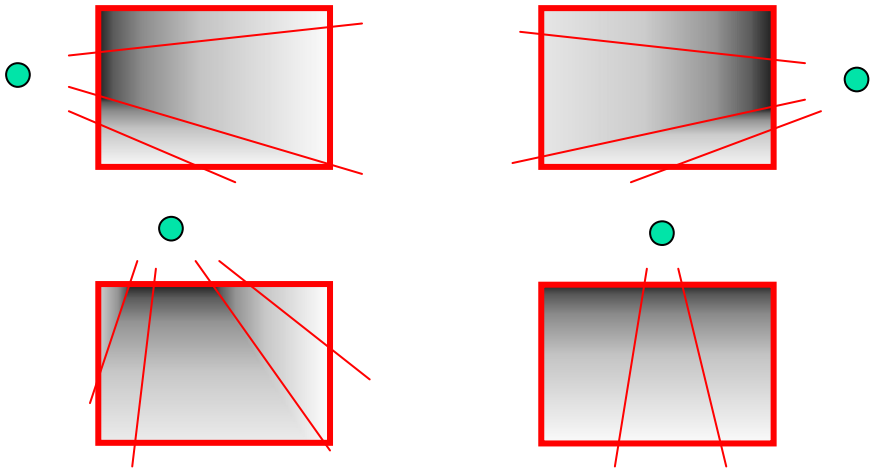


Figure 11- Examples of the heuristic rules to generate depth gradient planes: the green circle represents the vanishing point.

4.3.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).

Figure 11 shows as in horizontal planes the grey level (depth level) is constant along the rows while in vertical planes the grey level (depth level) is constant along the columns. The depth levels value is approximated by a *piece wise linear* and it depends from slopes *m1* and *m2* of *vanishing lines* generating *depth gradient plane* (Figure 12).
Figure 13 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.

4.4.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 color information. It is necessary, therefore, to analyze the positions, inside the image, of each region with respect to the others checking their dimensions. 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 (see figure 14).

4.5.Depth map generation by fusion

In this step *qualitative depth map* and *geometric depth map* are “fused” to generate the final depth map **M**. Let be $\mathbf{M_1(x,y)}$ the *geometric depth map* and $\mathbf{M_2(x,y)}$ the *qualitative depth map* after the consistency verification analysis of the regions. The “fusion” between $\mathbf{M_1(x,y)}$ and $\mathbf{M_2(x,y)}$ depends on the image category:

1. If the image belongs to the *indoor* category then **M(x,y)** coincides with $\mathbf{M_1(x,y)}$:

$$\mathbf{M(x,y)} = \mathbf{M_1(x,y)}\ \forall \mathbf{(x,y)}: \quad 0<=\mathbf{x}<=\mathbf{W}-1 \quad 0<=\mathbf{y}<=\mathbf{H}-1. \tag{10}$$

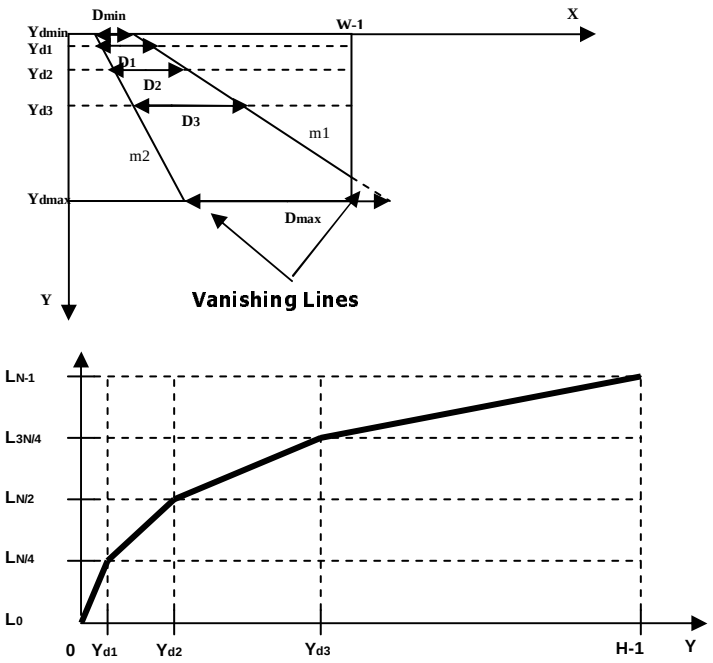


Figure 12 - Example of depth gradient assignment for an horizontal plane generated by two vanishing lines.

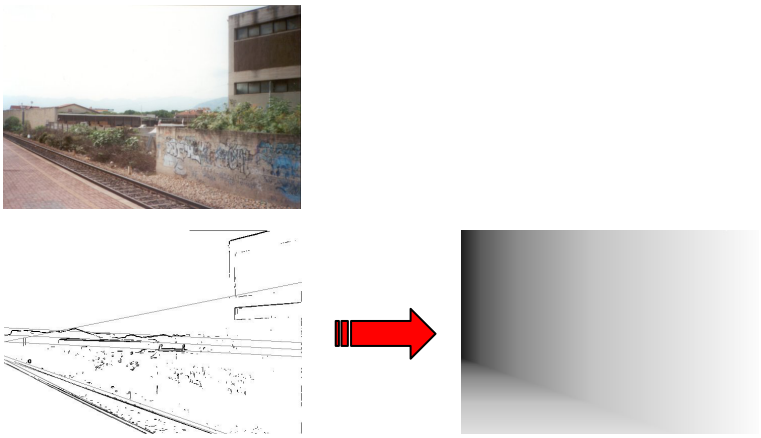


Figure 13 - Example of ***geometric depth map*** generation.

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:

$$\mathbf{M(x,y)} = \mathbf{M_1(x,y)} \forall \mathbf{(x,y)} \in \textit{Land} \text{ and } \forall \mathbf{(x,y)} \in \textit{Other} \quad (11)$$

$$\mathbf{M(x,y)} = \mathbf{M_2(x,y)} \ \forall \mathbf{(x,y)} \notin \textit{Land} \text{ and } \forall \mathbf{(x,y)} \notin \textit{Other} \quad (12)$$

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

$$\mathbf{M(x,y)} = \mathbf{M_2(x,y)} \ \forall \mathbf{(x,y)} \in \textit{Sky}. \quad (13)$$

$$\mathbf{M(x,y)} = \mathbf{M_1(x,y)} \ \forall \mathbf{(x,y)} \notin \textit{Sky}. \quad (14)$$

Figure 14 shows an example of depth map fusion.

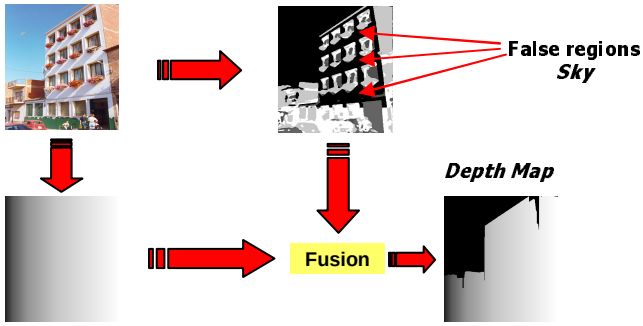


Figure 14 - Examples of depth map generation by fusion.

5. EXPERIMENTAL RESULTS

The overall methodology has been tested using a large dataset of images. Figure 15 shows a series of depth-map images obtained starting from typical images. Some images have been acquired by CMOS digital sensor in bayer pattern format [7]. The computational complexity is well suited for real time application. The experimental results confirm the robustness of the method in the classification stage even if the extension to further image categories is definitively needed.
Heuristics rules used for regions detection and/or for depth assignment allow to manage properly almost all involved category.

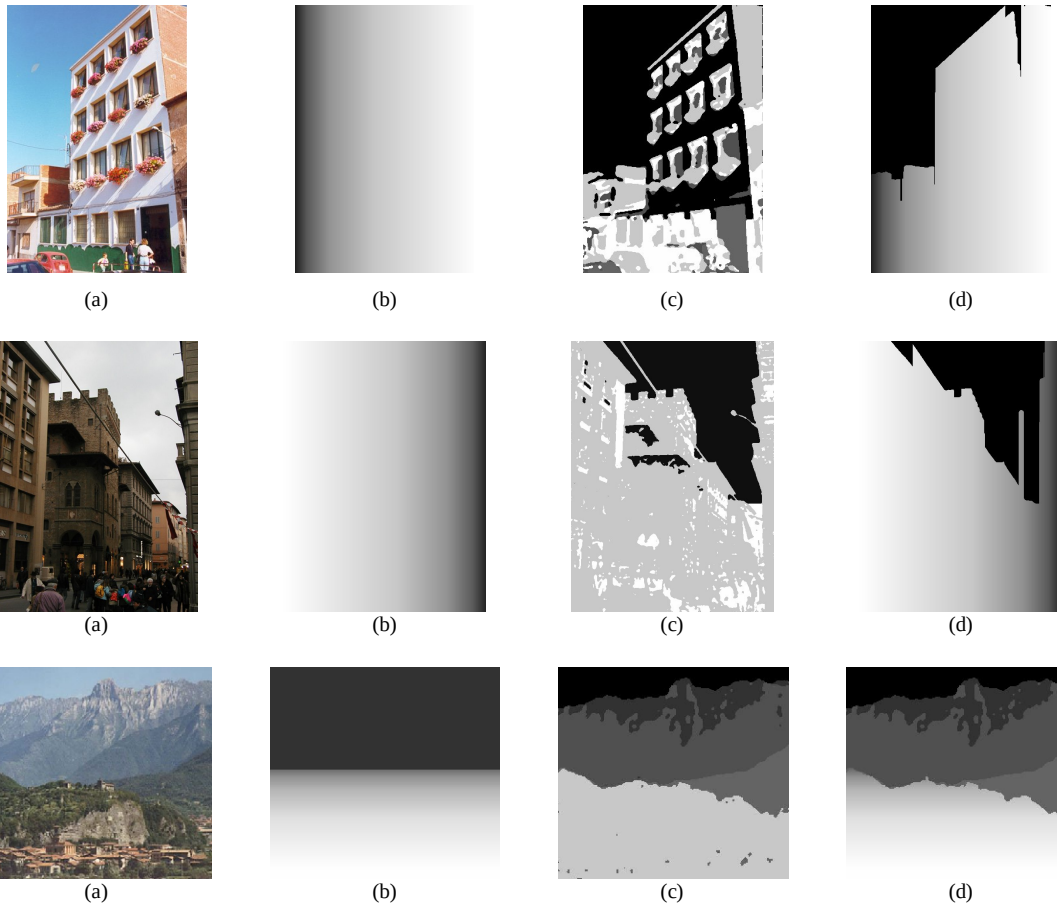
6. CONCLUSIONS AND FUTURE WORKS

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. 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) or vanishing lines detection could detect a possible second vanishing point and its relative vanishing lines. This method could be used in order to obtained the stereo pair image from a single view [14].

7. REFERENCES

1. P. Harman, "Home Based 3D Entertainment – An Overview", In Proc. Of IEEE *Intl. Conference on Image Processing*, p. 1-4, Vancouver 2000.
2. H. Murata, Y. Mori, S. Yamashita, A. Maenaka, S. Okada, K. Oyamada, S. Kishimoto, "A Real Time 2D to 3D Image Conversion Technique Using Computed Image Depth", *SID SYM*, V**ol. 29**, pp. 919-922, 1998.
3. P. Harman, J. Flack, S. Fox, M. Dowley, "Rapid 2D to 3D Conversion", In Proc. SPIE Vol. 4660, *Stereoscopic Displays and Virtual Reality Systems IX*, pp. 78-86, 2002.
4. Y. Matsumoto, H. Terasaki, K. Sugimoto, T. Arakawa, "Conversion System of Monocular Image Sequence to Stereo Using Motion Parallax", In Proc. of SPIE, Vol. 3012, *Stereoscopic Displays and Virtual Reality Systems IV*, pp. 108-112, May 1997.
5. P. Grossman, "Depth from focus", *Pattern Recognition Letters*, **Vol. 5**, No. 1, pp 63-69, Jan 1987.
6. M. Subbarao, "Parallel Depth Recovery by Changing Camera Parameters", In Proc. Of IEEE *Intl. Conf. On Computer Vision*, pp. 149-155, Florida, USA, 1988.
7. 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.

8. D. Brainard, “Bayesian Method for Reconstructing Color Images from Trichromatic Samples”, In Proc. of *IS&T 47th Annual Conference*, 1994.
9. 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.
10. J.R. Smith, Chung-Sheng Li, “Decoding Image Semantics Using Composite Region Templates”, In Proceedings of CVPR, *Workshop on Content-Based Access of Image and Video Libraries*, 1998.
11. J.R. Smith, Chung-Sheng Li, “Image Classification and Querying Using Composite Region Templates”, Journal of *Computer Vision and Image Understanding*, 1999.
12. J.R. Smith, Shih-Fu Chang, “Multi-stage Classification of Image from Futures and Related Text”, In Proc. of Fourth *DELOS workshop*, Pisa, Italy, August, 1997.
13. 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.
14. 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.

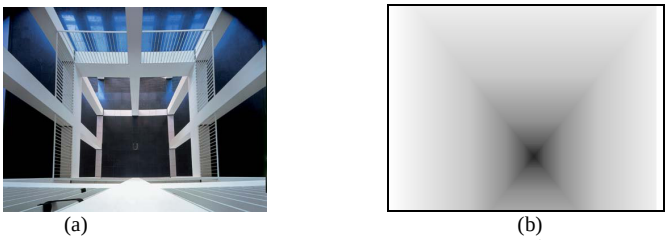


(a) (b) (c) (d)

(a) (b) (c) (d)

(a) (b) (c) (d)

Figure 15 - (a)Original Image; (b) *Geometric depth map*; (c) *Qualitative depth map*; (d) Final depth map.



(a) (b)

Figure 16 - (a) Original Indoor Image; (b) Geometric and final depth map.