실안개를 이용한 단일 영상으로부터의 깊이정보 획득 및 뷰 생성 알고리듬

*소용석, **현대영, ***이상욱

서울대학교 전기공학부 신호처리연구실

*ysoh611@ipl.snu.ac.kr **huanream@ipl.snu.ac.kr

***sanguk@ipl.snu.ac.kr

Depth estimation and View Synthesis using Haze Information

*Yongseok, Soh **Dae-Young, Hyun ***Sang-uk, Lee

Signal Processing Lab., School of Electrical Engineering,

Seoul National University

Abstract

Previous approaches to the 2D to 3D conversion problem require heavy computation or considerable amount of user input. In this paper, we propose a rather simple method in estimating the depth map from a single image using a monocular depth cue: haze. Using the haze imaging model, we obtain the distance information and estimate a reliable depth map from a single scenery image. Using the depth map, we also suggest an algorithm that converts the single image to 3D stereoscopic images. We determine a disparity value for each pixel from the original 'left' image and generate a corresponding 'right' image. Results show that the algorithm gives well refined depth maps despite the simplicity of the approach.

1. Introduction

Nowadays, due to the heavy demands of 3D stereo image contents and the development and manufacture of 3D display panels, the 2D to 3D conversion problem is a popular issue in the field of computer vision.

This ill-posed problem has been approached in many different ways such as using a supervised learning approach or using a sequence of images as input. In most cases the 2D to 3D conversion problem requires either heavy computation or considerable amount of user input.

In this paper, we propose an algorithm which generates a depth map using haze information, a monocular cue. This implies that we use a single 2D image as input. We target scenery image and show that by using the haze formation model we can obtain a reliable depth map. Using the depth map we obtained we synthesize a new image, which would correspond as the right view image and in result, give stereoscopic 3D images.

2. Related Works

In this section we will review a couple of approaches in estimating the depth map using a single 2D image. Also, we will give brief explanation regarding an algorithm that suggests a prior which is mentioned throughout this paper.

There are numerous approaches to solve the 2D to 3D conversion problem using all kinds of cues, and lots of on-going work exists as well.

The cues used by the human eye enable us to perceive depth in numerous ways. The cues used to recognize depth can be classified as monocular cues and binocular cues. Monocular cues are, for example, the sizes of objects, haze, perspectives, texture gradient etc. The binocular cue we use is called binocular disparity. It means we perceive depth by taking in two different images for each the left and right eye.

Depth from motion figures the relationship between the motion of the object and its distance from the camera by using optic flow [1]. This, of course, assumes that there is a sequence of images and the objects in the images move proportionally to their corresponding depth and assign depth to each pixel according to its movement.

An improved, high level algorithm introduces the idea of machine learning [2]. They introduce the 'feature vector' which contains information of all kinds of cues such as: texture variation, texture gradient, occlusion, haze etc for a patch of the image. By undergoing a training process for several images they use Markov Random Field to model the relation between the depth of a patch and the depths of its neighboring patches. This implies that a large amount of user input is necessary. And due to the training and classification stages, the conversion is a task which requires a good amount of computation.

Throughout this paper, we use a prior suggested by K. He *et al.*, called the dark channel prior. It is a simple and effective prior based on a key observ-vation - most local patches in haze-free images contain almost zero intensity in at least one color channel. By using this prior, we can manipulate the model used to describe the formation of haze images and obtain high quality haze-free images.

3. Proposed Algorithm

There are 2 steps in our proposed algorithm. The first step is the estimation of the depth map using dark channel prior, and the second step is the synthesis of the right view image using the depth map obtained from the first step.

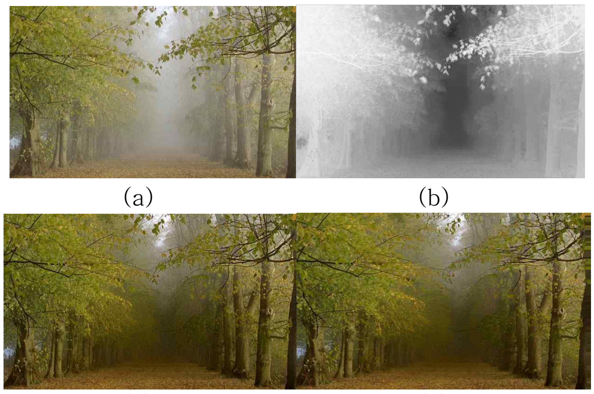
As mention previously in this paper, we will take a look at the equation which describes the formation of haze images.

$$\mathbf{I}(x) = \mathbf{J}(x)t(x) + \mathbf{A}(1 - t(x)) , \qquad (1)$$

where $\mathbf{I}$ is the observed intensity, $\mathbf{J}$ is the scene radiance and $\mathbf{A}$ is the global atmospheric light, and t is the transmission term. Here, we focus on the transmission term. If we assume that the atmosphere is homogeneous, the transmission term can be expressed as:

$$t(x) = e^{-\beta d(x)} , \qquad (2)$$

where β is the scattering coefficient of the atmosphere. By examination, we can see that the transmission is exponentially related with the depth of the scene. According to [3], equation (1) can be organized in respect to t as:



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

Figure1. Results of the proposed algorithm. (a) Input image. (b) Estimated depth map. (c) Left view image. (d) Right view image.

$$t(x) = 1 - \min_c\left(\min_{y \in \Omega(x)} \left(\frac{I^c(y)}{A^C} \right)\right) , \qquad (3)$$

where c represents the color channels of the image. Because using the dark channel requires operations in windows, the results have block artifacts. Therefore we apply a soft matting algorithm suggest in [4] to refine the results.

As mentioned above we can use equation (3) to obtain the depth value by calculating:

$$d(x) = -\frac{1}{\beta} \ln t(x) . \qquad (4)$$

The second step is the synthesis of the right view. The idea is simple and we approached this problem in a quite intuitive manner. We assign each pixel from the original image (left image) to a disparity value corresponding to the depth assigned by (4). In this process we make sure that the pixels with bigger disparity values are assigned first due to the fact that occlusion will occur and the object further away will be occluded by the object standing closer.

In the synthesized image, there are areas which are empty because of the discrete values in the depth map, especially on the border of objects. We call these areas 'holes' and fill them by repeating the pixels adjacent to the holes. In cases where the size of the holes is big, distortion problems may arise. Therefore to avoid such cases, we limit the disparity value to 10 percent of the horizontal scale of the image.

4. Results

In this section we provide and show the results generated by the proposed algorithm and evaluate the results. In Figure1, An image containing haze can be expressed as the sum of scene radiance and atmospheric light as shown in (a). By using the prior suggested by K. He *et al.*, and refining the results suggested in [4] we can obtain a well refined depth map such as (b). (c) is the left image and (d) is the synthesized right image. We can see that the proposed algorithm gives good results despite the simple approach. However, there are some cases where this algorithm may not work as well. The dark channel prior may not hold and give false implications of the atmospheric light. This may result in inaccurate depth maps.

5. Conclusion

In conclusion, this paper suggests an algorithm which converts a single 2D scenery image into 2 stereoscopic images using haze information. We have shown that this method, compared to other previous approaches, requires no user input and less computation but yet provides fair results. Although the algorithm is limited in images where haze is in existence, the approach taken in this paper is expected to be widely applied in the conversion problem. We intend to expand the algorithm by using other depth cues in the future.

6. Reference

[1]Y. Matsumoto, H. Terasaki, K. Sugimoto, and T. Arakawa, "Conversion System of Monocular Image Sequence to Stereo Using Motion Parallax," in *Proc. SPIE* Stereoscopic Displays and Virtual Reality Systems vol. 3012, IV, pp. 108-112, 1997.

[2]A. Saxena, S. Chung, and A.Y. Ng, "Learning depth from single monocular images," in *Proc. NIPS*, 2005.

[3]K. He, J. Sun, X. Tang, "Single image haze removal using dark channel prior," in *Proc. CVPR,* pp. 1956-1963, 2009.

[4] A. Levin, D. Lischinski, and Y. Weiss "A closed form solution to natural image matting," in *Proc. CVPR*, pp. 1:61-68, 2006