2D-TO-3D CONVERSION OF IMAGES USING EDGE INFORMATION

By

S. BHARATHI

Reg No: 1020106003

of

KUMARAGURU COLLEGE OF TECHNOLOGY,
COIMBATORE - 641 049.
(An Autonomous Institution affiliated to Anna University of Technology, Coimbatore)

A PROJECT REPORT

Submitted to the

FACULTY OF ELECTRONICS AND COMMUNICATION ENGINEERING

In partial fulfillment of the requirements for the award of the degree

of

MASTER OF ENGINEERING

IN

APPLIED ELECTRONICS

APRIL 2012

BONAFIDE CERTIFICATE

Certified that this project report titled "**2D-TO-3D CONVERSION OF IMAGES USING EDGE INFORMATION"** is the bonafide work of **Ms. S.BHARATHI (Reg. No. : 1020106003)** who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

(Dr. A. VASUKI) **(Dr. RAJESWARI MARIAPPAN)**

PROJECT GUIDE **HEAD OF THE DEPARTMENT**

The candidate with **Register No. 1020106003** was examined by us in Project Viva-Voce examination held on _ _ _ _ _ _

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

I would like to express my thanks and appreciation to the many people who have contributed to the successful completion of this project. Firstly I express my profound gratitude to our Chairman **Padmabhusan Arutselvar Dr. N. Mahalingam B.Sc., F.I.E.,** and Co-Chairman **Dr. B. K. Krishnaraj Vanavarayar, B.Com., B.L.,** for giving this opportunity to pursue this course.

I would like to thank **Dr. J. Shanmugam,** Director, for providing me an opportunity to carry out this project work.

I would like to thank **Dr. S. Ramachandran** Principal, for providing me an opportunity to carry out this project work.

I would like to thank **Dr. Rajeswari Mariappan** Head of the Department, Electronics and Communication Engineering, who gave her continual support for me throughout the course of this project.

My heartfelt thanks to my internal guide **Dr.A. Vasuki, Prof.,** Supervisor, for her technical guidance, constructive criticism and many valuable suggestions provided throughout the project work.

I would like to thank M**rs.R.Hemalatha M.E., Asst. Professor**, and Project Coordinator, for her contribution and innovative ideas at various stages of the project and for her help to successful completion of this project work.

I express my sincere gratitude to my family members, friends and to all my staff members of Electronics and Communication Engineering department for their support throughout the course of my project.

ABSTRACT

The three-dimensional (3D) video signal processing has received considerable attention in visual processing. Given advances in 3D display technology, humans aspire to experience more realistic and unique 3D effects. Although 3D displays enhance visual quality more than two-dimensional (2D) displays do, the depth information required for 3D displays is unavailable in the conventional 2D content. Therefore, converting 2D images into 3D ones has become an important issue in emerging 3D applications. Several approaches capable of generating 3D content require specific devices and are only effective in generating new content, making them infeasible for conversion of 2D contents. This necessitates the development of an efficient 2D-to-3D conversion system.

This work presents a novel algorithm that automatically converts 2D images into 3D ones. The proposed algorithm first segments the image into object groups by choosing an effective grouping method. This is done with the help of the edge information. The grouping is based on the pixels having similar colors and spatial locality. A depth map is then assigned based on a hypothesized depth gradient model. Next, the depth map is then assigned by cooperating with a cross bilateral filter. This is very much essential to generate visually comfortable depth maps efficiently and also to diminish the blocky artifacts. Then the filtered image is processed using a Depth Image Based Rendering (DIBR) method and 3D image is readily generated by fusion of images. Image registration is another algorithm applied for the fusion process of images. This algorithm is applied for the left and right views and the 3D output image is obtained with the help of the control points selected from the images. Thus the obtained 3D images from DIBR and registration process are compared and the results are analyzed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2D	---	Two- Dimensional
3D	---	Three- Dimensional
DIBR	---	Depth Image-Based Rendering
MST	---	Minimum Spanning Tree
CID	---	Computed Image Depth
FFT	---	Fast Fourier Transform
3D TV	---	Three-Dimensional Tele-Vision
MRI	---	Magnetic Resonance Imaging
SPECT	---	Single Photon Emission Computed Tomography
TFORM	---	TransFORMation
RGB	---	Red-Green-Blue

CHAPTER 1
INTRODUCTION

1.1 OVERVIEW OF THE PROJECT

With the development of 3D applications, the conversion of existing 2D images to 3D images becomes an important component of 3D content production. The dominant technique for such content conversion is to develop a depth map for each frame of 2D material. The use of a depth map as part of the 2D to 3D conversion process has a number of desirable characteristics:

1. The resolution of the depth map may be lower than that of the associated 2D image;
2. It can be highly compressed;
3. 2D compatibility is maintained; and
4. Real time generation of stereo, or multiple stereo pairs, is possible.

When observing the world, the human brain integrates various heuristic depth cues to generate the depth perception. The major depth perceptions are binocular depth cues from two eyes and monocular depth cues from a single eye. The disparity of binocular visual system helps human eyes to converge and accommodate the object at the right distance. Therefore, humans can also perceive depth from the single-view image/video.

One of the key steps in 2D to 3D conversion is how to generate a dense depth map. In recent years, a number of depth map generation algorithms have been proposed. Each algorithm has its own strengths and weakness. 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.

2D-to-3D depth generation algorithms generally face two challenges. One is the depth uniformity inside the same 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. 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.

To overcome these two challenges, this work presents a novel algorithm that uses a simple depth hypothesis to assign the depth of each group rather than retrieving the depth value directly from the depth cue. Firstly, the proposed algorithm chooses an effective grouping method in which grouping pixels have similar colors and spatial locality. Then the depth values are assigned according to the hypothesis depth value. Cross bilateral filter is applied to enhance the visual comfort. Experimental results indicate that the proposed algorithm can generate promising results with slight side effects.

1.2 SOFTWARE USED

- MATLAB R2009a.

1.3 ORGANIZATION OF THE REPORT

- **Chapter 2** discusses about the proposed conversion system and the algorithm used for processing the image through various stages. It shows the mathematical expressions used in each of the processing stage and explains each of the process used in brief with the appropriate figures wherever necessary.
- **Chapter 3** gives the detailed view of the image registration algorithm and its performance in generating the required output by means of the registration process of two images of same scene and exhibits the results obtained.
- **Chapter 4** shows the experimental results and analysis over the comparison of two algorithms and its results over their performance. It also discusses about the various comparison parameters and their results.
- **Chapter 5** gives the conclusion of the project comparing the results obtained and infers about the extension of the project for the future work.

CHAPTER 2
PROPOSED CONVERSION SYSTEM

There are several approaches capable of generating 3D content which include active depth sensor, triangular stereo vision and 3D graphics rendering. 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 which are infeasible for the 2D contents. So this is a time consuming method of manually editing the depth map for the content. Hence to make this generation efficient, this work proposes a novel 2D-to-3D conversion method based on use of the edge information. This conversion automatically generates the depth information from a single view image to the multi-view image using the estimated depth maps.

This 2D-to-3D depth generation algorithms usually face two challenges: One is the depth uniformity inside the same object. As the image consists of 2D pixel arrays, the information regarding the object grouping of those available pixels is lacking much. This implies that a better grouping of pixels results in a better depth uniformity inside the object. So here an effective grouping method called Block-based region grouping has been chosen. This method considers both the color similarity and also the spatial distance.

Whereas the other challenge involves retrieving the appropriate depth relationship among all the objects. Some methods integrate the depth information with the object grouping concept such as using motion parallax as the primary depth cue. This may lead to false depth information when the object is with different self motion vectors. Thus pixels of same object may be assigned with different depth values. In order to overcome these challenges, an automatic conversion system is proposed here. As a result of this conversion, the input 2D image is converted into visually comfortable 3D image without the presence of any artifacts enhancing the quality of the image in the display.

2.1. ALGORITHM

This 2D-to-3D conversion process can be explained with the following steps to generate a visually comfortable multi-view 3D output maintaining the pixel information.

Step 1: Initially the input image is segmented into blocks. Following this segmentation, the blocks are then segmented into groups using the edge information. With the help of an effective grouping method, the pixels having similar colors and spatial locality are grouped.

.

Step 2: 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.

Step 3: The corresponding depth for each segment is then assigned by using the hypothesized depth gradient. This is derived based on the linear perspective information. Because of the use of the depth gradient and texture followed by a depth hypothesis, the depth map obtained has a good quality resembling the input image.

Step 4: Thus the generated depth map now contains the blocky artifacts. So, those blocky artifacts are then removed using the cross bilateral filtering. This cross bilateral filter particularly removes most of the noise available in the depth map when compared with other filters.

Step 5: To enhance a comfortable visual quality of the smoothened image, the depth image-Based rendering (DIBR) method is applied. Thus a 3D image is displayed from the 2D image.

2.2. BLOCK DIAGRAM

The conversion system proposed here is illustrated in figure 1. Here we consider the input 2D image and partition it into various blocks and group them with the help of edge information. A depth map is generated using a depth hypothesis gradient. Some blocky artifacts are present in the generated depth map and so a cross bilateral filter is used here to remove the noise present in the depth map. Finally, the algorithm chosen is applied to generate a multi-view 3D image.

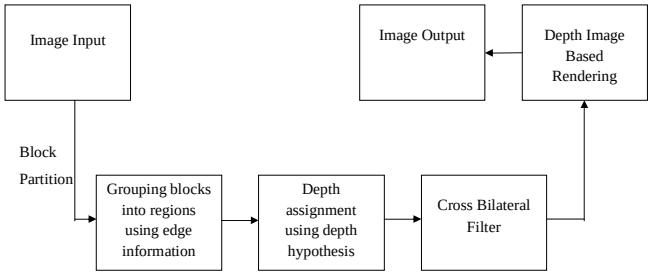


Figure 2.1 2D-TO-3D conversion system

2.3. BLOCK BASED REGION GROUPING

Computational complexity is reduced using a block-based algorithm, implying that each pixel in the same block has the same depth value. A 4-by-4 graph is used as an example. Each node is a 4-by-4 pixel block, and each node is four-connected. The value of each link is assigned as the absolute difference of the mean of neighboring blocks:

$$\text{Diff}(a, b) = |\text{Mean}(a) - \text{Mean}(b)| \quad (2.1)$$

where a and b denote two neighboring blocks, respectively, and $Mean(a)$ represents the mean color of a. A smaller value implies a higher similarity between the two blocks. Following calculation of the absolute difference of the mean of the neighboring blocks, the blocks are segmented into multiple groups by using the minimum spanning tree segmentation.

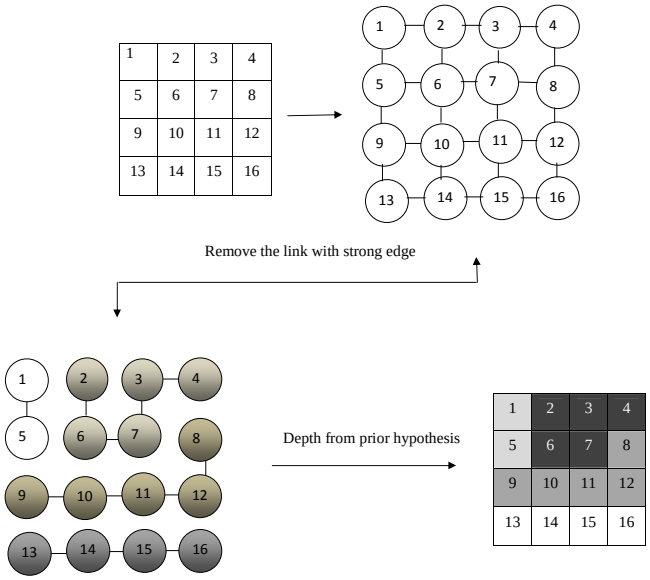


Figure 2.2 Flow of block-based region grouping

The flow of block-based region grouping is shown in the figure 2.2. Initially, a minimum spanning tree (MST) is constructed. The links of stronger edges are removed to generate multiple grouped regions. Notably, MST algorithm is used to identify the coherence among blocks with both the color difference and the connectivity without generating many small groups. MST algorithm, which preserves the link connectivity, has an excellent result in spatial locality.

The MST preserves the connectivity of the image graph and provides a link to all nodes (regions) at a minimum total edge cost. By deleting MST edges with the largest values, isolated clusters that correspond to segmented regions, are formed. Recursive MST versions, incorporating some form of global information by recalculating links and costs, will result in a much increased computational complexity. The pseudo code of segmentation using the minimum spanning tree is given below:

Table 2.1 Pseudo code of MST algorithm

```
Minimum Spanning Tree Segmentation( )
        initial: T = { }
        while T does not form a spanning tree:
                find the smallest edge in E that form a spanning tree for T
                T= T union {E}
        sorting the edges in T by the edge strength
```

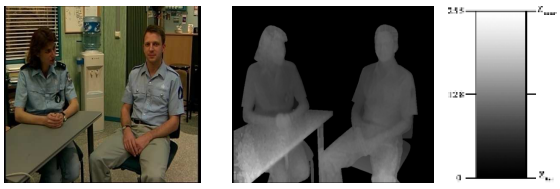
Thus with the above pseudo code, the strongest edges are removed and only the weakest edges are considered. The MST is constructed using well known greedy algorithms like Dijkstra's algorithm, Prim's algorithm and Kruskal's algorithm and the upper bound of its computational complexity is $O(N^2 \log N)$, N the number of graph nodes. Although among graph based segmentation algorithms, the computational efficiency of the MST is recognized, there are several problems associated with it. These are related to the fact that the MST does not include

global information in the merging operations, which are decided locally by weight values. Nevertheless results are in some cases quite satisfactory.

A situation where MST fails is when a series of regions, which differ slightly from each other, bridge a large gap in the feature space. Since the edge weights are small, the graph will not be cut and very different objects in terms of feature space characteristics will be merged. Recursive implementations of the MST algorithm, recalculating edge costs, can tackle the problem increasing complexity. With the efficient linkage preserving property, the minimum spanning tree segmentation method can generate excellent grouping results. The proposed depth generation algorithm can also be substituted by other automatic or manual segmentation with satisfactory grouping results.

2.4. DEPTH FROM PRIOR HYPOTHESIS

Depth Map or Depth Image: 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.



(a) (b)

Figure 2.3 Figure showing (a) 2D image and (b) Depth Image (map)

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. Each signal represents different depth information.

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:

1) A computed image depth (CID) method to find out the depth relations from the part of an image.

2) A feasible depth from focus algorithm to obtain the depth map from multiple different focus distance images.

3) Another still image analysis method which utilizes the machine learning algorithm to find the feature points for depth mapping.

4) Generation of a depth map, which bases on several steps including the generation of gradient planes, depth gradient assignment, consistency verification of the detected region, and finally the depth map generation.

5) A vanishing line detection method which uses block-based algorithm to extract the vanishing lines and point.

6) A motion detector and grouping method to assign different depth to each group.

7) Use the motion segmentation, motion estimation, and object tracking algorithms to find out the object and its possible depth map.

8) A visual modeling method by the motion parallax cue.

9) Some reviews of the motion parallax based 3D reconstruction methods.

10) A simple line tracing method.

Depth map generation methods can also be classified mainly into single-frame and multi-frame methods. The above mentioned methods belong to the category of single-frame methods. Although capable of exploiting various monotonic depth cues and appropriate for different cases, these single-frame methods are unreliable when the selected cues are weak in the input image. Moreover, the above single-frame methods process each frame individually without considering the temporal coherence and may produce flickering in the depth maps.

In this conversion process, following the generation of the block groups, the corresponding depth for each block is assigned by the hypothesized depth gradient. The process includes the generation of gradient planes, depth gradient assignment, consistency verification of the detected region, and finally the depth map generation. When each scene change is detected, the linear perspective of the scene is analyzed by a line detection algorithm using Hough transform.

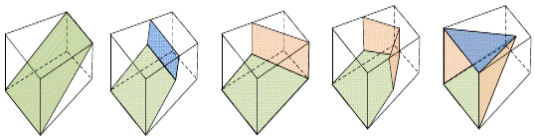


Figure 2.4 Multiple scene modes using depth from geometry

The initial depth gradient hypothesis is derived based on the linear perspective information. The above figure 2.4 shows multiple scene modes using the depth from geometry. Thus the depth gradient and texture are executed here and a depth hypothesis is assigned to obtain the required depth map which has a good quality inferring all the information from that of the input image. Thus compared with the other depth map generation methods, this method of using a depth hypothesis gives a better depth map.

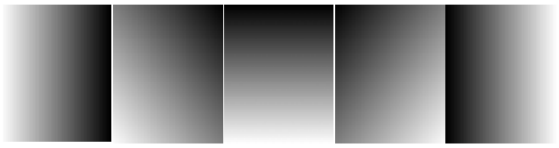
The depth value of a given block group *R* is assigned by:

$$\text{Depth}(R)=128+255\left\{\sum_{pixel(x,y)\in R} W_{rl}\frac{x-\frac{width}{2}}{width}+W_{ud}\frac{y-\frac{height}{2}}{height}\right\}/\text{pixel\_num}(R) \qquad (2.2)$$

where $|Wrl| + |Wud| = 1$.

A larger value of the assigned depth implies a closer pixel to the user. The above equation suggests that the assigned depth value is the gravity center of the block group, explaining why each block group belongs to the same depth. The absolute value and sign of *Wrl*

and *Wud* can adjust the left-to-right and top-to-bottom depth gradient weight. The following figure shows the different depth map gradients:



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

Figure 2.5 Five major depth map gradients: (a) left; (b) left-down; (c) bottom-up; (d) right-down; (e) right

The exact depth map gradients have to be found from different categories as shown in the figure 2.5. 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.

2.5. 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 (e.g. distance between hair orientations).

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.

Among the variants of the bilateral filter, this conversion method chooses the cross bilateral filtering. In computational photography applications, it is often useful to decouple the data to be smoothed defining the edges to be preserved. 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:

$$Depth_f(x_i) = \frac{1}{N(x_i)} \sum_{x_j \in \Omega(x_i)} e^{-0.5\left[\frac{|x_j - x_i|}{\sigma_s^2} + \frac{|u(x_j) - u(x_i)|^2}{\sigma_r^2}\right]} Depth(x_j) \quad (2.3)$$

$$N(x_i) = \sum_{x_j \in \Omega(x_i)} e^{-0.5\left[\sigma_s^2 |x_j - x_i| + \sigma_r^2 |u(x_j) - u(x_i)|^2\right]} \quad (2.4)$$

where $u(xi)$ denotes the intensity value of the pixel xi, $\Omega(xi)$ represents the neighboring pixels of xi, $N(xi)$ refers normalization factor of the filter coefficients and $Depth_f$ 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.

2.6. 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.

For hole-filling, average filter is commonly used. However, the average filter does not preserve edge information of the interpolated area. Therefore, using average filter results in obvious artifacts at highly-textured area. This artifact is well known as rubber and sheet artifact. Another novel method uses Gaussian filter to smooth the whole depth image before 3D image warping. The method claims that the depth map after smoothing is blurred, there are fewer hole to be filled. However, if we smooth the whole depth map before warping, the computation time and the warped image quality is bad.

Hence with an efficient edge dependent depth filter, this depth image-based rendering algorithm is processed. Hole-location and hole-size are detected before 3D image warping. And the edge dependent Gaussian filter is used to reduce the hole-size efficiently. Then we use edge dependent interpolation to fill the small holes, thus the edge information of synthesized image is preserved. So the expected output is free from noise which has be smoothened for the further process of generation of multi-view output for different types of images by considering the generated depth map during the processing of the image.

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.

Depth-Image-Based-Rendering for advanced 3D TV System can be illustrated by the following block diagram. This system includes three parts, pre-processing of depth map, 3D

image Warping and Hole-Filling. Smoothing filter is first applied to smooth the depth image. Then 3D image warping generates the left and right view according to the smoothed depth map and intermediate view. If there are still holes, hole-filling is applied to fill color into these holes.

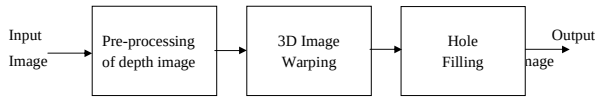


Figure 2.6 Depth Image Based Rendering block diagram

2.6.1 Pre-Processing of Depth Image

This is the first step in this depth image based rendering algorithm while processing the image for generating multi-view output image which is visually comfortable. Pre-Processing of depth image is usually a smoothing filter. Because depth image with horizontal sharp transition would result in big holes after warping, smoothing filter is applied to smooth sharp transition so as to reduce the number of big hole.

However, if we blur the whole depth image, we will not only reduce big holes but also degrade the warped view. This is because the depth map of non-hole area is smoothed. Hence the resulted smoothing image has to be passed to the next stage to generate different views of the image with the help of the depth map generated from the depth hypothesis.

2.6.2 3D Image Warping

3D image warping maps intermediate view pixel by pixel to left or right view according to pixel depth value. In the other words, 3D image warping transforms pixel location according to depth value. The 3D image warping formula is as following:

$$x_r = x_c + \left(\frac{t_x}{2} \frac{f}{z}\right) \quad (2.5)$$

$$x_r = x_c - \left(\frac{t_x}{2} \frac{f}{z}\right) \quad (2.6)$$

The x_l is the horizontal coordinate of the left view, and x_r is the horizontal coordinate of the right view. Besides, x_c is the horizontal coordinate of the intermediate view. Z is depth value of current pixel, f is camera focal length and t_x is eye distance. The formula shows that 3D warping maps pixel of intermediate view to left and right view in horizontal direction.

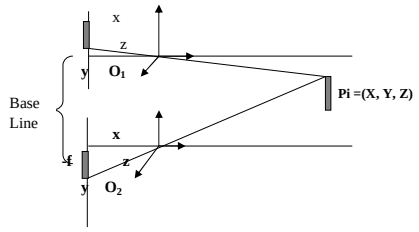
2.6.3 Hole Filling

Average filter interpolation method is a common method for Hole-Filling in DIBR. However, using average filter only would result in artifacts at highly-textured areas. Besides, hole-size in DIBR is so huge that it is needed to using average filter with large window size. At the same time, average filter with large window size cannot preserve edge information for the reason that edge information is blurred.

The holes that are formed because of the warping process are filled by this process by giving the colors to the respected regions and generating the left and right view images with the help of the depth map obtained previously. These three process are essential for the effective generation of the final output image of this depth image based rendering algorithm.

The disparity can be calculated based on the known depth information as shown in figure 2.4. 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.

Thus this depth image-based rendering algorithm is used to generate multiple views of a single 2D image by preserving the edge and pixel information. The required fusion process is carried over to generate the stereoscopic view of the input image to feel the realistic scenes.



$$X_1 = -f\frac{X}{Z}, X_2 = -f\frac{X-B}{Z} = X_1 - f\frac{B}{Z} \quad (2.7)$$

$$\Rightarrow Z = \frac{fB}{X_1 - X_2} = \frac{fB}{D} \Rightarrow Disparity\ D = \frac{fB}{Z} \quad (2.8)$$

Figure 2.7 Equations of depth and disparity mapping for depth image based rendering

Thus by the means of warping process, the left view and the right view of the image are obtained. By the smoothing process, the numbers of big holes are reduced and the sharp transitions are smoothened by the help of the smoothing filter. Also the non-hole areas are also smoothened. The depth image is warped by means of pixels of intermediate view to generate the multi-view of the image.

CHAPTER 3

IMAGE REGISTRATION

Image registration is the process of aligning two or more images of the same scene. Typically, one image, called the *base* image or *reference* image, is considered the reference to which the other images, called *input* images, are compared. The object of image registration is to bring the input image into alignment with the base image by applying a spatial transformation to the input image. The differences between the input image and the output image might have occurred as a result of terrain relief and other changes in perspective when imaging the same scene from different viewpoints. Lens and other internal sensor distortions, or differences between sensors and sensor types, can also cause distortion.

A spatial transformation maps locations in one image to new locations in another image. Determining the parameters of the spatial transformation needed to bring the images into alignment is the key to the image registration process. The Image Processing Toolbox software provides tools to support point mapping to determine the parameters of the transformation required to bring an image into alignment with another image. In point mapping, you pick points in a pair of images that identify the same feature or landmark in the images. Then, a spatial mapping is inferred from the positions of these control points.

Image registration is often used as a preliminary step in other image processing applications. For example, you can use image registration to align satellite images of the earth's surface or images created by different medical diagnostic modalities (MRI and SPECT). After registration, you can compare features in the images to see how a river has migrated, how an area is flooded, or to see if a tumor is visible in an MRI or SPECT image. Thus this registration process is an existing algorithm for registering any input image over a base image overlaying all the required parameters in the input image with that of the base image.

3.1 ALGORITHM

The image registration process overlays the input image over that of the base or the reference image to bring the parameters to match to the maximum extent. By means of specific type if transformation, this process aligns two or more images of the same scene. The process registers the image through various steps given as follows:

Step 1: Initially the two images named the base image or reference image and the input image are to be read. Those images need not be read in the workspace. They may in any alignment but of same scenes.

Step 2: Next the control points should be chosen from both the images. An interactive tool is helpful in selecting the pairs of points in the images. These control points should be chosen as if they are the corresponding pairs in both the images.

Step 3: Then we have to save the control point pairs to the MATLAB workspace. This is essential to register the base image and the input image. These points should be atleast in the required number to proceed to the next steps.

Step 4: Now there is an optional step that uses cross-correlation to adjust the position of the control points that are selected with cpselect function. If the control points are locked, then only the matched points will be considered.

Step 5: Then the type of transformation has to be specified and its parameters have to be inferred. Different types of transformations are applicable based on the selection of regions of the image.

Step 6: Finally, in this image registration process, transform the input image to bring it into alignment with the base image. Based on the spatial transformation, the number of control points can be increased to get more clear and to the maximum an appropriate output image of the registration process.

3.2 GRAPHIC ILLUSTRATION

The following figure provides the graphical flow of the image registration process:

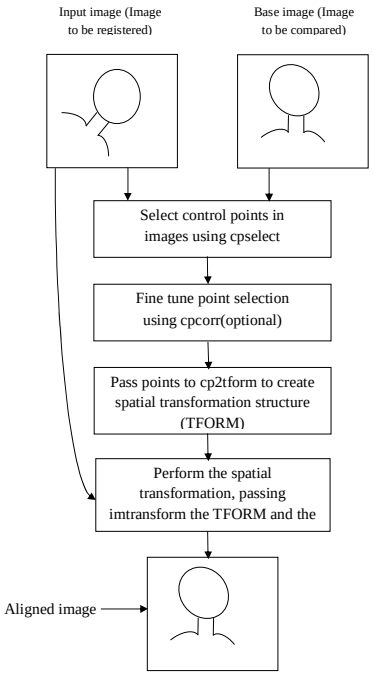


Figure 3.1 Image Registration Process Flow

This graphical flow illustrates the various stages of the image registration process. Consider the two images named the base image or the reference image and the input image to be registered over that of the base image. By means of the control point selection tool, pairs of corresponding control points are picked up from both the images which may be the images of same scene. With the help of these control points, the correlation process can be implemented for making the control points to match approximately to the much extent.

The cpselect function is used to select the control points from both the images. The fine tuning of the control points can be made by using the function named cpcorr. This function mostly locks the control points that were selected in the corresponding manner in the images. The spatial transformation structure has to be created by choosing the appropriate transformation among different categories. This is achieved by using the cp2tform function. Then finally, imtransform function is used to register the input image over the base image or the reference image.

3.3 READ THE IMAGES

In this step, two images are to be read. One is the base of the reference image and the other is the input image. Ortho photos can also be considered for the registration process. In the orthophoto, each pixel center corresponds to a definite geographic location, and every pixel is 1 meter square in map units. A visible-color RGB image can also be taken. Even the aerial image is geometrically uncorrected: it includes camera perspective, terrain and building relief, internal (lens) distortions, and it does not have any particular alignment or registration with respect to the earth.

Thus any form of images can be read but should belong to the same scene. The images can belong to different types like aerial, ortho photo, RGB image and so on. The main difference between these types of images is the various kinds of parameters available in the images being taken for the registration process. Thus once the images are read, then they are ready for the control points selection.

3.4 CHOOSE CONTROL POINTS IN THE IMAGES

The toolbox provides an interactive tool, called the Control Point Selection Tool that you can use to pick pairs of corresponding control points in both images. Control points are landmarks that you can find in both images, like a road intersection, or a natural feature. To start this tool, enter cpselect at the MATLAB prompt, specifying as arguments the input and base images.

cpselect(input image, base image)

The Control Point Selection Tool displays two views of both the input image and the base image in which you can pick control points by pointing and clicking. This figure shows the Control Point Selection Tool with four pairs of control points selected. The number of control point pairs you pick is at least partially determined by the type of transformation you want to perform. The type of the transformation gives information about the minimum number of points required by each transformation.

3.5 SAVE THE CONTROL POINTS TO THE WORKSPACE

In the Control Point Selection Tool, click the **File** menu and choose the **Export Points to Workspace** option. For example, the following set of control points in the input image represent spatial coordinates; the left column lists *x*-coordinates, the right column lists *y*-coordinates.

```
input_points =   118.0000   96.0000
                 304.0000   87.0000
                 358.0000  281.0000
                 127.0000  292.0000
```

This is essential for the images to register as the control points play a main role in registering the image one over another. Thus this forms an essential step for the further process to be continued so as to match the corresponding pairs of the control points in the images to be registered.

3.6 FINE-TUNE THE CONTROL POINT PAIR PLACEMENT

This is an optional step that uses cross-correlation to adjust the position of the control points you selected with cpselect. To use cross-correlation, features in the two images must be at the same scale and have the same orientation. They cannot be rotated relative to each other. Because the Concord image is rotated in relation to the base image, cpcorr cannot tune the control points. This is an optional step to be included in the process of the image registration for locking the selected pairs of control points that have been selected.

3.7 TRANSFORMATION AND ITS TYPES

In this step, you pass the control points to the cp2tform function that determines the parameters of the transformation needed to bring the image into alignment. cp2tform is a data-fitting function that determines the transformation based on the geometric relationship of the control points. cp2tform returns the parameters in a geometric transformation structure, called a TFORM structure.

When you use cp2tform, you must specify the type of transformation you want to perform. The cp2tform function can infer the parameters for five types of transformations. You must choose which transformation will correct the type of distortion present in the input image. Images can contain more than one type of distortion.

There are different types of transformation to be mentioned and are as follows: The cp2tform function can infer the parameters for the following types of transformations, listed in order of complexity.

- 'nonreflective similarity'
- 'affine'
- 'projective'
- 'polynomial' (Order 2, 3, or 4)
- 'piecewise linear'
- 'lwm'

The first four transformations, 'nonreflective similarity', 'affine', 'projective', and 'polynomial' are global transformations. In these transformations, a single mathematical expression applies to an entire image. The last two transformations, 'piecewise linear' and 'lwm' (local weighted mean), are local transformations. In these transformations, different mathematical expressions apply to different regions within an image.

When exploring how different transformations affect the images you are working with, try the global transformations first. If these transformations are not satisfactory, try the local transformations: the piecewise linear transformation first, and then the local weighted mean transformation.

The choice of transformation type affects the number of control point pairs you need to select. For example, a nonreflective similarity transformation requires at least two control point pairs. A polynomial transformation of order 4 requires 15 control point pairs. The predominant distortion in the aerial image results from the camera perspective. Under some cases, image registration can correct for camera perspective distortion by using a projective transformation. The projective transformation also rotates the image into alignment with the map coordinate system underlying the base digital orthophoto image.

3.8 TRANSFORM THE IMAGE

As the final step in image registration, transform the input image to bring it into alignment with the base image. You use `imtransform` to perform the transformation, passing it the input image and the `TFORM` structure, which defines the transformation. `imtransform` returns the transformed image. Thus this overlaying process registers the base image or the reference image over that of the input image to give a final view of the combined image.

CHAPTER 4

EXPERIMENTAL RESULTS AND ANALYSIS

The proposed method generates good and comfortable visual quality images since it is not restricted in horizontal line tracing. The proposed 2D grouping method has better results than the other single view based algorithm especially in vertically standing objects. In contrast with the conventional motion-based algorithms in generating depth from multiple frames, this conversion method uses only a single image with only slight side effects.

In finding the complexity of algorithm, a larger number of blocks in the frame imply a longer computational time since the MST segmentation algorithm has a highly sequential dependency. Thus, smaller the block size, shorter the computation time. However, a larger block size implies a lower depth map quality .This work also evaluated the visual quality of the proposed algorithm by comparing the four images. The block sizes are considered small for enabling the less computation complexity by adopting the convenient block-based algorithm.

Figure 4.1 shows some experimental analysis, including data on four sets of the original images, depth maps and red-cyan images. Initially, the input image 'image1' gives a clear depth map regarding the nearest and the farthest objects. Thus this depth map enables greatly to give a clear and finely visualizing multi-view 3D display. The results are displayed in the first set of images. Considering the next image 'image2', the depth map is slightly blurred because of the distant objects in the image. So the display of the output image giving two different views is also little blurred as if reproduced from that of the depth map of the image.

Similarly, the next image, 'church' is also giving the clear view of the nearest and the farthest objects in its depth map and so the 3D display is clearly visualized when compared with that of the input image. Finally the image 'house' gives a clear depth map as if the initial image considered. So the output image is comfortably displayed preserving the edge information with the left and right view displaying additional information.

4.1 EXPERIMENTAL RESULTS OF DIBR ALGORITHM

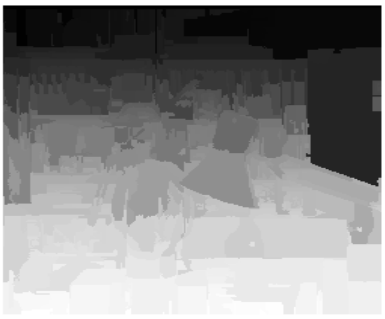
The experimental results were simulated and are analyzed step by step for the following four sets of images: 'image1.bmp';'image2.bmp';'church.bmp';'house.bmp'. Different stages of output are displayed here for these images for analysis.

'image1.bmp'

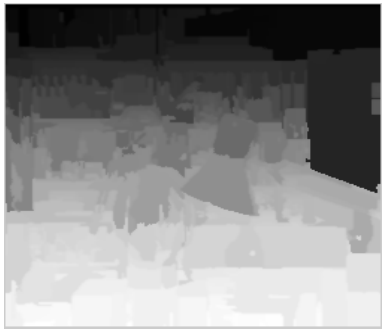
(a)

(b)

Figure 4.1 (a) 2D input (original) Image (b) Edge mapped image



(c)

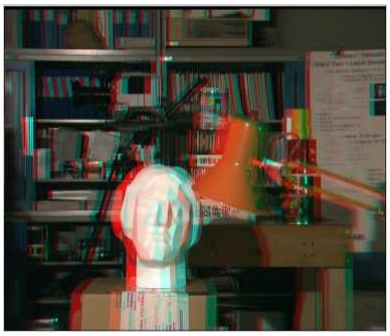


(d)

Figure 4.2 (c) Depth map assigned image (d) Bilateral Filtered image

(e)

(f)

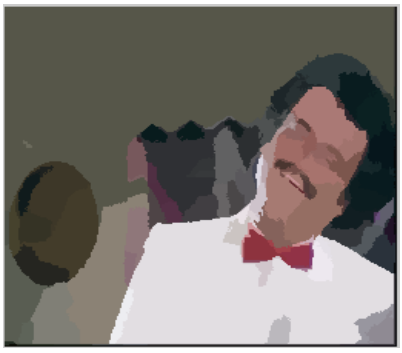


(g)

Figure 4.3 Multi-view display (e) Left-view image (f) Right-view image; (g) 3D output image

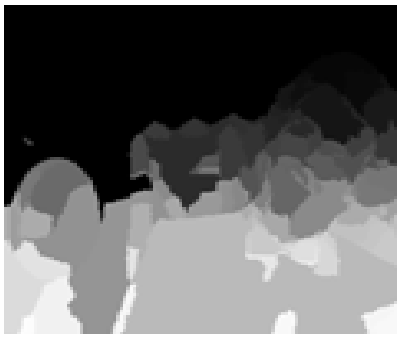
‘image2.bmp’

(a)

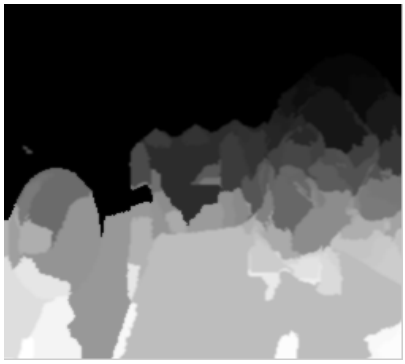


(b)

Figure 4.4 (a) 2D input (original) Image (b) Edge mapped image



(c)



(d)

Figure 4.5 (c) Depth map assigned image (d) Bilateral Filtered image

(e) (f)

(g)

Figure 4.6 Multi-view display (e) Left-view image (f) Right-view image; (g) 3D output image

‘church.bmp’

(a)

(b)

Figure 4.7 (a) 2D input (original) Image (b) Edge mapped image

(c)

(d)

Figure 4.8 (c) Depth map assigned image (d) Bilateral Filtered image

(e)

(f)

(g)

Figure 4.9 Multi-view display (e) Left-view image (f) Right-view image; (g) 3D output image

'house.bmp'

(a)

(b)

Figure 4.10 (a) 2D input (original) Image (b) Edge mapped image

(c)

(d)

Figure 4.11 (c) Depth map assigned image (d) Bilateral Filtered image

(e) **(f)**

(g)

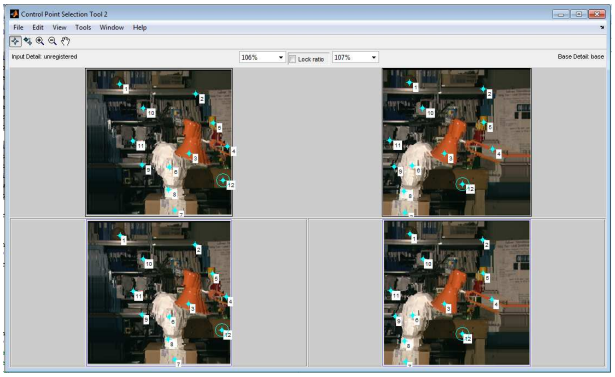
Figure 4.12 Multi-view display (e) Left-view image (f) Right-view image; (g) 3D output image

4.2 EXPERIMENTAL RESULTS OF IMAGE REGISTRATION

The process of image registration for fusing the tow images of the same scene has been experimentally carried over and the results are displayed for the furhter analysis. Given the results for four sets of images: 'image1.bmp';'image2.bmp';'church.bmp';'house.bmp'.

'image1.bmp'

(a) **(b)**



(c)

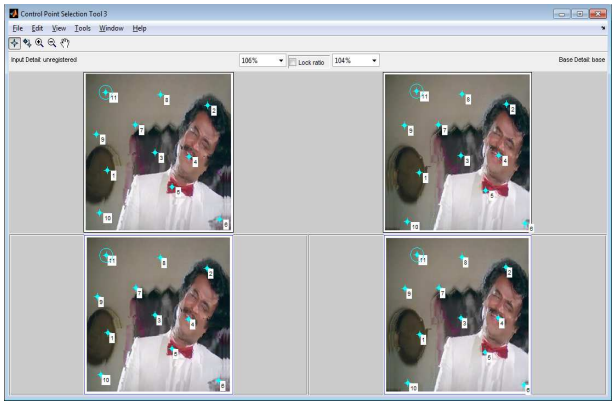
Figure 4.13 (a) Left-view Image (b) Right-view Image (c) Control Point Selection

(d)

'image2.bmp'

(a) **(b)**

Figure 4.14 (d) 3D output (a) Left-view Image (b) Right-view Image



(c)

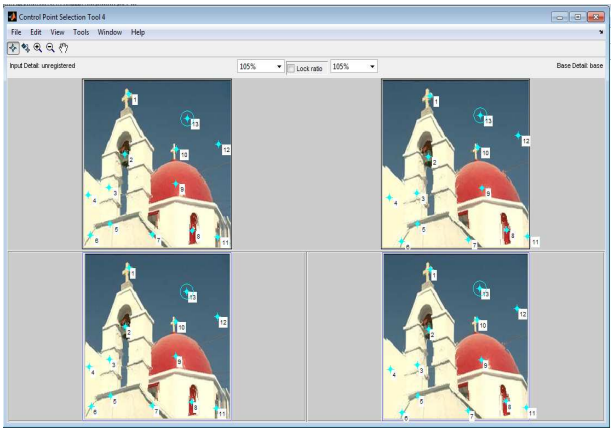
(d)

Figure 4.15 (c) Control Point Selection (d) 3D output

'church.bmp'

(a)

(b)



(c)

Figure 4.16 (a) Left-view Image (b) Right-view Image (c) Control Point Selection

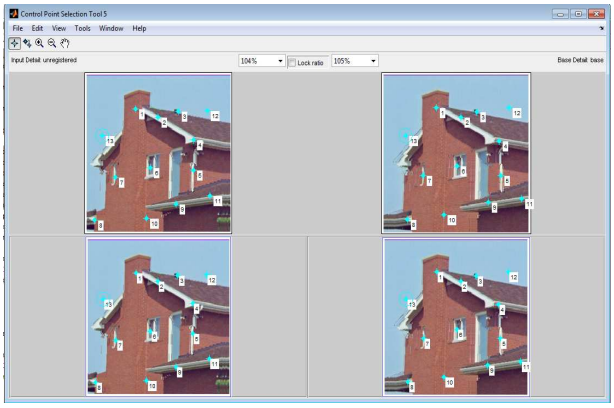
(d)

'house.bmp'

(a)

(b)

Figure 4.17 (d) 3D output (a) Left-view Image (b) Right-view Image



(c)

(d)

Figure 4.18 (c) Control Point Selection (d) 3D output

Comparing the four output images, the output image is having a good depth based on the depth information obtained from the input image. The more clear the depth map, higher the quality of the output image. The results are clearly shown in the four set of images in the experimental results. For example, in 'image1', the depth of the input image is obtained clearly. And so the output image is also having a clear depth.

The depth maps of all the images are giving more importance to the foreground information compared to that of the background ones. The results are clearly shown in the four set of images in the experimental results. The results also indicate that the images violating the initial hypothesis still generate an acceptable quality of image. Thus the proposed method is assuring of high quality and comfortable visual images giving a 3D display from those data available in the input 2D images.

Since the block-based region grouping is considered to be the good grouping method by considering the blocks, the edge mapped image obtained from the input image is so approximate that it resembles so close to that of the input image. The particular depth gradient named the bottom-up depth gradient that removes the most of the noise generated in the depth map has been chosen as one of the parameters.

Then a depth hypothesis is assigned by means of a mathematical formula which considers the width and height of the block. And so the depth map obtained is having a better quality when compared with the other depth map generation methods. The blocky artifacts have to be removed from that of the depth map obtained. The cross bilateral filter has been chosen for the filtering process which smoothens the images making it a clear depth map for the next processing stage.

Finally, the depth image based rendering method is applied over the depth map and the input view to obtain the different views of the image like the left image and the right image giving a multi-view display. By means of the Image Registration process, the appropriate control points are selected and the appropriate parameters are combined in the registered image giving out the correct 3D output in the display. This algorithm also ensures the preserving of edge and pixel information in the final output image.

4.3 COMPARISON RESULTS

Here the results of the DIBR algorithm and the Image Registration process are compared. The results imply that the DIBR algorithm is giving a better and visually comfortable result when compared with that of the Image Registration process. The overlaying process of the image registration requires the maximum number of parameters to be chosen for the more similar result.

4.3.1 DIBR vs Image Registration

The algorithm fuses the different views of the image obtained and generated the 3D display as the output image. Whereas the image registration process is an existing algorithm for registering the tow or more images of the same scene one over another. The algorithm giving the 3D output is compared with that of the two images registered together with the matching parameters. The results giving a clear visual display imply that the 3D output obtained from that of the algorithm is looking better when compared with that of the registered image.

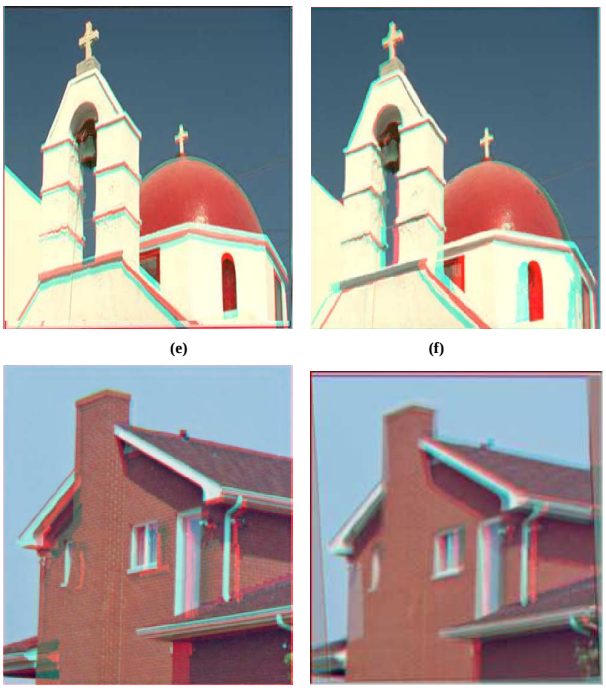
In the image registration process, based on the number of control points selected, the registered image will be fused over the base image. In case of correlation if used in the process, the control points will be locked only under the condition that they are matched. Thus the image registration process is an iterative process to be repeated to obtain the clear registered image. The comparative results can be discussed in the following figure 5.1.

Comparing the four sets of images named ‘image1.bmp’ , ‘image2.bmp’ , ‘church.bmp’, ’ house.bmp’, the DIBR algorithm is giving a clear and visually comfortable output images when comparing with the results of the image registration process. The results are discussed and display below:

(a) **(b)**

(c) **(d)**

Figure 4.19 Comparative Results: (a), (c) are DIBR results; (b), (d) are Image Registration results.



(e) **(f)**

(g) **(h)**

Figure 4.20 Comparative Results: (a), (c) are DIBR results; (b), (d) are Image Registration results.

This DIBR algorithm is quality scalable depending on the block size. Hence the result processed by means of DIBR algorithm is showing a visually comfortable output when compared with that of the image registration process.

4.3.2 Run Time vs Block Size

This DIBR algorithm depends much on the block size of the image considered. Without any particular parameter, the block size has been chosen in a random manner. So it is clear from the result that smaller the block size better the depth quality and larger the block size lower the computational complexity.

Table 4.1 Run time versus Block size

BLOCK SIZE	4 x 4	8 x 8	16 x 16	32 x 32
AVERAGE RUN TIME (ms)	13710	3659	2417	825

Thus from the above table, it is inferred that as the block size increases the computational complexity is lowered along with that of the run time. The smaller block size gives us a better depth quality when compared with that of the larger ones. Thus this algorithm is highly promising for various emerging 3D applications.

Also depending on the size of the block chosen, many parameters can be compared such that run time as shown above, along with that of the computational complexity, depth map quality and so on. This gives a gist of the algorithm processing over the image and the required parameters to be concentrated during the processing of the image. The parameters are compared as follows:

4.3.3 Parameters Comparison

The following table compares various parameters like computational complexity, depth map quality and average run time in terms of whether the block size is smaller or larger. The results are given in the table 5.2 implying that smaller the block size, the obtained depth map quality is fine to produce a visually comfortable output.

Table 4.2 Parameters Comparison

BLOCK SIZE	COMPUTATIONAL COMPLEXITY	DEPTH MAP QUALITY	AVERAGE RUN TIME
SMALLER	HIGH	BETTER	HIGH
LARGER	LOW	GOOD	LOW

Larger the block size, even though the computational complexity is lower because of the respective region grouping algorithm, the depth map quality is slightly less when compared with that of the smaller block size.

CHAPTER 5

CONCLUSION

This work has presented a novel 2D-to-3D conversion algorithm. The proposed algorithm utilizes edge information to group the image into coherent regions. A simple depth hypothesis is adopted to assign the depth for each region and a cross 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. The Image Registration process overlays the two images based on the parameters selected and gives out a better result combining the parameters of both the images of the same scene. This is an existing algorithm which when compared with that of the proposed algorithm, gives out a less comfortably visual 3D output in the final display of the image.

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] 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

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

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

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

[5] 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

[6] 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

[7] D. Kim, D. Min, and K. Sohn, “A Stereoscopic Video Generation Method Using Stereoscopic Display Characterization and Motion Analysis”, in *IEEE Trans. On Broadcasting*, Vol. 54, Issue 2, pp. 188-197, 2008

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

[9] 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

[10] S. Battiato, S. Curti, E. Scordato, M. Tortora, and M. La Cascia,”Depth map generation by image classification”, *Three-Dimensional Image Capture and Applications VI*, vol. 5302, pp. 95-104, 2004

[11] G. Economou, V. Pothos and A. Ifantis, “Geodesic distance and MST based image segmentation”, in *Proc. European Signal Processing Conf*, 2004

[12] 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*

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

[14] Digital image processing using MATLAB by Rafael C. Gonzalez, Richard E.Woods, Steven L.Eddins.