CONVERSION OF 2D IMAGES TO 3D

A PROJECT REPORT

Submitted by

GAYATHRI.J

Register No: 13MCO09

in partial fulfillment for the requirement of award of the degree

of

MASTER OF ENGINEERING

in

COMMUNICATION SYSTEMS

Department of Electronics and Communication Engineering

KUMARAGURU COLLEGEOF TECHNOLOGY
(An autonomous institution affiliated to Anna University, Chennai)

COIMBATORE-641049

ANNA UNIVERSITY: CHENNAI 600 025

APRIL-2015

BONAFIDE CERTIFICATE

Certified that this project report titled **“CONVERSION OF 2D IMAGES TO 3D”**is the bonafide work of **GAYATHRI.J[Reg. No. 13MCO09]** who carried out the research under my supervision. Certified further that, to the best of my knowledge the work reported herein does not form part of any other project or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE
PROF. S. GOVINDARAJU
PROJECT SUPERVISOR
Department of ECE
Kumaraguru College of Technology
Coimbatore-641 049

SIGNATURE
Dr. RAJESWARI MARIAPPAN
HEAD OF THE DEPARTMENT
Department of ECE
Kumaraguru College of Technology
Coimbatore-641 049

The candidate with university **Register No.13MCO09** is examined by us in the project viva-voce examination held on...........................

ACKNOWLEDGEMENT

First, I would like to express my praise and gratitude to the Lord, who has showered his grace and blessings enabling me to complete this project in an excellent manner.

I express my sincere thanks to the management of Kumaraguru College of Technology and Joint Correspondent **Shri. Shankar Vanavarayar** for the kind support and for providing necessary facilities to carry out the work.

I would like to express my sincere thanks to our beloved Principal **Dr.R.S.Kumar Ph.D.,** Kumaraguru College of Technology, who encouraged me with his valuable thoughts.

I would like to thank **Dr.Rajeswari Mariappan Ph.D.,** Head of the Department, Electronics and Communication Engineering, for her kind support and for providing necessary facilities to carry out the project work.

In particular, I wish to thank with everlasting gratitude to the project coordinator **Ms.R.Hemalatha M.E.,(Ph.D.,)** Associate Professor, Department of Electronics and Communication Engineering ,for her expert counselling and guidance to make this project to a great deal of success.

I am greatly privileged to express my heartfelt thanks to my project guide **Prof.S.Govindaraju, M.E.,** Associate Professor, Department of Electronics and Communication Engineering, throughout the course of this project work and I wish to convey my deep sense of gratitude to all teaching and non-teaching staff of ECE Department for their help and cooperation.

Finally, I thank my parents and my family members for giving me the moral support and abundant blessings in all of my activities and my dear friends who helped me to endure my difficult times with their unfailing support and warm wishes.

ABSTRACT

The three dimensional(3D) image displays have become a trend in the visual processing field. 3D displays provide better visual experience than conventional 2D displays. The conversion of existing 2D images to 3D images becomes an important component of 3D content production. However, the depth information required for 3D displays is not available in the conventional 2D content. The manual creation of 3D images is time consuming and expensive. Several approaches require specific devices to generate 3D images thus making it infeasible for conversion of 2D contents. Therefore, an automatic 2D to 3D conversion is necessary.

The main aim of this project is to present an efficient algorithm for 2D to 3D conversion based on the depth information. The image is segmented into groups based on the pixel intensity using k-means segmentation algorithm. The grouping is based on the pixels having similar colors and spatial locality. Then the depth values are assigned as per the hypothesis depth value. Cross bilateral filter is then used to enhance the visual comfort and also to eliminate the blocky artifacts that are present in the depth map. Then the filtered image is processed using a Depth Image Based Rendering(DIBR) method to generate the 3D image. The key step in the process is the generation of depth map since the visual quality of the reconstructed image depends on the dense depth information.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVATIONS

2D	-	Two- Dimensional
3D	-	Three- Dimensional
DIBR	-	Depth Image Based Rendering
CID	-	Computed Image Depth
MST	-	Minimum Spanning Tree
MTF	-	Modulation Transfer Function

CHAPTER 1

INTRODUCTION

3D reconstruction has turned out to be an essential need in areas of medical imaging and artistic applications, product design, reverse engineering and rapid prototyping and many others. They also find many applications in broadcasting, gaming, photography, camcorders and education. 2D to 3D conversion is the process of transforming 2D film to 3D form, which in almost all cases is stereo, so it is the process of creating imagery for each eye from one 2D image. A 2D image has only height and width. On the other hand 3D image adds the perception of depth. 2D-to-3D conversion adds the binocular disparity depth cue to digital images perceived by the brain, thus, if done properly, greatly improving the immersive effect while viewing stereo image in comparison to 2D image. However, in order to be successful, the conversion should be done with sufficient accuracy and correctness: the quality of the original 2D images should not deteriorate, and the introduced disparity cue should not contradict to other cues used by the brain for depth perception. If done properly and thoroughly, the conversion produces stereo image of similar quality to "native" stereo which is shot in stereo and accurately adjusted and aligned in post-production. Several techniques are available for the reconstruction of 3D images. The conversion process of existing 2D images to 3D is fulfilling the growth of high quality stereoscopic images. The dominant technique for such content conversion is to develop a depth map for each frame of 2D image.

1.1 STEREOSCOPY:

Stereoscopy is a technique for creating or enhancing the illusion of depth in an image by means of stereopsis for binocular vision. Any stereoscopic image is called stereogram. Originally, stereogram referred to a pair of stereo images which could be viewed using a stereoscope. Most stereoscopic methods present two offset images separately to the left and right eye of the viewer. These two-dimensional images are then combined in the brain to give the perception of 3D depth. This technique is distinguished from 3D displays that display an image in three full dimensions, allowing the observer to increase information about the 3-dimensional objects being displayed by head and eye movements. Stereoscopy creates the illusion of three-dimensional depth from given two-dimensional images. Human vision, including the perception of depth, is a complex process which only begins with the acquisition of visual information taken in through the eyes; much processing ensues within the brain, as it strives to make intelligent and meaningful sense of the raw information provided. One of the very important visual functions that occur within the brain as it interprets what the eyes see is that of assessing the relative distances of various objects from the viewer, and the depth dimension of those same perceived objects. The brain makes use of a number of cues to determine relative distances and depth in a perceived scene

Stereoscopy is the production of the illusion of depth in a photograph, movie, or other two-dimensional image by presenting a slightly different image to each eye, and thereby adding the first of these cues (stereopsis) as well. Both of the 2D offset images are then combined in the brain to give the perception of 3D depth. It is important to note that since all points in the image focus at the same plane regardless of their depth in the original scene, the second cue, focus, is still not duplicated and therefore the illusion of depth is incomplete. There are also primarily two effects of stereoscopy that are unnatural for the human vision: first, the mismatch between convergence and accommodation, caused by the difference between an object's perceived position in front of or behind the display or screen and the real origin of that light and second, possible crosstalk between the eyes, caused by imperfect image separation by some methods.

Although the term "3D" is ubiquitously used, it is also important to note that the presentation of dual 2D images is distinctly different from displaying an image in three full dimensions. The most notable difference is that, in the case of "3D" displays, the observer's head and eye movement will not increase information about the 3-dimensional objects being displayed. Holographic displays or volumetric display are examples of displays that do not have this limitation. Similar to the technology of sound reproduction, in which it is not possible to recreate a full 3-dimensional sound field merely with two stereophonic speakers, it is likewise an overstatement of capability to refer to dual 2D images as being "3D". The accurate term "stereoscopic" is more cumbersome than the common misnomer "3D", which has been entrenched after many decades of unquestioned misuse. Although most stereoscopic displays do not qualify as real 3D display, all real 3D displays are also stereoscopic displays because they meet the lower criteria as well. Most 3D displays use this stereoscopic method to convey images.

Stereoscopy is used in photogrammetry and also for entertainment through the production of stereograms. It is useful in viewing images rendered from large multi-dimensional data sets such as are produced by experimental data. An early patent for 3D imaging in cinema and television was granted to physicist Theodor V. Ionescu in 1936. Modern industrial three-dimensional photography may use 3D scanners to detect and record three-dimensional information. The three-dimensional depth information can be reconstructed from two images using a computer by corresponding the pixels in the left and right images. Solving the Correspondence problem in the field of Computer Vision aims to create meaningful depth information from two images.

Traditional stereoscopic photography consists of creating a 3D illusion starting from a pair of 2D images, a stereogram. The easiest way to enhance depth perception in the brain is to provide the eyes of the viewer with two different images, representing two perspectives of the same object, with a minor deviation equal or nearly equal to the perspectives that both eyes naturally receive in binocular vision. To avoid eyestrain and distortion, each of the two 2D images should be presented to the viewer so that any object at infinite distance is perceived by the eye as being straight ahead, the viewer's eyes being neither crossed nor diverging. When the picture contains no object at infinite distance, such as a horizon or a cloud, the pictures should be spaced correspondingly closer together. The principal advantages of side-by-side viewers is the lack of diminution of brightness, allowing the presentation of images at very high resolution and in full spectrum color, simplicity in creation, and little or no additional image processing is required. Under some circumstances, such as when a pair of images are presented for freeviewing, no device or additional optical equipment is needed.

1.2 3D VIEWERS:

There are two categories of 3D viewer technology, active and passive. Active viewers have electronics which interact with a display. Passive viewers filter constant streams of binocular input to the appropriate eye.

Shutter systems:

A shutter system works by openly presenting the image intended for the left eye while blocking the right eye's view, then presenting the right-eye image while blocking the left eye, and repeating this so rapidly that the interruptions do not interfere with the perceived fusion of the two images into a single 3D image. It generally uses liquid crystal shutter glasses. Each eye's glass contains a liquid crystal layer which has the property of becoming dark when voltage is applied, being otherwise transparent. The glasses are controlled by a timing signal that allows the glasses to alternately darken over one eye, and then the other, in synchronization with the refresh rate of the screen.

Polarization systems:

To present stereoscopic pictures, two images are projected superimposed onto the same screen through polarizing filters or presented on a display with polarized filters. For projection, a silver screen is used so that polarization is preserved. On most passive displays every other row of pixels are polarized for one eye or the other. This method is also known as being interlaced. The viewer wears low-cost eyeglasses which also contain a pair of opposite polarizing filters. As each filter only passes light which is similarly polarized and blocks the opposite polarized light, each eye only sees one of the images, and the effect is achieved.

Color anaglyph systems:

Anaglyph 3D is the name given to the stereoscopic 3D effect achieved by means of encoding each eye's image using filters of different colors, typically red and cyan. Red-cyan filters can be used because our vision processing systems use red and cyan comparisons, as well as blue and yellow, to determine the color and contours of objects. Anaglyph 3D images contain two differently filtered colored images, one for each eye. When viewed through the "color-coded" "anaglyph glasses", each of the two images reaches one eye, revealing an integrated stereoscopic image. The visual cortex of the brain fuses this into perception of a three dimensional scene or composition.

Interference filter systems:

This technique uses specific wavelengths of red, green, and blue for the right eye, and different wavelengths of red, green, and blue for the left eye. Eyeglasses which filter out the very specific wavelengths allow the wearer to see a full color 3D image. It is also known as spectral comb filtering or wavelength multiplex visualization or super-anaglyph. Dolby 3D uses this principle. The Omega 3D/Panavision 3D system has also used an improved version of this technology. In addition to the passive stereoscopic 3D system, Omega Optical has produced enhanced anaglyph 3D glasses. The Omega’s red/cyan anaglyph glasses use complex metal oxide thin film coatings and high quality annealed glass optics.

Pulfrich method:

The Pulfrich effect is based on the phenomenon of the human eye processing images more slowly when there is less light, as when looking through a dark lens. Because the Pulfrich effect depends on motion in a particular direction to instigate the illusion of depth, it is not useful as a general stereoscopic technique. For example, it cannot be used to show a stationary object apparently extending into or out of the screen; similarly, objects moving vertically will not be seen as moving in depth. Incidental movement of objects will create spurious artifacts, and these incidental effects will be seen as artificial depth not related to actual depth in the scene.

Chromadepth system:

The ChromaDepth procedure of American Paper Optics is based on the fact that with a prism, colors are separated by varying degrees. The ChromaDepth eyeglasses contain special view foils, which consist of microscopically small prisms. This causes the image to be translated a certain amount that depends on its color. If one uses a prism foil now with one eye but not on the other eye, then the two seen pictures – depending upon color – are more or less widely separated. The brain produces the spatial impression from this difference. The advantage of this technology consists above all of the fact that one can regard ChromaDepth pictures also without eyeglasses (thus two-dimensional) problem-free (unlike with two-color anaglyph). However the colors are only limitedly selectable, since they contain the depth information of the picture. If one changes the color of an object, then its observed distance will also be changed.

1.3 STEREO WINDOW

For any branch of stereoscopy the concept of the stereo window is important. If a scene is viewed through a window the entire scene would normally be behind the window, if the scene is distant, it would be some distance behind the window, if it is nearby, it would appear to be just beyond the window. An object smaller than the window itself could even go through the window and appear partially or completely in front of it. The same applies to a part of a larger object that is smaller than the window. The goal of setting the stereo window is to duplicate this effect. To understand the concept of window adjustment it is necessary to understand where the stereo window itself is. In the case of projected stereo, including "3D" movies, the window would be the surface of the screen. With printed material the window is at the surface of the paper. When stereo images are seen by looking into a viewer the window is at the position of the frame. In the case of Virtual Reality the window seems to disappear as the scene becomes truly immersive.

Computer animated 2D films made with 3D models can be re-rendered in stereoscopic 3D by adding a second virtual camera if the original data is still available. This is technically not a conversion; therefore, such re-rendered films have the same quality as films originally produced in stereoscopic 3D. With the increase of films released in 3D, 2D to 3D conversion has become more common. The majority of non-CGI stereo 3D blockbusters are converted fully or at least partially from 2D footage. The reasons for shooting in 2D instead of stereo are financial, technical and sometimes artistic. Stereo post-production workflow is much more complex and not as well-established as 2D workflow, requiring more work and rendering. Professional stereoscopic rigs are much more expensive and bulky than customary monocular cameras. Some shots, particularly action scenes, can be only shot with relatively small 2D cameras. Stereo cameras can introduce various mismatches in stereo image, such as vertical parallax, tilt, color shift, reflections and glares in different positions, that should be fixed in post-production anyway because they ruin the 3D effect. This correction sometimes may have complexity comparable to stereo conversion. Stereo cameras can betray practical effects used during filming. The same scene filmed in stereo would reveal that the objects were not the same distance from the camera. By their very nature, stereo cameras have restrictions on how far the camera can be from the filmed subject and still provide acceptable stereo separation. For example, the simplest way to film a scene set on the side of a building might be to use a camera rig from across the street on a neighboring building, using a zoom lens. However, while the zoom lens would provide acceptable image quality, the stereo separation would be virtually nil over such a distance. Even in the case of stereo shooting, conversion can frequently be necessary. Besides the mentioned hard-to-shoot scenes, there are situations when mismatches in stereo views are too big to adjust, and it is simpler to perform 2D to stereo conversion, treating one of the views as the original 2D source.

Without respect to particular algorithms, all conversion workflows should solve the following tasks:

- Allocation of "depth budget" – defining the range of permitted disparity or depth, what depth value corresponds to the screen position, the permitted distance ranges for out-of-the-screen effects and behind-the-screen background objects. If an object in stereo pair is in exactly the same spot for both eyes, then it will appear on the screen surface and it will be in zero parallax. Objects in front of the screen are said to be in negative parallax, and background imagery behind the screen is in positive parallax. There are the corresponding negative or positive offsets in object positions for left and right eye images.
- Control of comfortable disparity depending on scene type and motion – too much parallax or conflicting depth cues may cause eye-strain and nausea effects

- Filling of uncovered areas – left or right view images show a scene from a different angle, and parts of objects or entire objects covered by the foreground in the original 2D image should become visible in a stereo pair. Sometimes the background surfaces are known or can be estimated, so they should be used for filling uncovered areas. Otherwise the unknown areas must be filled in by an artist or inpainted, since the exact reconstruction is not possible.

The fundamental principle underlying 2D-to-3D conversion techniques rests on the fact that stereoscopic viewing involves binocular processing by the human visual system of two slightly dissimilar images. The slight differences between the left-eye and right-eye images (horizontal disparities) are transformed into distance information such that objects are perceived at different depths and outside of the 2D display plane.

Most semiautomatic methods of stereo conversion use depth maps and depth-image-based rendering. The idea is that a separate auxiliary picture known as the "depth map" is created for each frame or for a series of homogenous frames to indicate depths of objects present in the scene. The depth map is a separate grayscale image having the same dimensions as the original 2D image, with various shades of gray to indicate the depth of every part of the frame.

The major steps of depth-based conversion methods are:

- Depth budget allocation – how much total depth in the scene and where the screen plane will be.
- Image segmentation – creation of mattes or masks, usually by rotoscoping. Each important surface should be isolated. The level of detail depends on the required conversion quality and budget.
- Depth map creation – Each isolated surface should be assigned a depth map. The separate depth maps should be composed into a scene depth map. This is an iterative process requiring adjustment of objects, shapes, depth, and visualization of intermediate results in stereo. Depth micro-relief, 3D shape is added to most important surfaces to prevent the "cardboard" effect when stereo imagery looks like a combination of flat images just set at different depths.
- Stereo generation based on 2D+Depth with any supplemental information like clean plates, restored background, transparency maps, etc. When the process is complete, a left and right image will have been created. Usually the original 2D image is treated as

the center image, so that two stereo views are generated. However, some methods propose to use the original image as one eye's image and to generate only the other eye's image to minimize the conversion cost. During stereo generation, pixels of the original image are shifted to the left or to the right depending on depth map, maximum selected parallax, and screen surface position.

- Reconstruction and painting of any uncovered areas not filled by the stereo generator.

Depth image is a 2D image that gives depth value to a point on an object in real scene according to its image coordinates. Some of the desirable characteristics of depth map are;

- The resolution of the depth map may be lower than that of the associated 2D image
- It can be highly compressed
- 2D compatibility is maintained
- Real time generation of stereo, or multiple stereo pairs, is possible

The human brain integrates the heuristic depth cues for the generation of the depth perception. Depth perceptions are mainly classified as binocular and monocular cues. Binocular cues are based on the recipient of information from both the eyes whereas monocular cues are from single eye. Monocular cues which include focus/defocus, motion parallax, relative height/size, and texture gradient provide various depth perceptions based on human experience. Therefore, humans are able to perceive depth from the single-view image/video. The key step in 2D to 3D conversion process is the generation of a dense depth. In recent years, a number of depth generation algorithms have been proposed according to the principle of human visual system. Several methods capable of generating 3D content require specific devices and are only effective in generating new content, making them infeasible for 2D contents. The main disadvantage has been the manual conversion techniques used to create depth maps, which results in a slow and costly process. 2D contents that require time-consuming manual editing of the depth information necessitates the development of an efficient 2D-to-3D conversion system.

Depth map generation methods can classified mainly into single-frame and multi-frame methods. Single-frame methods include depth assignment using image classification, machine learning, depth from focus/defocus, depth from geometric perspective, depth from texture gradient, and depth from relative height. Multi-frame based methods include triangular vision from stereo/multi-view and depth from motion. Depth from motion parallax is a temporal relation of depth from triangular vision. 2D-to-3D depth generation algorithms generally face two challenges. One is the depth uniformity inside the same object. The other involves retrieving an appropriate depth relationship among all objects.

CHAPTER 2

LITERATURE SURVEY

[1] DEPTH MAP GENERATION FOR 2D-TO-3D CONVERSION BY SHORT-TERM MOTION ASSISTED COLOR SEGMENTATION:

Yu-Lin Chang, Chih-Ying Fang, Li-Fu Ding, Shao-Yi Chen, and Liang-Gee Chen

In this paper the authors presented a novel depth map generation method – the short-term motion assisted color segmentation, which combines the pictorial, monocular and binocular depth cues of human vision. In this paper, the three depth cues are considered together to produce a temporally and spatially smooth depth map. A short-term motion assisted color segmentation for depth map generation is presented. It utilizes the color information in the image to find out the objects. Color segmentation, motion segmentation, and background registration technique are the main methods for segmentation. In order to deal with the color variance in the video sequence, the motion segmentation helps to extract the object boundary from the moving areas. The background registration registers the background information in the memory and subtracts the background from the captured image. Then the foreground objects are found. This method outputs a smooth depth map for the 2Dto- 3D conversion either in the spatial or temporal domain. It contains four parts: motion/edge detection, K-means algorithm for color segmentation, connected component, and motion/image segment adaptation. Thus a short-term motion assisted color segmentation, which is a combination of an online short-term motion segmentation and color segmentation to produce smooth depth maps both in the spatial and temporal domain is presented. However, this method faces the moving cameras problem and the tuning of various different type image sequences in the future. It should be combined with the depth from geometry perspective and other depth cues to produce more accurate depth map.

[2] A NOVEL 2D-TO-3D CONVERSION SYSTEM USING EDGE INFORMATION

Chao-Chung Cheng, Student Member, IEEE, Chung-Te Li, and Liang-Gee Chen, Fellow, IEEE

This work presents an algorithm that automatically converts 2D videos into 3D ones. The proposed algorithm utilizes the edge information to segment the image into object groups. A depth map is then assigned based on a hypothesized depth gradient model. Next, the depth map is block-based assigned by cooperating with a cross bilateral filter to generate visually comfortable depth maps efficiently and also diminish the block artifacts. A multiview video can be readily generated by using a depth imagebased rendering method. This method is based on the fact that the edge of an image has a high probability of being the edge of the depth map. After the pixels are grouped together, a relative depth value can be assigned to each region using an initial depth hypothesis. Next, the blocky artifact is removed using cross bilateral filtering. Finally, multi-view images are rendered by depth image-based rendering (DIBR) and display on a 3D display. This algorithm is quality-scalable depending on the block size. Smaller block size will result in better depth detail and large block size will have lower computational complexity. Capable of generating a comfortable 3D effect, the proposed algorithm is highly promising for 2D-to-3D conversion in 3D applications.

[3] A BLOCK-BASED 2D-TO-3D CONVERSION SYSTEM WITH BILATERAL FILTER

Chao-Chung Cheng, Chung-Te Li, Po-Sen Huang, Tsung-Kai Lin, Yi-Min Tsai, and Liang-Gee Chen Graduate Institute of Electronics Engineering, National Taiwan University, Taiwan, R.O.C.

This paper describes an automatic and robust system to convert 2D videos to 3D videos. combines two major depth generation modules, the depth from motion and depth from geometrical perspective. A block-based algorithm is applied and cooperates with the bilateral filter to diminish block effect and generate comfortable depth map. After generating the depth map, the multi-view video is rendered to 3D display. The depth fusion module fuses the depth map D(x, y) produced by DMP and the depth map produced by DGP according to weighting factors *Wm* and *Wp*. The DMP and DGP modules are all implemented by block-based algorithm to ease the hardware implementation. A cross bilateral filter is then applied to remove block artifact of depth map. The DIBR renders multiple views with various view points for 3D displays. This method is a hardware oriented algorithm and is suitable for VLSI design.

[4] GEODESIC DISTANCE AND MST BASED IMAGE SEGMENTATION

George Economou, Vassilios Pothos and Apostolos Ifantis

In this work, the integration of spatial proximity information in graph based segmentation algorithms is carried out. This is done by means of the geodesic distance. Distance calculation and the implementation of the method are carried out using the minimal spanning tree (MST), constructed on a watershed image partition. Distance, defined over the MST edges, presents a measure of both spatial and feature coherence. It is incorporated in

MST based color image segmentation applications, by means of a new density feature, which is computed with spatial locality restrictions. Density estimation is proposed by means of the minimal spanning tree and applied for image segmentation. It is in general helpful in detecting clusters that are internally homogeneous and constrained by a spatial neighbourhood structure. Besides image segmentation, this novel spatially local density estimate can find other applications as i.e. in densitybased sampling for data mining.

[5] 3D-TV CONTENT GENERATION: 2D-TO-3D CONVERSION

Wa James Tam, Liang Zhang

In this paper the authors provided an overview of the fundamental principle underlying 2D-to-3D conversion techniques, a cursory look at a number of approaches for depth extraction using a single image, and a highlight of the potential use of surrogate depth maps in depth image based rendering for 2D-to-3D conversion. The ability of the human visual system is exploited to combine reduced disparity information that are located mainly at edges and object boundaries with pictorial depth cues to produce an enhanced sensation of depth over 2D images. Generation of depth maps is an important prerequirement for depth image based rendering which is a useful technique for 2D-to-3D conversion of images and video. Techniques are required to avoid user interaction as far as possible and in reducing computational complexity.

[6] EFFICIENT DEPTH IMAGE BASED RENDERING WITH EDGE DEPENDENT DEPTH FILTER AND INTERPOLATION

Wan-Yu Chen, Yu-Lin Chang, Shyh-Feng Lin, Li-Fu Ding, and Liang-Gee Chen

An efficient depth image based rendering with edge dependent depth filter and interpolation is presented in this paper. Depth-Image-Based-Rendering(DIBR) is a key technology in advanced three dimensional television system. The main problem to be solved in DIBR system is to reduce the number of big hole while keeping the subjective view quality. There are three features proposed in this method: edge dependent depth filter, edge oriented interpolation, and vertical edge rectification. The hole-location is detected by special edge filter and the pre-processing of depth map is operated on the detected edges. The edge dependent interpolation method is used to preserve edge information of the interpolated area. Vertical edge rectification provides depth accuracy along vertical line. Thus this method fills the holes without defects by edge dependent Gaussian filter and interpolation.

[7] DEPTH-MAP GENERATION BY IMAGE CLASSIFICATION

S. Battiato, S. Curti, M. La Cascia, M. Tortora, E. Scordato

This paper describes about a technique to estimate depth information from a single input image. It based on a new image classification technique able to classify digital images 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 input image is processed by the following steps: Bayer to approximated-RGB color conversion, Color-based segmentation, Rule-based regions detection to find specific areas (e.g. sky, land, mountain, etc.), Image classifications to discriminate between outdoor with or without geometric elements and indoor images, approximated depth map estimation. This method includes 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.

[8] CONVERSION 2D IMAGE TO 3D BASED ON SQUEEZE FUNCTION AND GRADIENT MAP

Lee Sang-Hyun, Park Dae-Won, Jeong Je-Pyong and Moon Kyung-Il

This paper proposes an algorithm that automatically converts 2D images into 3D ones. It is composed of the estimation of depth levels by using modulation transfer function (MTF) squeeze model and determination of gradient map related to each depth level. The grouping is based on the pixels having similar colors and spatial locality. Based on a depth gradient map, a depth level is assigned. Next, the depth map is assigned by cooperating with a cross bilateral filter to diminish the blocky artifacts. The key step in the conversion process is the generation of a dense depth. This algorithm uses a simple depth hypothesis to assign the depth of each group instead of retrieving the depth value directly from the depth cue based on region of interest. Cross bilateral filter is then used to enhance the visual comfort. This method is based on ROI boundary by MTF squeeze function and estimation of depth level through gradient map. Smaller block size will result in better depth detail and that of the larger block size will have lower computational complexity.

[9] DEPTH MAP GENERATION USING LOCAL DEPTH HYPOTHESIS FOR 2D-TO-3D CONVERSION

Na-Eun Yang, Ji Won Lee, and Rae-Hong Park

In this paper the authors proposed an interactive method of depth map generation from a single image for 2D-to-3D conversion. Using a hypothesis of depth variation can reduce the human effort to generate a depth map. The only thing required from a user is to mark some salient regions to be distinguished with respect to depth variation. This algorithm makes hypothesis of each salient region and generates a depth map of an input image. This method is a semi-automatic and simple method with a little intervention from user. It consists of four parts: scene grouping, local depth hypothesis generation, depth assignment and depth map refinement. The visual quality of the proposed method depends on result of scene grouping. In depth assignment step, the depth is assigned according to scene grouping result. Because detail structures of the depth map follow the result of scene grouping, the scene grouping should segment an image to faithfully represent distinct detail structures. This system takes more computation than other methods because it needs additional salient segmentation process. However, these days cloud computing service is being started. Cloud computing services can reduce the load of hardware implementation. It delivers computing data via Internet to make server compute the process. Therefore, this method can be applied to such consumer devices or applications.

[10] 2D-TO-3D STEREOSCOPIC CONVERSION: DEPTH-MAP ESTIMATION IN A 2D SINGLE-VIEW IMAGE

Jaeseung Ko, Manbae Kim, and Changick Kim

In this paper the authors focus on automatic conversion method that estimates the depth information of a single-view image based on degree of focus of segmented regions and then generates a stereoscopic image. Because of the limited ability of human-visual system or camera system for depth of focus, while closer regions are shown clearly, the regions which are far from viewer tend to be blurred. In low depth of field images , the focused region can be assumed to be a foreground closer to the viewers and the defocused regions are regarded as background which is far from the viewers. This focus cue can be measured using higher-order statistics (HOS). HOS information is combined with region information obtained by existing image segmentation technique to yield more accurate foreground extraction results. Then the relative depth map can be refined through the post-processing, and finally synthesize new left-view image and right-view image. These synthesized images can be used to generate 3D content, which is stereoscopic image.

[11] BILATERAL FILTERING FOR GRAY AND COLOR IMAGES

C. Tomasi, R. Manduchi

In this paper, the authors proposed a non-iterative scheme for edge preserving smoothing that is inoniterative and simple. Bilateral filtering smooths images while preserving edges, by means of a nonlinear combination of nearby image values. The method is noniterative, local, and simple. It combines gray levels or colors based on both their geometric closeness and their photometric similariv, and prefers near values to distant values in both domain and range. In contrast with filters that operate on the three bands of a color image separately, a bilateral filter can enforce the perceptual metric underlying the CIE-Lab color space, and smooth colors and preserve edges in a way that is tuned to human perception. Also, in contrast with standardjltering, bilateral filtering produces no phantom
colors along edges in color images, and reduces phantom colors where they appear in the original image. The generality of bilateral filtering is analogous to that of traditional filtering,
which is called domain filtering. The explicit enforcement of a photometric distance in the range component of a bilateral filter makes it possible to process. color images in a perceptually appropriate fashion.

In this chapter a survey of various papers related to 3D conversion system is done. Several algorithms are proposed for better visual quality. For the conversion process many algorithms related to depth generation is studied and a detailed study of DIBR and stereoscopic generation is made.

CHAPTER 3

GENERATION OF DEPTH MAP

3.1 DEPTH MAP

In 3D computer graphics a depth map is an image or image channel that contains information relating to the distance of the surfaces of scene objects from a viewpoint. Each depth image stores depth information as 8-bit grey values with the grey level 0 indicating the furthest value and the grey level 255 specifying the closest value. The extraction of depth is the crucial one in the conversion process. The greatest difference between 2D and 3D image is the depth information. The object can jump out of the screen and look like a real life due to the depth information. If we extract these depth signals and integrate them together, we will build a strong foundation to make 3D images of better and higher quality. The depth generation algorithms are roughly classified into three categories which utilize different kinds of depth cues: the binocular, monocular and pictorial depth cues.

3.1.1 MONOCULAR CUES

Motion parallax:

When an observer moves, the apparent relative motion of several stationary objects against a background gives hints about their relative distance. If information about the direction and velocity of movement is known, motion parallax can provide absolute depth information. This effect can be seen clearly when driving in a car. Nearby things pass quickly, while far off objects appear stationary.

Depth from motion:

When an object moves toward the observer, the retinal projection of an object expands over a period of time, which leads to the perception of movement in a line toward the observer. Another name for this phenomenon is depth from optical expansion. The dynamic stimulus change enables the observer not only to see the object as moving, but to perceive the distance of the moving object. Thus, in this context, the changing size serves as a distance cue.

Relative size:

If two objects are known to be the same size (e.g., two trees) but their absolute size is unknown, relative size cues can provide information about the relative depth of the two objects. If one subtends a larger visual angle on the retina than the other, the object which subtends the larger visual angle appears closer.

Familiar size:

Since the visual angle of an object projected onto the retina decreases with distance, this information can be combined with previous knowledge of the object's size to determine the absolute depth of the object. For example, people are generally familiar with the size of an average automobile. This prior knowledge can be combined with information about the angle it subtends on the retina to determine the absolute depth of an automobile in a scene.

Texture gradient:

Fine details on nearby objects can be seen clearly, whereas such details are not visible on faraway objects. Texture gradients are grains of an item. For example, on a long gravel road, the gravel near the observer can be clearly seen of shape, size and colour. In the distance, the road's texture cannot be clearly differentiated.

Defocus blur:

Selective image blurring is very commonly used in photographic and video for establishing the impression of depth. This can act as a monocular cue even when all other cues are removed. It may contribute to the depth perception in natural retinal images, because the depth of focus of the human eye is limited. In addition, there are several depth estimation algorithms based on defocus and blurring.

Kinetic depth effect:

If a stationary rigid figure, for example cube, is placed in front of a point source of light so that its shadow falls on a translucent screen, an observer on the other side of the screen will see a two-dimensional pattern of lines. But if the cube rotates, the visual system will extract the necessary information for perception of the third dimension from the movements of the lines, and a cube is seen. This is an example of the kinetic depth effect. The effect also occurs when the rotating object is solid, provided that the projected shadow consists of lines which have definite corners or end points, and that these lines change in both length and orientation during the rotation.

Perspective:

The property of parallel lines converging in the distance, at infinity, allows us to reconstruct the relative distance of two parts of an object, or of landscape features. An example would be standing on a straight road, looking down the road, and noticing the road narrows as it goes off in the distance.

Absolute size:

Even if the actual size of the object is unknown and there is only one object visible, a smaller object seems further away than a large object that is presented at the same location.

Aerial perspective :

Due to light scattering by the atmosphere, objects that are a great distance away have lower luminance contrast and lower color saturation. Due to this, images seem hazy the farther they are away from a person's point of view. In computer graphics, this is often called "distance fog." The foreground has high contrast; the background has low contrast. Objects differing only in their contrast with a background appear to be at different depths. The color of distant objects are also shifted toward the blue end of the spectrum (e.g., distant mountains). Some painters, employ "warm" pigments (red, yellow and orange) to bring features forward towards the viewer, and "cool" ones (blue, violet, and blue-green) to indicate the part of a form that curves away from the picture plane.

Accommodation:

This is an oculomotor cue for depth perception. When we try to focus on far away objects, the ciliary muscles stretch the eye lens, making it thinner, and hence changing the focal length. The kinesthetic sensations of the contracting and relaxing ciliary muscles (intraocular muscles) is sent to the visual cortex where it is used for interpreting distance/depth. Accommodation is only effective for distances less than 2 meters.

Occlusion:

Occlusion (also referred to as interposition) happens when near surfaces overlap far surfaces. If one object partially blocks the view of another object, humans perceive it as closer. However, this information only allows the observer to create a "ranking" of relative nearness. The presence of monocular occlusions consist of the object's texture and geometry. Monocular occlusions are able to reduce the depth perception latency both in natural and artificial stimuli.

Curvilinear perspective:

At the outer extremes of the visual field, parallel lines become curved, as in a photo taken through a fisheye lens. This effect, although it is usually eliminated from both art and photos by the cropping or framing of a picture, greatly enhances the viewer's sense of being positioned within a real, three-dimensional space. Classical perspective has no use for this so-called "distortion," although in fact the "distortions" strictly obey optical laws and provide perfectly valid visual information, just as classical perspective does for the part of the field of vision that falls within its frame.

Lighting and shading:

The way that light falls on an object and reflects off its surfaces, and the shadows that are cast by objects provide an effective cue for the brain to determine the shape of objects and their position in space.

Elevation:

When an object is visible relative to the horizon, we tend to perceive objects which are closer to the horizon as being farther away from us, and objects which are farther from the horizon as being closer to us.

3.1.2 BINOCULAR CUES

Binocular parallax:

By using two images of the same scene obtained from slightly different angles, it is possible to triangulate the distance to an object with a high degree of accuracy. Each eye views a slightly different angle of an object seen by the left and right eyes. This happens because of the horizontal separation parallax of the eyes. If an object is far away, the disparity of that image falling on both retinas will be small. If the object is close or near, the disparity will be large.

Convergence:

This is a binocular oculomotor cue for distance/depth perception. Because of stereopsis the two eyeballs focus on the same object. In doing so they converge. The convergence will stretch the extraocular muscles. As happens with the monocular accommodation cue, kinesthetic sensations from these extraocular muscles also help in depth/distance perception. The angle of convergence is smaller when the eye is fixating on far away objects. Convergence is effective for distances less than 10 meters.

Shadow stereopsis:

The retinal images with no parallax disparity but with different shadows are fused stereoscopically, imparting depth perception to the imaged scene. This is referred to as shadow srereopsis.

3.2 APPLICATIONS OF DEPTH MAP

Depth maps have a number of uses, including:

- Simulating the effect of uniformly dense semi-transparent media within a scene - such as fog, smoke or large volumes of water.
- Simulating shallow depths of field - where some parts of a scene appear to be out of focus. Depth maps can be used to selectively blur an image to varying degrees. A shallow depth of field can be a characteristic of macro photography and so the technique may form a part of the process of miniature faking.
- Z-buffering and z-culling, techniques which can be used to make the rendering of 3D scenes more efficient. They can be used to identify objects hidden from view and which may therefore be ignored for some rendering purposes. This is particularly important in real time applications such as computer games, where a fast succession of completed renders must be available in time to be displayed at a regular and fixed rate.
- Shadow mapping - part of one process used to create shadows cast by illumination in 3D computer graphics. In this use, the depth maps are calculated from the perspective of the lights, not the viewer.
- To provide the distance information needed to create and generate autostereograms and in other related applications intended to create the illusion of 3D viewing through stereoscopy .
- Subsurface scattering - can be used as part of a process for adding realism by simulating the semi-transparent properties of translucent materials such as human skin.

3.3 METHODS TO GENERATE DEPTH MAP

A number of devices capable of capturing depth maps in real-time, in synchronism with the 2D source are now commercially available. These include 3DV's 'Z-Cam' and other sensors based on scanning lasers. These systems both enable live broadcasting and eliminate the need for content conversion. Whilst live recording will most certainly be the dominant process in the future, there are still significant challenges in educating the existing 2D content creators in this new art as well as the costs associated with equipping studios with such technology.

In the meantime the conversion of 2D content, either pre-existing or recorded specifically for the purpose of display on a 3D screen, is a commercially viable alternative. Given the vast library of existing 2D material the consumer is ensured of both compelling and current content. Conversion from 2D to 3D, of pre-existing content, based on the generation of depth maps is now an established process. The main disadvantage of the technique has been the manual nature of the majority of techniques used to create depth maps, which resulted in a slow and costly process.

There are a number of manual techniques that are currently used to produce depth maps, which include:

- Hand drawn object outlines manually associated with an artistically chosen depth value; and
- Semi-automatic outlining with corrections made manually by an operator.

Each of these has a number of drawbacks. Hand drawing produces high quality depth maps but is very time consuming and expensive. Semi-automatic outlining is generally unreliable where complex outlines are encountered. Although the fully automated recovery of depth from monocular image sequences is possible under certain conditions, the operational constraints associated with such techniques limit their commercial viability.

Depth maps can be generated using direct methods, such as using a ZCam that directly measures the distance of objects in a scene by employing the time required to bounce a beam of infra-red light back to a sensor located on the camera. Another direct method is to project a structured light pattern onto the scene so that the depths of the various objects could be recovered by analyzing the distortions of the light pattern created by the 3D shape of objects in the scene. However, aside from requiring specialized hardware, the direct methods have other drawbacks such as restrictive scene lighting and the need to have objects within a restrained distance.

There are a lot of depth generation methods proposed to retrieve the depth map from different kinds of images and videos. The brief introduction is as follows:

- A computed image depth (CID) method to find out the depth relations from the part of an image.
- A feasible depth from focus algorithm to obtain the depth map from multiple different focus distance images.
- Another still image analysis method which utilizes the machine learning algorithm to find the feature points for depth mapping.
- Generation of a depth map based on several steps including the generation of gradient planes, depth gradient assignment, consistency verification of the detected region, and finally the depth map generation
- A vanishing line detection method which uses block-based algorithm to extract the vanishing lines and point.
- A motion detector and grouping method to assign different depth to each group.
- Use the motion segmentation, motion estimation, and object tracking algorithms to find out the object and its possible depth map.
- A visual modeling method by the motion parallax cue.
- Some reviews of the motion parallax based 3D reconstruction methods.
- A simple line tracing method.

2D to 3D depth generation algorithms generally face two challenges. One is the depth uniformity inside the same object. Because the image/video consists of 2D pixel arrays, information about the object grouping relation of pixels is lacking. A better grouping of pixels implies a better outcome for the depth uniformity inside the object. An effective grouping method should consider both color similarity and spatial distance. The other challenge involves retrieving an appropriate depth relationship among all objects. Some previous works, integrate depth information with the object grouping concept. However, these methods use motion parallax as the primary depth cue. These methods obtain false depth information when the object with different self motion vectors. The pixels belong to the same object that may be assigned with different depth values. Generating a depth map from single 2D images is an ill-posed problem. Not all the depth cues can be retrieved from an image or multiple consecutive frames. To overcome these two challenges, the algorithm uses a simple depth hypothesis to assign the depth of each group rather than retrieving the depth value directly from the depth cue.

CHAPTER 4
EXISTING AND PROPOSED METHOD

4.1 EXISTING METHOD:

Many approaches are available to generate 3D images which include active depth sensor, triangular stereo vision, and 3D graphics rendering. They are mainly categorized into active and passive methods. Active methods use active sensors, such as structured light and time-of-flight sensor, to retrieve depth maps. Triangular stereo vision requires multiple cameras to record the content. These methods require specific devices to generate the new content. So this is a time consuming method of manually editing the depth map for the content. A typical 2D-to-3D conversion system which automatically generates depth information from a single view image/video to the multi-view image/video using the estimated depth maps. The major depth perceptions are binocular depth cues from two eyes and monocular depth cues from a single eye. Monocular cues, including focus/defocus, motion parallax, relative height/size, and texture gradient, provide various depth perceptions based on human experience. Therefore, humans can also perceive depth from the single-view image/video.

Depth map generation methods can classified mainly into single-frame and multi-frame methods. Single-frame methods include depth assignment using image classification , machine learning, depth from focus/defocus, depth from geometric perspective, depth from texture gradient, and depth from relative height. Battiato generated depth maps using multiple cues from a single image. However, the above methods are unreliable for cases missing during the training phase. The computed image depth (CID) method divides a single image into several sub-blocks and uses contrast and blurriness information to generate depth information for each block.

Multi-frame based methods include triangular vision from stereo/multi-view and depth from motion. Depth from motion parallax is a temporal relation of depth from triangular vision. The correspondences of relative frames/views are estimated to retrieve the depth from the disparity. In the depth from motion parallax approaches, the depth-induced motion vector (MV) is converted to disparity vector (DV). Another approach, modified time difference method (MTD), generates stereo pairs without a depth map. The modified time difference frame pairs are selected as left-eye and right-eye images to generate a stereoscopic effect.

4.1.1 DISADVANTAGES OF EXIXTING SYSTEM

- Active methods require specific devices and are only efficient in generating new content making them infeasible for 2D contents.
- The reconstruction process is complex and turns out to be a high budget project.
- 2D contents require time consuming manual editing of depth information and this necessitates the development of an efficient 2D to 3D conversion system

4.2 PROPOSED SYSTEM

An automatic conversion of 2D to 3D images using depth information is proposed. The main aim is to reconstruct the 3D image with better visual quality using a single view of 2D image. Depth image based rendering technology is used to provide a 3D vision without requiring the left and right view of the image. With this technology the intermediate view of the image is sufficient to reconstruct the 3D content. The work here describes an efficient 2D-to-3D conversion method based on the use of depth information. Importantly, the edge of an image has a high probability as it can be the edge of the depth map. The key step in 2D to 3D conversion process is the generation of a dense depth. In recent years, a number of depth generation algorithms have been proposed according to the principle of human visual system. Every algorithm has its own strengths and weakness. Most depth extraction algorithms make use of a certain depth cue but few of them combines two or more depth cues to generate depth map.

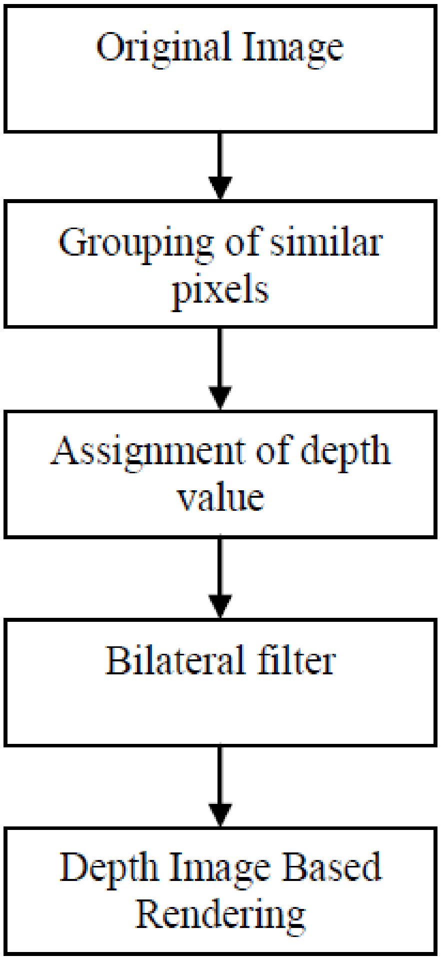

Original Image
Grouping of similar pixels
Assignment of depth value
Bilateral filter
Depth Image Based Rendering

Figure 4.1 Conversion algorithm

The algorithm for the conversion process is as follows. Initially the input image is segmented into groups of similar pixels using k-means segmentation algorithm. With the help of an effective grouping method, the pixels having similar colors and spatial locality are grouped. After the pixels are grouped together, a relative depth value can be assigned to each region. Importantly, the edge of an image has a high probability of being the edge of the depth map. Once the pixels are grouped together, the depth of each segment is assigned with the help of an initial depth hypothesis. Next, the blocky artifacts have to be removed using cross bilateral filtering. Finally, multi-view images are obtained by the method of DIBR. As a result, the input 2D image is converted into visually comfortable 3D image without the presence of artifacts enhancing the quality of the image in the display. The algorithm mainly focuses on the generation of dense depth since the visual quality of the content is highly dependent on the depth map.

4.3 K-MEANS SEGMENTATION

K-Means algorithm is an unsupervised clustering algorithm that classifies the input data points into multiple classes based on their inherent distance from each other. The algorithm assumes that the data features form a vector space and tries to find natural clustering in them. The points are clustered around centroids µi∀i = 1 . . . k which are obtained by minimizing the objective

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} (x_j - \mu_i)^2 \qquad (1)$$

where there are k clusters S_i , i = 1, 2, . . . , k and μ_i is the centroid or mean point of all the points $x_j \in S_i$. The algorithm takes a 2 dimensional image as input. Various steps in the algorithm are as follows:

- Compute the intensity distribution(also called the histogram) of the intensities.
- Initialize the centroids with k random intensities.
- Repeat the following steps until the cluster labels of the image does not change anymore.
- Cluster the points based on distance of their intensities from the centroid intensities.

$$C^{(i)} = \arg min_j \left\| x^{(i)} - \mu_j \right\|^2 \qquad (2)$$

- Compute the new centroid for each of the clusters.

$$\mu_i = \frac{\sum_{i=1}^{m} 1\{C_{(i)}=j\} x^{(i)}}{\sum_{i=1}^{m} 1\{C_{(i)}=j\}} \qquad (3)$$

where k is a parameter of the algorithm (the number of clusters to be found), i iterates over the all the intensities, j iterates over all the centroids and μ_i are the centroid intensities.

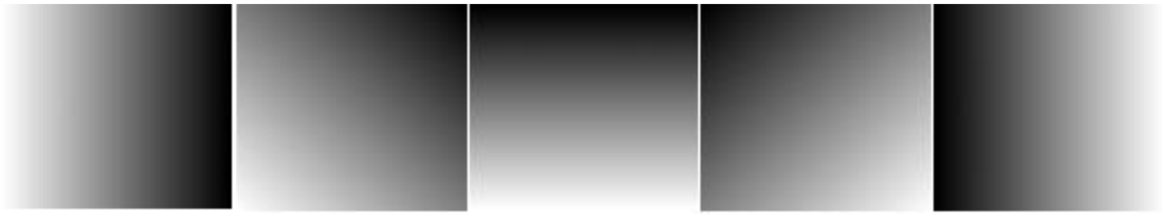
4.4 DEPTH FROM PRIOR HYPOTHESIS

The extraction of depth is the crucial one in the conversion process. The greatest difference between 2D and 3D image is the depth information. The object can jump out of the screen and look like a real life due to the depth information. If we extract these depth signals and integrate them together, we will build a strong foundation to make 3D images of better and higher quality. Following generation of the block groups, the corresponding depth for each block is assigned by the hypothesized depth gradient. The initial depth gradient hypothesis is derived based on the linear perspective information. The depth value of a given block group *G* is assigned by:

$$Depth(G) = 128 + 255\left\{\sum_{p(x,y)\in G}\{Wrl\frac{x-\frac{w}{2}}{w} + Wud\frac{y-\frac{h}{2}}{h}\}\right\}/pixel\_num(G) \qquad (4)$$

where $|W_{rl}| + |W_{ud}| = 1$.

This implies that a pixel closer to the user is assigned with a larger depth value. The depth map is usually represented by greyscale. In other words the closest point will be brighter when compared to the farthest point in the region. The depth values are assigned in such a way that it forms the gravity centre of the block group. That is why each block in the same group is assigned with a same depth value. Based on the geometrical perspective of image the orientation of hypothesized gradients can be derived. The following figure shows the different depth map gradients:



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

Fig.4.2. Depth gradients: (a)left, (b)left-down, (c)bottom-up, (d)right-down, (e)right

The orientation of hypothesized depth gradient can be derived from analysis of a geometrical perspective of the images. Analysis results indicate that the bottom-up mode is the most important mode in the real world. The bottom-up mode is mainly chosen because it prevents most of the noise occurrence in the depth map to be obtained during the depth assignment using a hypothesis depth gradient. If the linear perspective fails to detect the scene mode, the bottom-up mode is selected as the default mode. Thus the depth map is generated by choosing the appropriate parameters in cases like the depth gradients, texture of the image and also the mathematical expression for obtaining a better depth image from the block groupings. The obtained depth map has a small extent of blocky artifacts during the generation of the depth image which has to be filtered before applying the required algorithm.

4.5 BILATERAL FILTERING

After the depth map has been generated, a recursive depth filtering operation has to be performed to extract the image features like color, luminance, edge, etc. According to the previous results, if discontinuity happens right at the edge boundary, a perfect 3D effect will be delivered without causing visual fatigue. In order to achieve this goal, we have to pass it through an edge-preserving smoother. To smooth the depth map preserving the edge information, a cross bilateral filter is required. This filter is particularly chosen to maintain the pixel information as if that of the input image.

The bilateral filter is a non-linear filter used for smoothening the images. It has been adopted for several applications such as image denoising, relighting and texture manipulation, dynamic range compression, illumination correction, and photograph enhancement. It has also been adapted to other domains such as mesh fairing, volumetric denoising, optical flow and motion estimation, and video processing. This large success stems from several origins. First, its formulation and implementation are simple: a pixel is simply replaced by a weighted mean of its neighbors. And it is easy to adapt to a given context as long as a distance can be computed between two pixel values.

The bilateral filter is also non-iterative, thereby achieving satisfying results with only a single pass. This makes the filter’s parameters relatively intuitive since their effects are not cumulated over several iterations. The bilateral filter has proven to be very useful, however it is slow. It is nonlinear and its evaluation is computationally expensive since traditional accelerations, such as performing convolution after an FFT, are not applicable. Brute-force computation is on the order of tens of minutes. Nonetheless, solutions have been proposed to speed up the evaluation of the bilateral filter. Unfortunately, most of these methods rely on approximations that are not grounded on firm theoretical foundations, and it is difficult to evaluate the accuracy that is sacrificed. The cross bilateral filter is a variant of the classical bilateral filter. This filter smoothes the image to locate the edges to preserve. The depth map generated by block-based region grouping contains blocky artifacts. Here, the blocky artifacts are removed by using the cross bilateral filter, as expressed in the following equation:

$$Depth_f(x_i) = \frac{1}{N(x_i)} \sum_{x_j \subset \Omega(x_i)} e^{-0.5\left(\frac{\left|x_j - x_i\right|^2}{\sigma_x^2} + \frac{\left|u(x_j) - u(x_i)\right|^2}{\sigma_r^2}\right)} Depth(x_j), \qquad (5)$$

$$N(x_i) = \sum_{x_j \subset \Omega(x_i)} e^{-0.5\left(\sigma_x^{-2}\left|x_j - x_i\right|^2 + \sigma_r^{-2}\left|u(x_j) - u(x_i)\right|^2\right)}$$

where $u(x_\mathrm{i})$ denotes the intensity value of the pixel *xi*, $\Omega(x_\mathrm{i})$ represents the neighboring pixels of x_i, $N(x_\mathrm{i})$ refers normalization factor of the filter coefficients and Depthf is the filtered depth map. The cross bilateral filter smoothens the depth map properly while preserving the object boundaries. The blocky artifact in the generated depth map is effectively removed while the sharp depth discontinuities along the object boundary are preserved.

4.6 DEPTH IMAGE-BASED RENDERING

Depth-Image-Based-Rendering (DIBR) is a key technology in advanced three dimensional television systems (3D TV System). Traditional 3D TV system requires the transmission of two video streams, the left and right view, to construct 3D vision. Unlike the traditional method, the advanced three dimensional television system proposed a novel technology “DIBR” to provide 3D vision. DIBR uses intermediate view and intermediate depth map to render left and right view. In this way, broadcast content providers only have to transmit the video and gray level depth map of the intermediate view. It has been proven that the coding efficiency is better than the transmission of two view color video stream. Another advantage is the 2D/3D selectivity. Users can change 3D vision into 2D vision only by displaying intermediate view. The more important is that left and right views are rendered according to the users’ parallax. Therefore, users can watch more comfortable 3D video by adjusting parallax of the rendered video.

Depth image is a 2D image that gives depth value to a point on an object in real scene according to its image coordinates. Once intermediate image and depth image is given, any nearby image can be synthesized by mapping pixel coordinates one by one according to its depth value. However, there is an essential problem in DIBR that occlusion holes appear after pixel to pixel mapping. Holes appear due to sharp horizontal changes in depth image, thus the location and size of holes differ from frame to frame.

The filtered depth map has a comfortable visual quality because the cross bilateral filter generates a smooth depth map inside the smooth region with similar pixel values and preserves sharp depth discontinuity on the object boundary. Following filtering by the cross bilateral filter, the depth map is used to generate the left/right or multi-view images using depth image-based rendering (DIBR) for 3D visualization. The disparity can be calculated based on the known depth information. Depending on the configurations of 3D displays, e.g., the stereoscopic or the multi-view displays, images for different view angles can be readily generated from the color image and the depth map using DIBR.

4.6.1STEREOSCOPIC IMAGE GENERATION

Parallax Generation

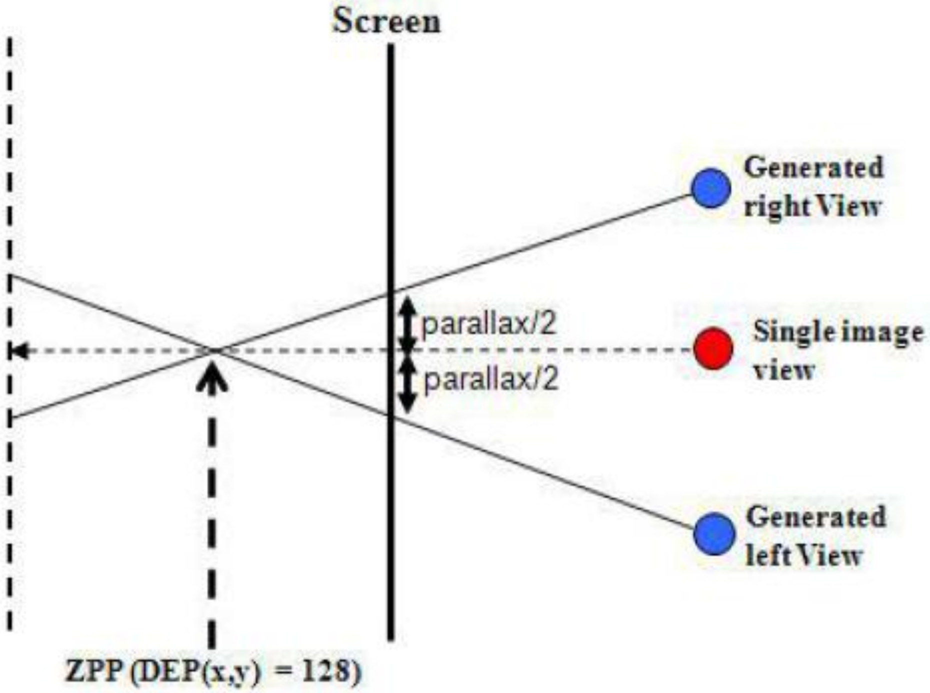


Fig.4.3. Right view and Left view generation.

To synthesize the left-view and the right-view images using the estimated depth map, we compute the parallax value for each pixel in an image from the estimated depth map, and then shift each pixel by corresponding parallax values in an input image. The parallax value at (x, y), *Parallax*(*x*, *y*) is computed from depth map as follows;

$$Parallax(x,y) = M \times \left(1 - \frac{dept\,h(x,y)}{128}\right) \quad (6)$$

where M denotes the maximum parallax value and *depth*(*x*, *y*) is the estimated depth value at (x,y). From this equation, the zero parallax plane (ZPP) is set to the region which has depth value of 128, so that the regions which have more than depth value of 128 have the negative parallax value and the regions which have lower than depth value of 128 have positive parallax value. Since we assign the negative parallax on the foreground regions, when the viewer see the stereoscopic image generated, they can feel that foreground regions are protruded out of the screen.

Stereoscopic Generation

The input image is considered as the center view of stereoscopic pair. We shift each pixel of the input image by the amounts of parallax(x, y)/2 to right direction to generate the right-view image. The left-view image can be obtained by the same process. In this way the left and right view of an image can be obtained using a single view.

CHAPTER 5

SIMULATION RESULTS

5.1 SOFTWARE REQUIREMENTS

The following are the minimum software requirements for this project.

OPERATING SYSTEM	: Windows XP service pack 3
PROGRAMMING TOOL	: MATLAB R2013a

5.2 SIMULATION RESULTS

The step by step process of the algorithm is simulated and the outputs for the set of images are displayed as follows;

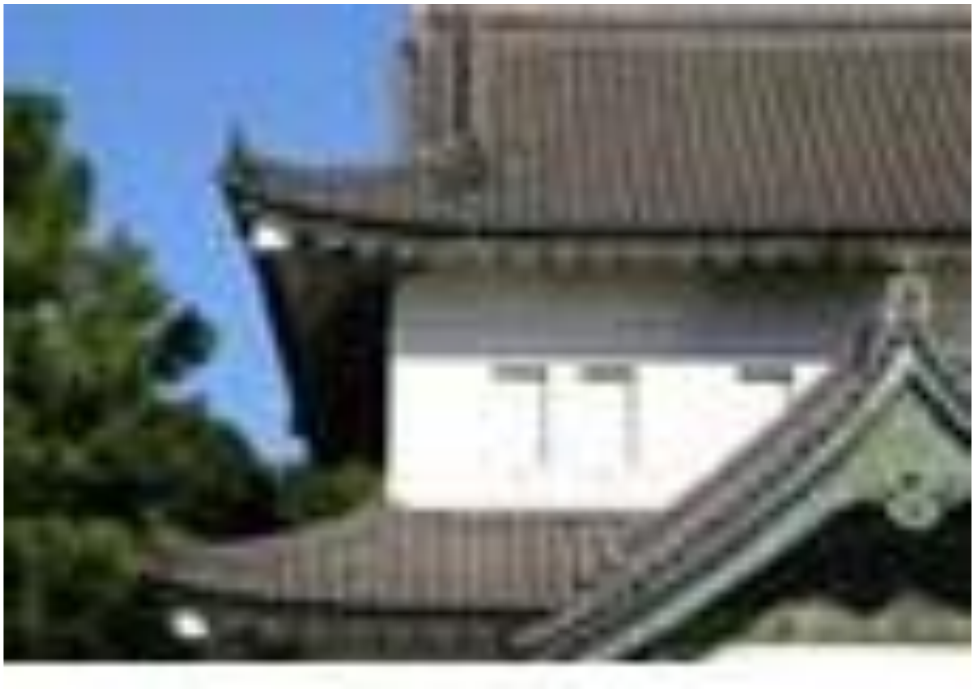


Fig.5.1.(a). Input image-‘house.jpg’

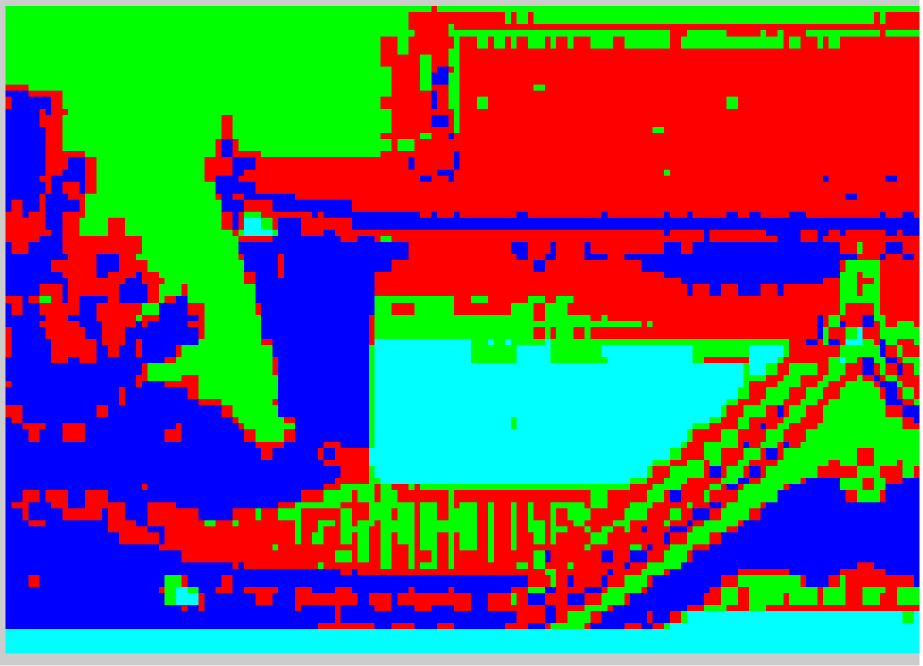


Fig.5.1.(b) K-means segmentation output

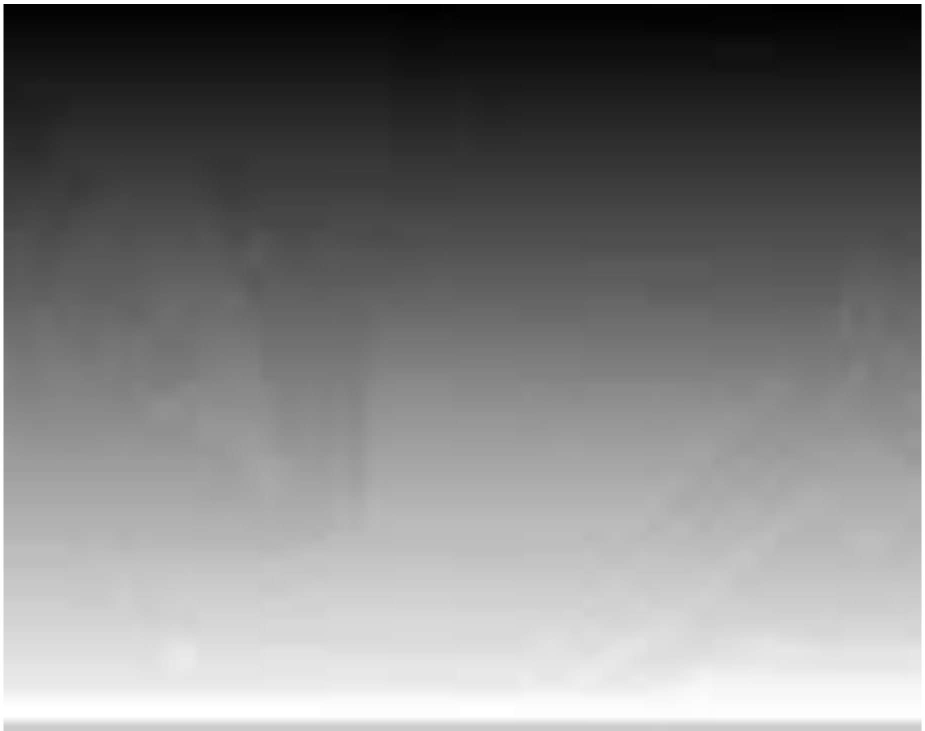


Fig5.1.(c) Depth map assigned image

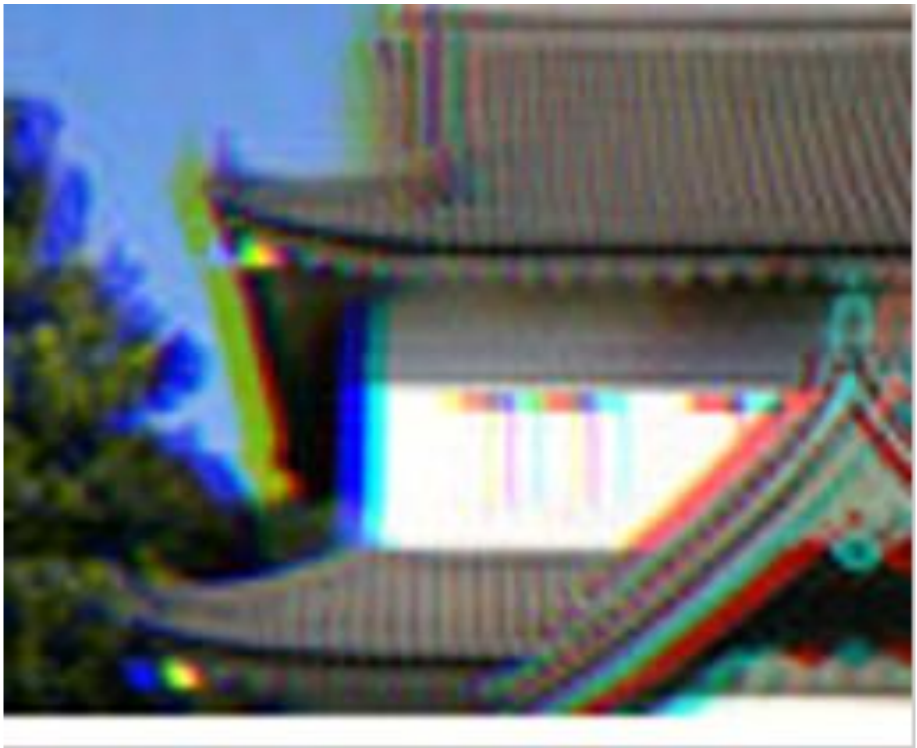


Fig.5.1.(d) 3D output

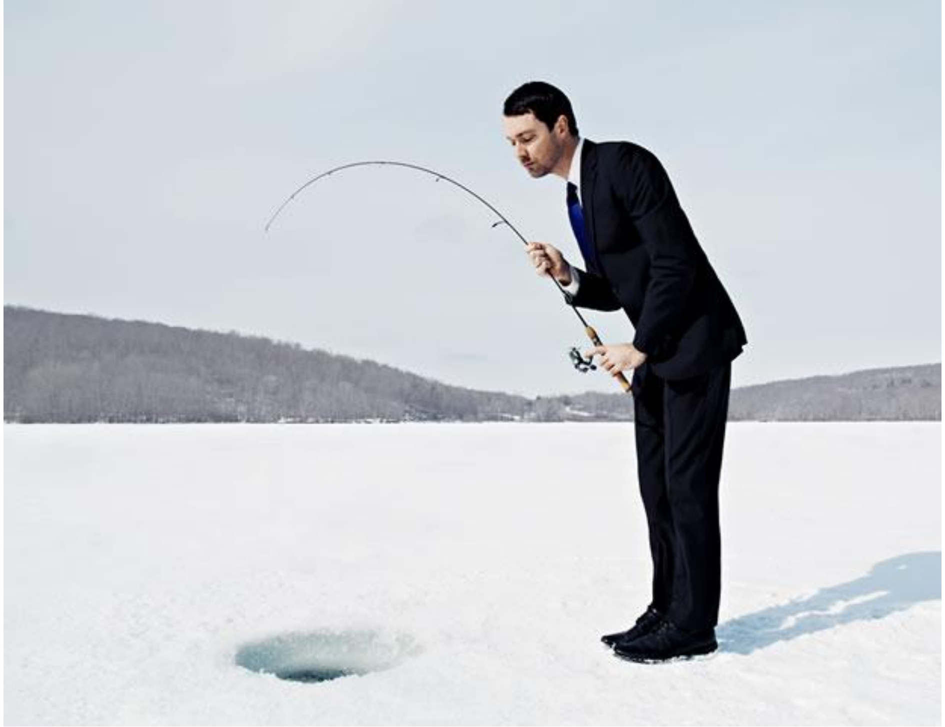


Fig.5.2.(a) Input image-‘Man.jpg’

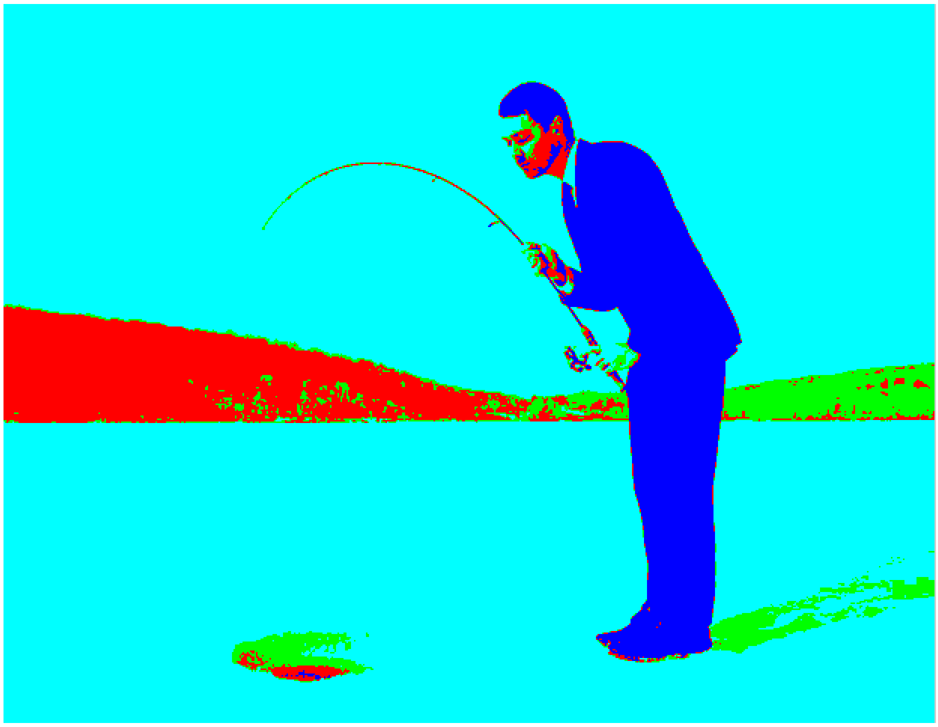


Fig.5.2.(b) K-means segmentation output

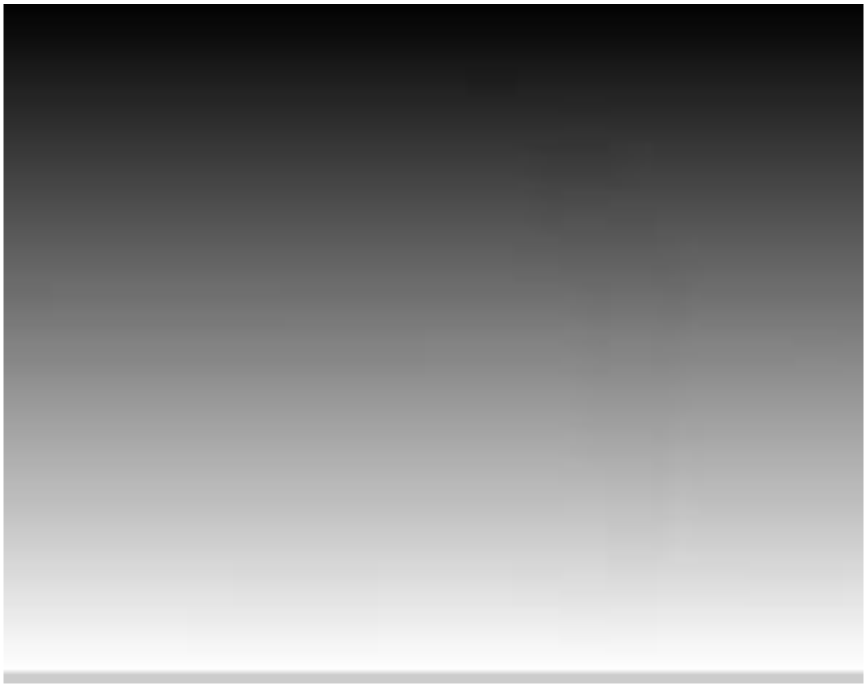


Fig.5.2.(c) Depth assigned image

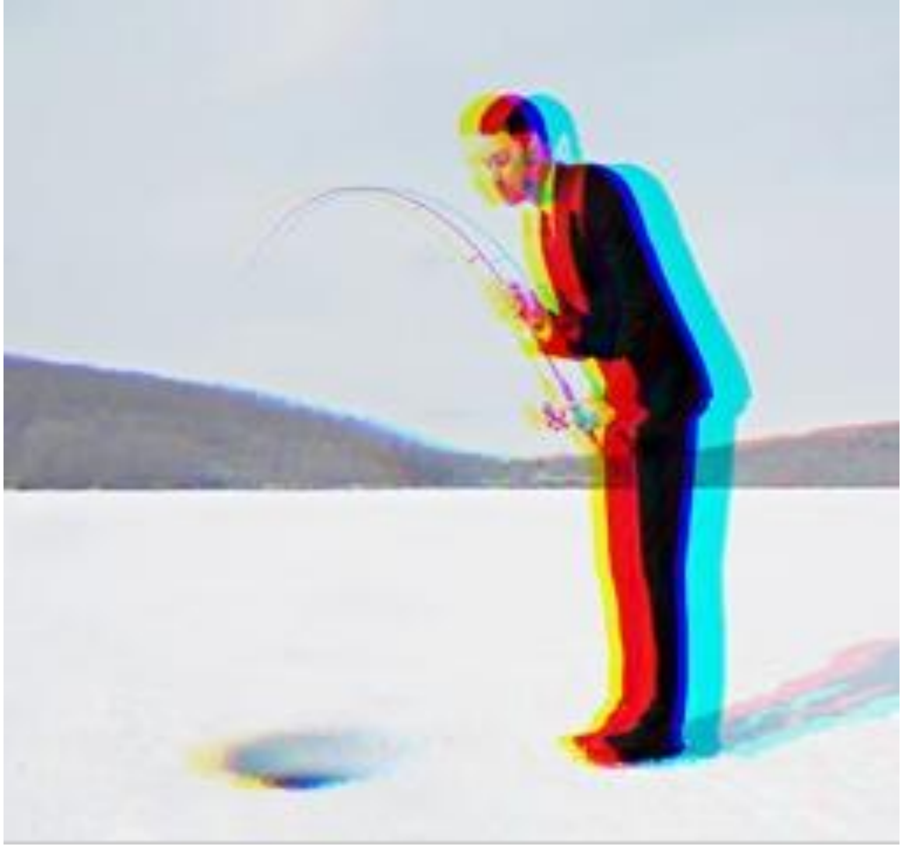


Fig.5.2.(d) 3D output

Thus the algorithm is implemented and the step by step output is displayed. Initially the input image is segmented into groups in order to generate the depth map. Instead of assigning depth value to each pixel, it is assigned to every pixel in the same group to reduce the complexity. Then multiple views are generated to produce a stereoscopic view. The quality of the stereoscopic view mainly depends on the depth map. So the depth map can be further refined to enhance the quality.

CHAPTER 6

CONCLUSION AND FUTURE WORK

This work has presented an automatic 2D-to-3D conversion algorithm. The proposed algorithm utilizes k-means segmentation to group the image into similar pixel. The depth levels are assigned with the initial depth hypothesis. A simple depth hypothesis is adopted to assign the depth for each region and a bilateral filter is subsequently applied to remove the blocky artifacts. The proposed algorithm is quality-scalable depending on the block size. Smaller block size will result in better depth detail and large block size will have lower computational complexity. Capable of generating a comfortable 3D effect, the proposed algorithm is highly promising for 2D-to-3D conversion in 3D applications. This highly promising algorithm for 3D applications is now able to convert any of the 2D images into quality-scalable 3D image. This algorithm can also be further applicable to the videos to generate a multi-view video as the display bringing out the good quality in the output that has been obtained. Thus it is highly satisfying for various emerging 3D applications for the image/visual processing field.

REFERENCES

[1] Y-L. Chang "Depth Map Generation For 2D-To-3D Conversion By Short-Term Motion Assisted Color Segmentation" in Proceedings of ICME, 2007

[2] Chao-Chung Cheng, Chung-Te Li and Liang-Gee Chen, "A 2D-To-3D Conversion System Using Edge Information" in Proceedings of IEEE International Conference on Consumer Electronics, 2010

[3] C.-C. Cheng, C.-T. Li, P.-S. Huang, T.-K. Lin, Y.-M. Tsai, and L.-G.Chen, "A blockbased 2D-to-3D conversion system with bilateral filter" in Proceedings of IEEE International Conferenc5e on Consumer Electronics, 2009

[4] G. Economou, V. Pothos and A. Ifantis, "Geodesic distance and MST based image segmentation", Proceedings European Signal Processing Conference, (2004).

[5] W. J. Tam, and L. Zhang, "3D-TV content generation: 2D-to-3D conversion," in Proc. ICME, pp. 1869-1872, 2006

[6] W.-Y. Chen and Y.-L. Chang and S.-F. Lin and L.-F. Ding and L.-G. Chen."Efficient depth image based rendering with edge dependent depth filter and interpolation," in *Proc. ICME,* pp. 1314-1317, 2005

[7] S. H. Lee, D. W. Park, J. P. Jeong and K. I. Moon, "Conversion 2D Image to 3D Based on Squeeze function and Gradient Map", in proceedings of International Journal of Software Engineering and its Applications,2014

[8] S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach", MIT Technical Report (MIT-CSAIL-TR-2006-073), (2006).

[9] Na Eun Yang, Ji Won Lee, and Rae-Hong Park, "Depth Map Generation Using Local Depth Hypothesis For 2D To 3D Conversion", in International Journal of Computer Graphics & Animation (IJCGA) Vol.3, No.1, January 2013

[10] Jaeseung Ko, Manbae Kim, and Changick Kim, "2D to 3D Stereoscopic Conversion: Depth Map Estimation in a 2D Single View Image" in proceedings of SPIE Vol.6696

[11] 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, 2002.

[12] Y. J. Jung, A. Baik, J. Kim, and D. Park, “A novel 2D-to-3D conversion technique based on relative height depth cue,” in SPIE Electronics Imaging, Stereoscopic Displays and Applications XX, 2009.

[13] H. Murata et al. “Conversion of two-dimensional images to three dimensions,” in Proc. SID Digest of Technical Papers, pp. 859-862, 1995.

[14] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images”, Proc. ICCV, (1998) January, pp. 839-846.

[15] D. Kim, D. Min and K. Sohn, “A Stereoscopic video generation method using stereoscopic display characterization and motion Analysis”, IEEE Trans. On Broadcasting, vol. 54, no. 2, (2008), pp. 188-197.

[16] Sung-Yeol Kim, Sang-Beom Lee, and Yo-Sung Ho, ”Three-dimensional natural video system based on layered representation of depth maps,” in IEEE Transactions on Consumer Electronics, 2006

[17] Yi Min Tsai, Yu Lin Chang and Liang Gee Chen, “Block based Vanishing Line and Vanishing Point Detection for 3D Scene Reconstruction” in International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS),2006.

[18] Jung, C. & Jiao, L. C. “Disparity-map-based rendering for mobile 3D TVs”, IEEE *Trans. Consumer Electronics*, Vol. 57, No. 3, 2011.

[19] Lai, S.-H., Fu, C.-W., & Chang, S. “A generalized depth estimation algorithm with a single image”, *IEEE Trans. Pattern Anal. Machine Intell*, Vol. 14, No. 4, 1992.

[20] Shuo, S. & Sim, T., “Defocus map estimation from a single image”, *Pattern Recognition*, Vol. 44, No. 9, 2011

[21] Saxena, A., Chung, S. H., & Ng, A. Y. “Learning depth from single monocular images”, *Advances in Neural Information Processing Systems 18,* Y. Weiss & B. Sch. Ed. Cambridge, MIT Press, 2006.

[22] Han, K. & Hong, K. “Geometric and texture cue based depth-map estimation for 2D to 3D image conversion”, in *Proc. IEEE Int. Conf. Consumer Electronics*, pp651-652, Las Vegas, NV, USA, 2011.

[23] Felzenszwalb, P. F., & Huttenlocher, D. P. “Efficient graph-based image segmentation”, *Int. J. Computer Vision*, Vol. 59, No. 2, 2004.

[24] Lie, W.-N, Chen, C.-Y., & Chen, W.-C, “2D to 3D video conversion with key-frame depth propagation and trilateral filtering,” *Electron. Lett*, Vol. 47, No. 5, 2011.

[25] Gonzalez, R. C. & Woods, R. E., *Digital Image Processing*, *Third edition*, Upper Saddle River, Pearson Education,2010.