Semi-Automatic 2D-to-3D Image Conversion Techniques for Touchscreen Device Applications

Dr. Lai-Man Po

Department of Electronic Engineering
City University of Hong Kong

Date: 15 October 2012

Seminar on “Latest Market Development on Display Technologies for Smart TV”

Content

- Recent Development of 3D display Technology
- Basic Concepts of 2D-to-3D Image Conversion
- Why Semi-Automatic 2D-to-3D Image Conversion?
- **A User-Guided Semi-Automated 2D-to-3D Image Conversion Algorithm for Touch Screen Devices**
- Demo on LG Autostereoscopic 3D Smartphone
- Conclusions

Development of TV Technology

- Black/White TV (1929)

1929, BBC

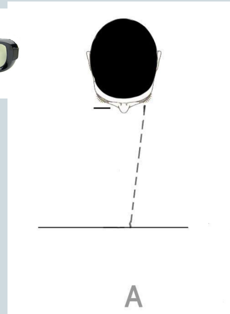
- Color TV (1954)

1954, U.S

- Digital TV (2001)
  - Standard-Definition (720x480)
  - High-Definition (1920x1080)

2001, Korea

- 3DTV (Stereoscopic)
  - Shutter Glasses
    - Alternately darken over the eyes



  - Polarization Glasses
    - Each eye receives only the intended image by polarization

Recent Development of TV Technology

Three directions to further improve the TV technology

- Increase the intelligence:
  - Smart TV with many new functionality such as Web Browser, Skype, Facebook, etc.
- Increase the resolution
  - 4K UHDTV with resolution of 3840x2160 (8.3 Mpixels)
  - 8K UHDTV with resolution of 7689 x 4320 (33.2 Mpixels)
- Increase the number of views
  - Autostereoscopic Multiview display (Non-Glasses)

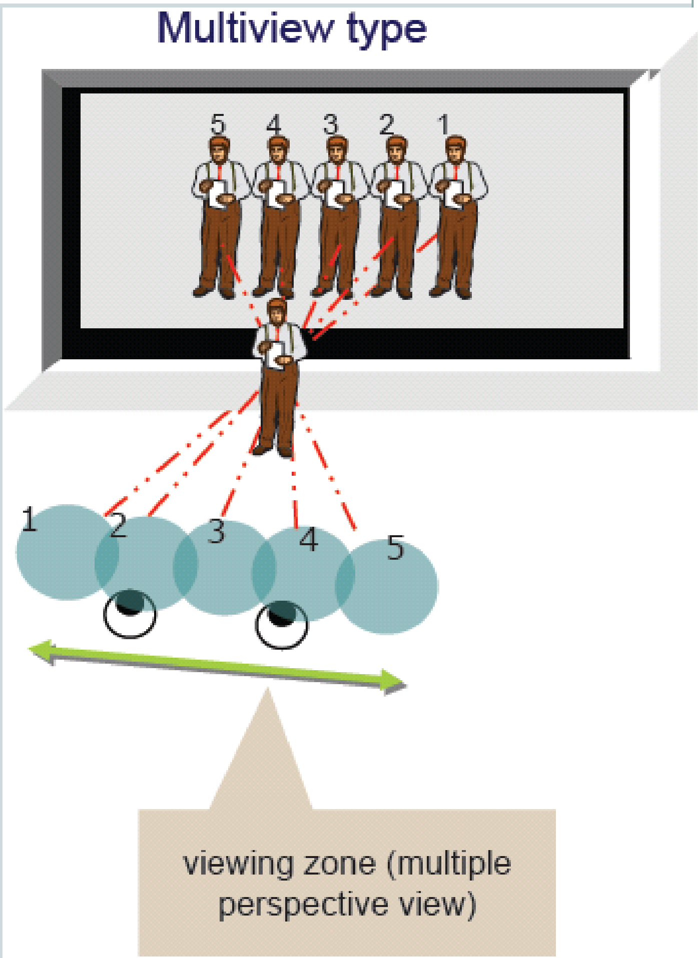
Ultra-High Definition TV (UHDTV)

- Increase the resolution
  - 4K UHDTV with resolution of **3840x2160** (8.3 Mpixels)
  - 8K UHDTV with 8K resolution of **7689 x 4320** (33.2 Mpixels)

LG unveils 84-inch 'ultra definition' 4K TV it's bringing to CES 2012

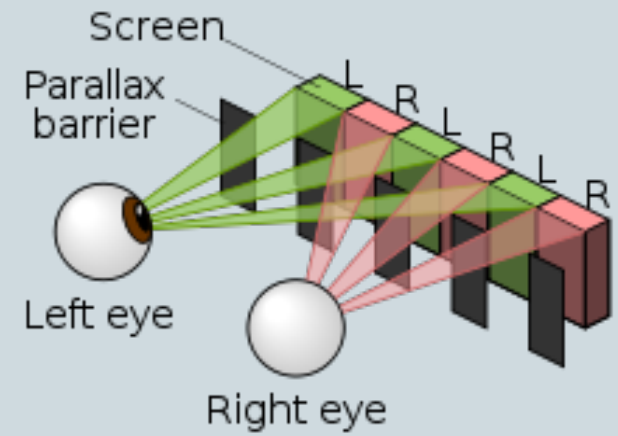
Autostereoscopic Multiview TV

- **Increase the number of views**
- **Non-Glasses & Multiview**
- Technologies for autostereoscopic
  - Parallax barriers
  - Lenticular lenses
- Redirect incoming imagery to several viewing regions
- Different viewing positions perceive a different image with each eye to provide a stereo image.

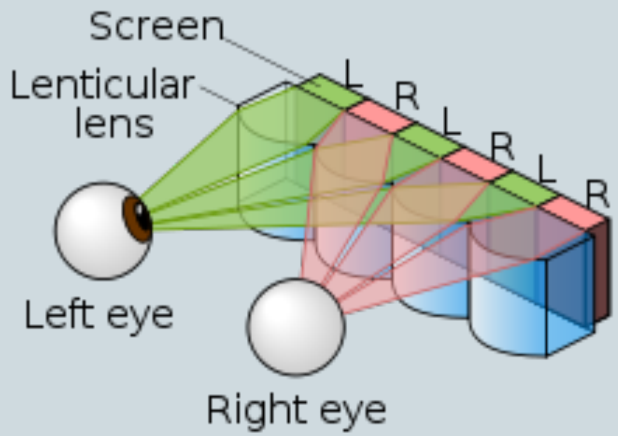


Two-View Autostereoscopic Displays

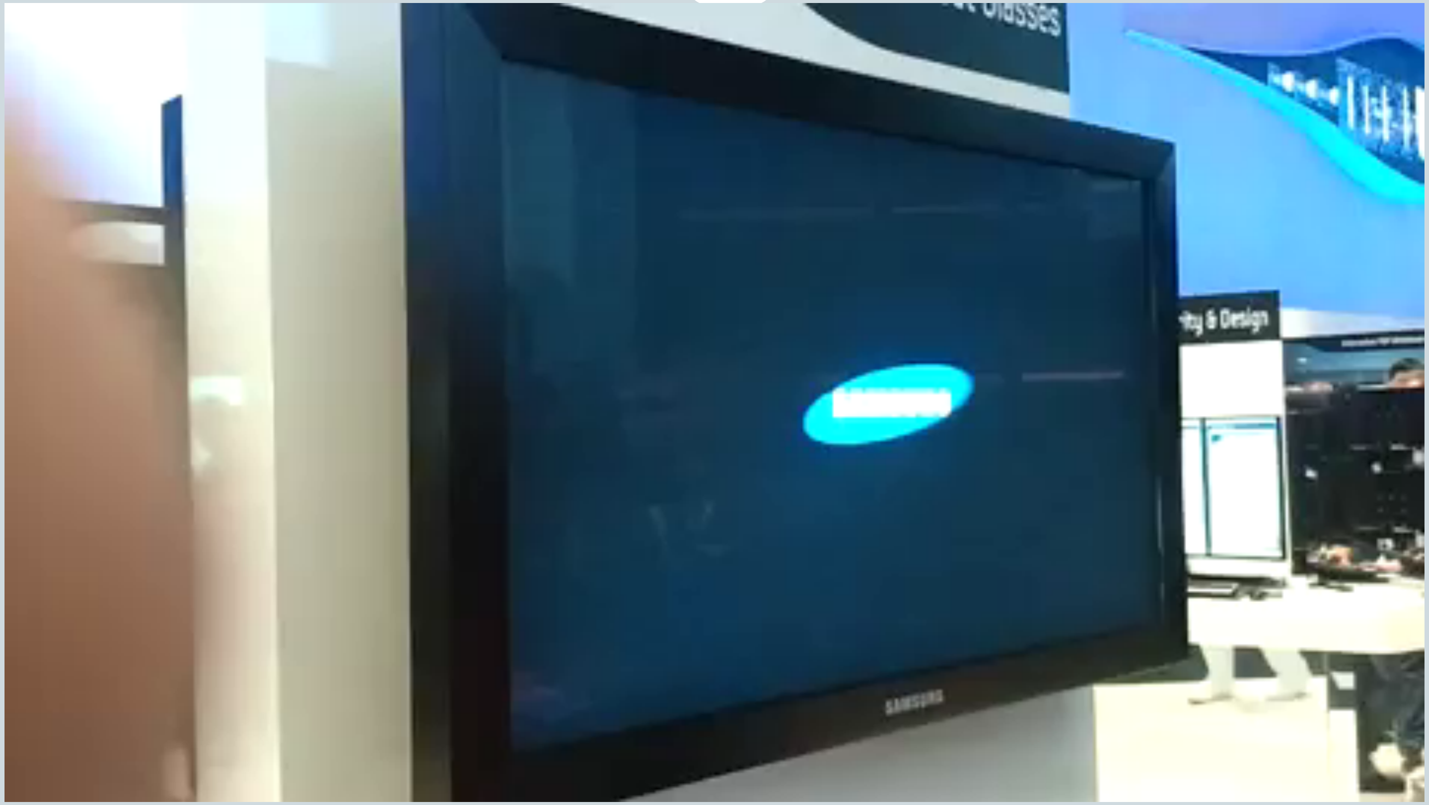
- **Parallax Barrier**
  - A barrier mask is placed in front of the pixel raster so that each eye sees light from only every second pixel column.



- **Lenticular Arrays**
  - An array of cylindrical lenslets is placed in front of the pixel raster, directing the light from adjacent pixel columns to different viewing slots at the ideal viewing distance so that each of the viewer's eyes sees light from only every second pixel column



Samsung 2012 3D Flat TV without Glasses



- This is just a prototype for autostereoscopic 3DTV and not yet available to commercial market.

Autostereoscopic 3D Handheld Devices are Available

- 3D Tablet

- 3D Photo Frame

- 3D Smartphone

LG Optimus 3D P920 Smartphone

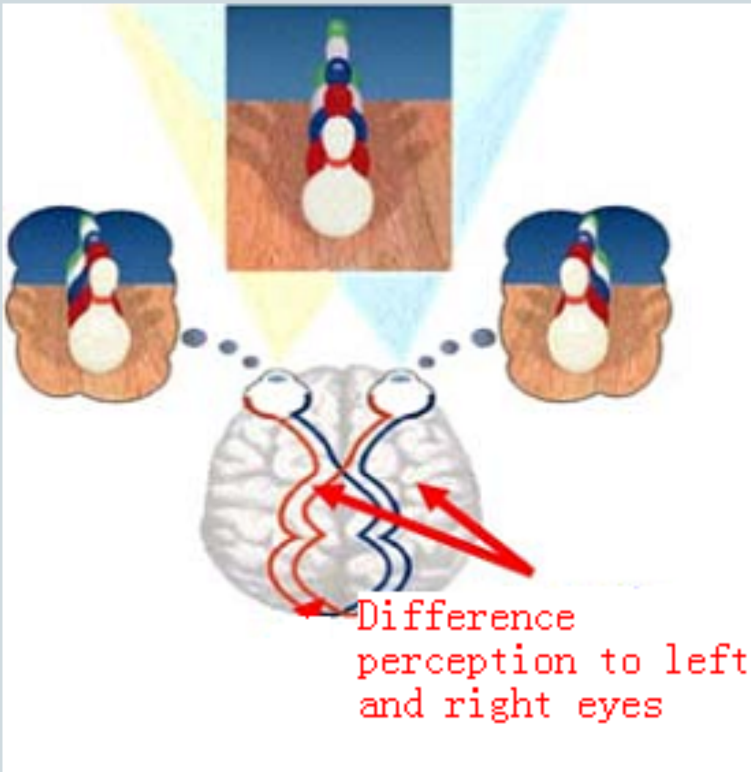
Why 2D to 3D Image Conversion?

- Although 3D handheld devices are available, we don't have lots of 3D photos that can be displayed on them.
- Even the 3D camera is also available, we cannot retake the pictures at these special moments of our life.
- It should be very interesting and meaningful
  - To convert your favorite pictures from 2D to 3D and show them on the autostereoscopic 3D display
  - That offer a more realistic sense of the scene to the viewer.

Basic Concepts of 2D to 3D Image Conversion

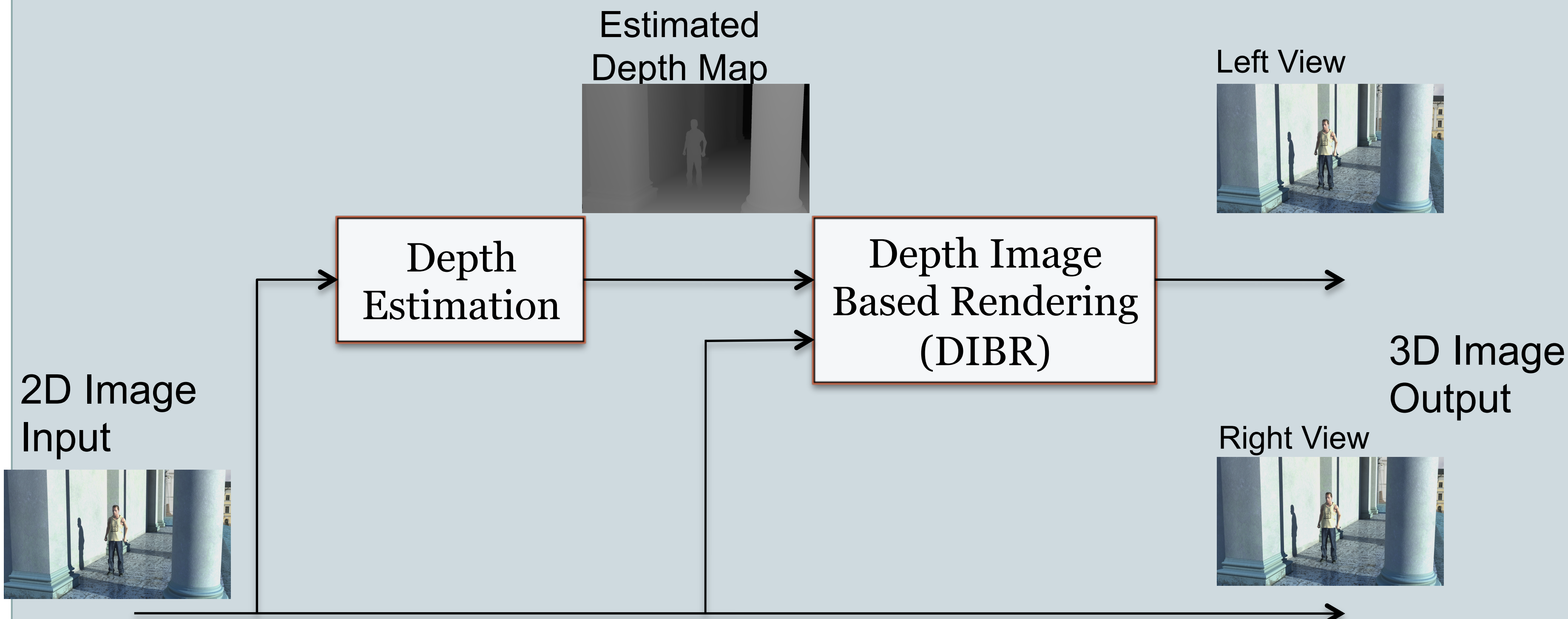
3D Perception of Human

- Binocular parallax
  - Small spatial displacement between the left and right eyes



Block Diagram of 2D-to-3D Image Conversion

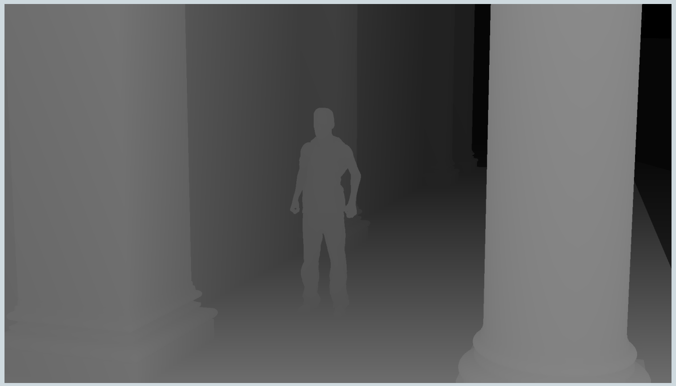
- The main purpose of the 2D-to-3D image conversion is to generate the second view image based on the structure of the input 2D image, which involve two processes: (1) Depth Estimation and (2) Depth Image Based Rendering (DIBR)



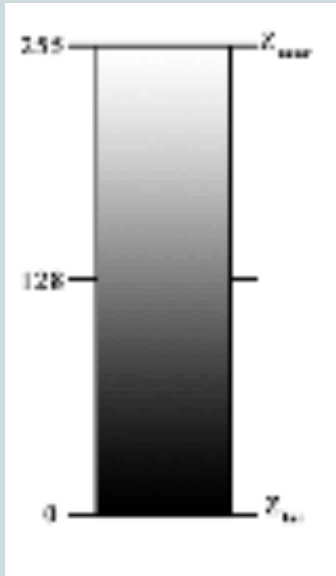
What is Depth Map?

- The depth map is a grayscale image, indicating how far or near objects are from the screen.
- Each depth image stores depth information as 8-bit grey values with the grey level 0 indicated the furthest value and the grey level 255 specifying the closest value.
- A bright value would indicate a close object and a dark value would indicate an object is far away from the screen.

2D Color Image

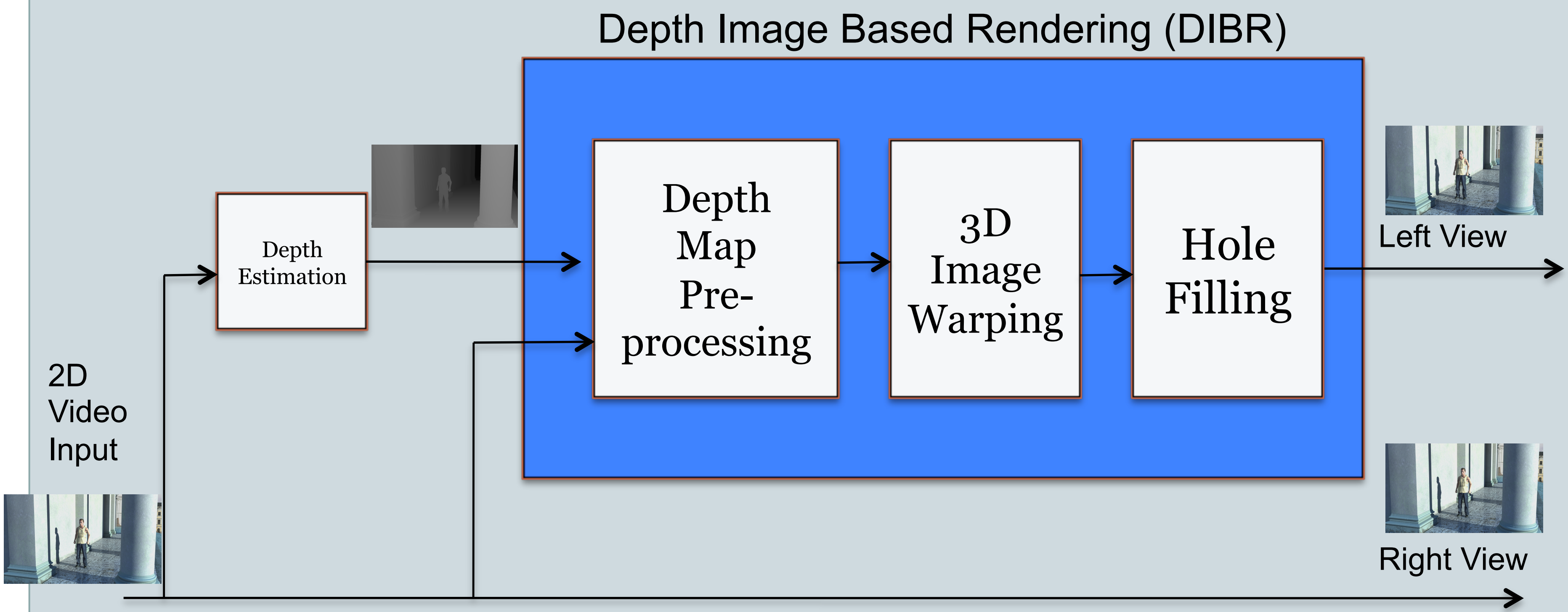


Depth Map or Depth image



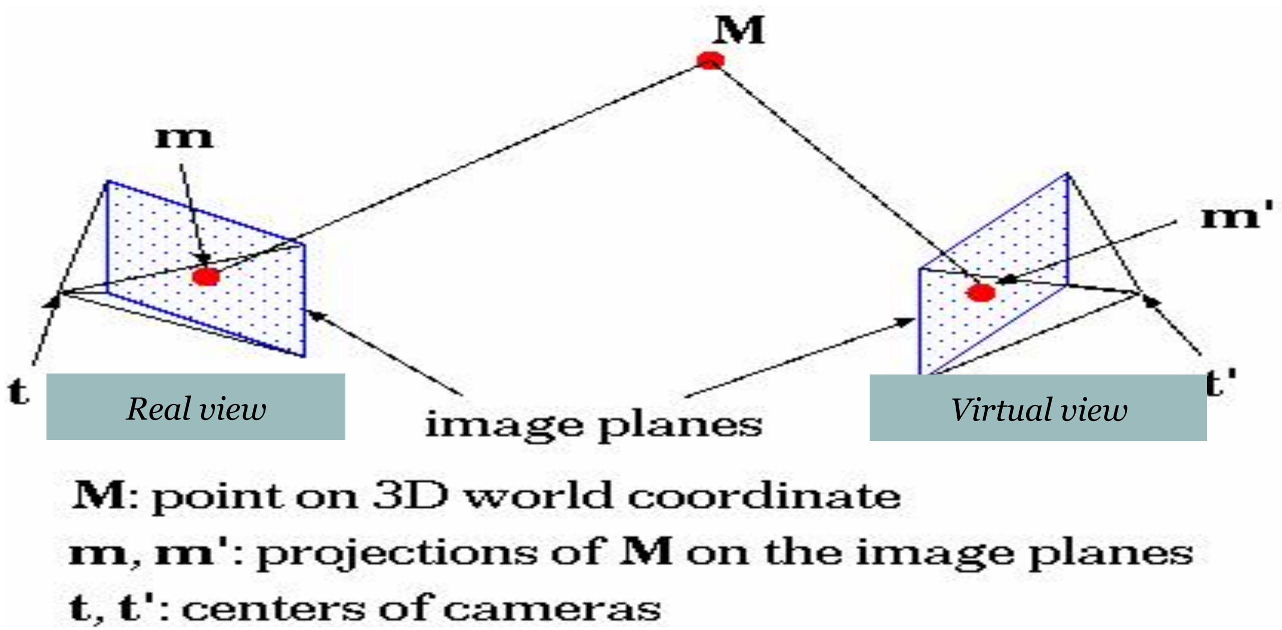
Depth Image Based Rendering (DIBR)

- To generate the 3D image, DIBR is used to syntheses the second view video based on the estimated depth map and the 2D image input.
- DIBR consists of three processes

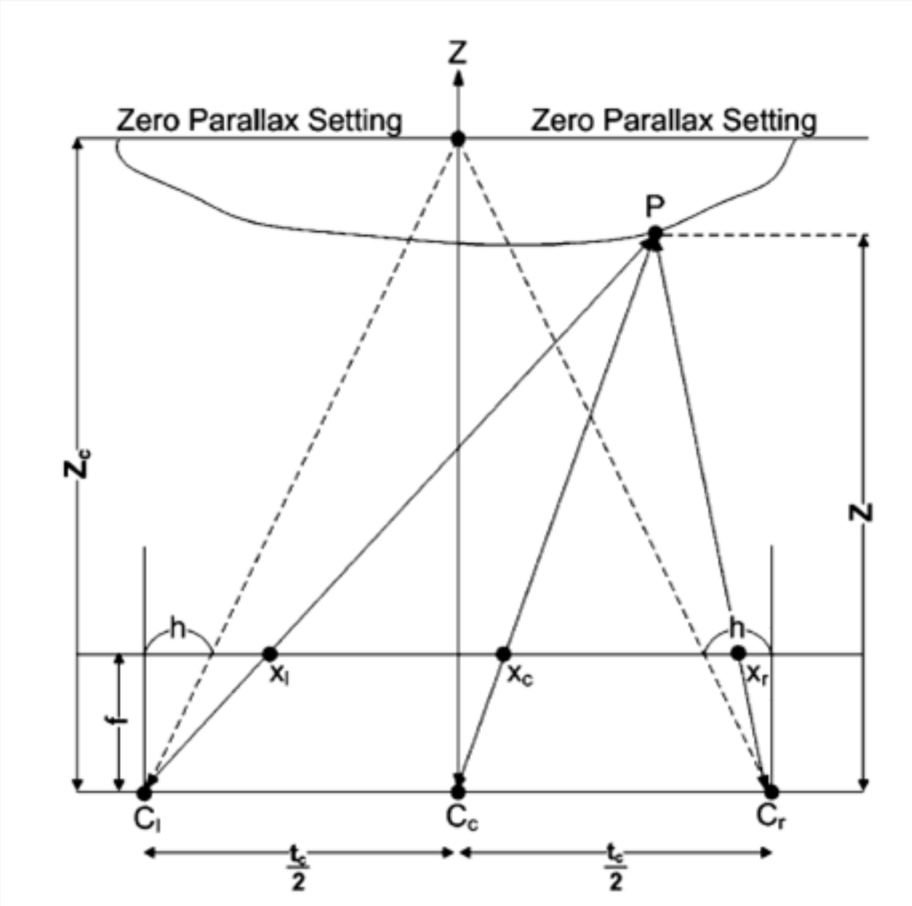


3D Warping

- The process includes two steps:
- Original image points (e.g. m(x,y)) from the real view are re-projected into the 3D world
- The 3D space points (e.g. M(X,Y,Z)) are projected into the image plane of the "virtual" view (e.g. m'(x',y')).



3D Warping



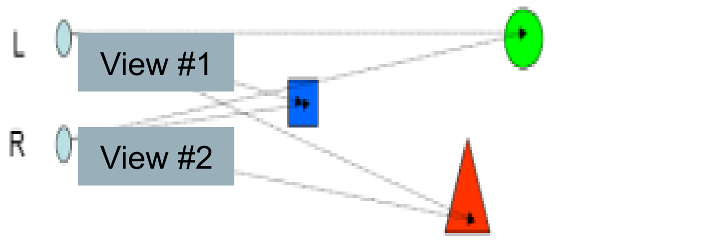
$$x_l = x_c + \frac{t_c f}{2Z(x_c, y)} + h$$

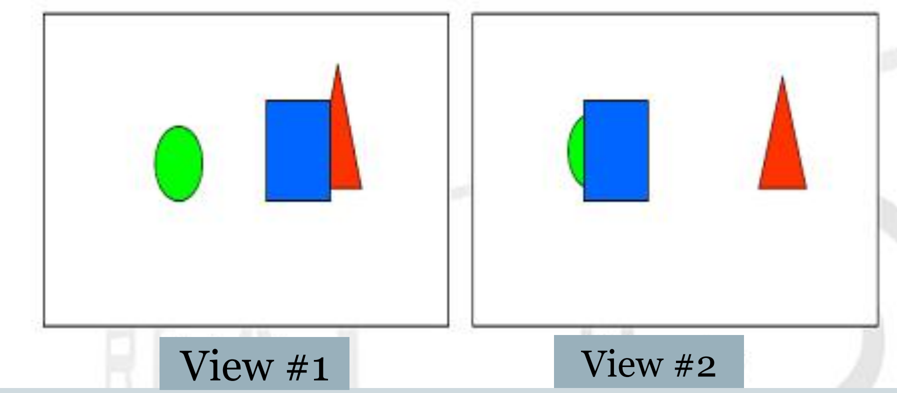
$$x_r = x_c - \frac{t_c f}{2Z(x_c, y)} - h$$

$$h = -\frac{t_c f}{2Z_c}$$

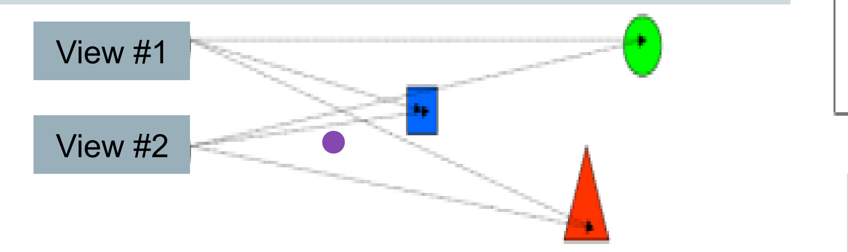
Major Challenges in DIBR

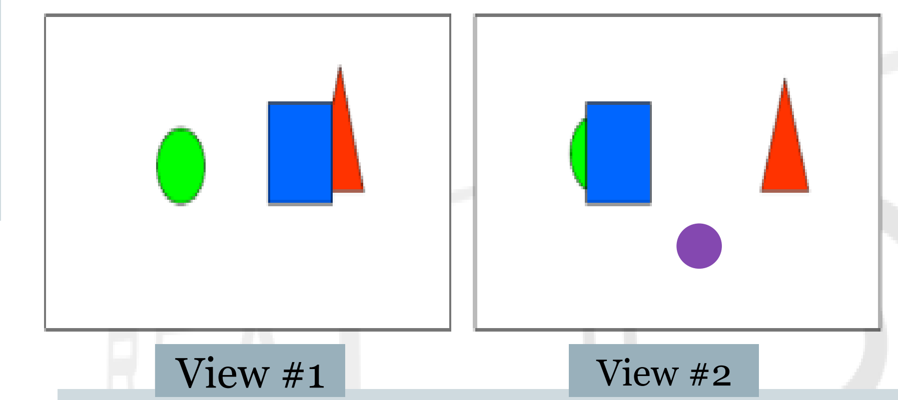
- **Facts:**
  - Two different points in the image plane at the real view can be warped to the same location in the virtual view. [Occlusion]





- Occluded area in the real view may become visible in the virtual view. This is called disocclusion





Holes Created in 3D Image Warping

Depth Map

Right View Image

3D Image Warping

Left View Image created by 3D Image Warping with holes due to disoculsion.

Hole-Filling of Warped Image

Left View Image after hole-filling

Enlarged Left View Image after hole-filling

Depth Estimation from a Single Image

- Depth Estimation from a single image is very difficult.
- It is because 3D information is naturally lost in 2D image.
- It is also a very computational intensive process.

Conventional Depth Estimation Methods

- **Use of Color Channel** : Used the Cr channel of an image after converting to YCbCr as the depth map.
- **Vanishing Point based Depth Estimation** : The main idea is to find out the vanishing point that is the farthest point of the whole image.
- **Object Classification Method** : Use an object classifier based on the Bayesian learning algorithm to classify objects into different types, and then assign relative depth values to these objects.
- **Supervised Learning** : Use a supervised learning approach by training a classifier with a set of images and their corresponding ground-truth depth maps. Depth maps are predicted as a function of the input image by using a set of filters to generate a feature vector.

Pros and Cons of Conventional Approaches

Pro :

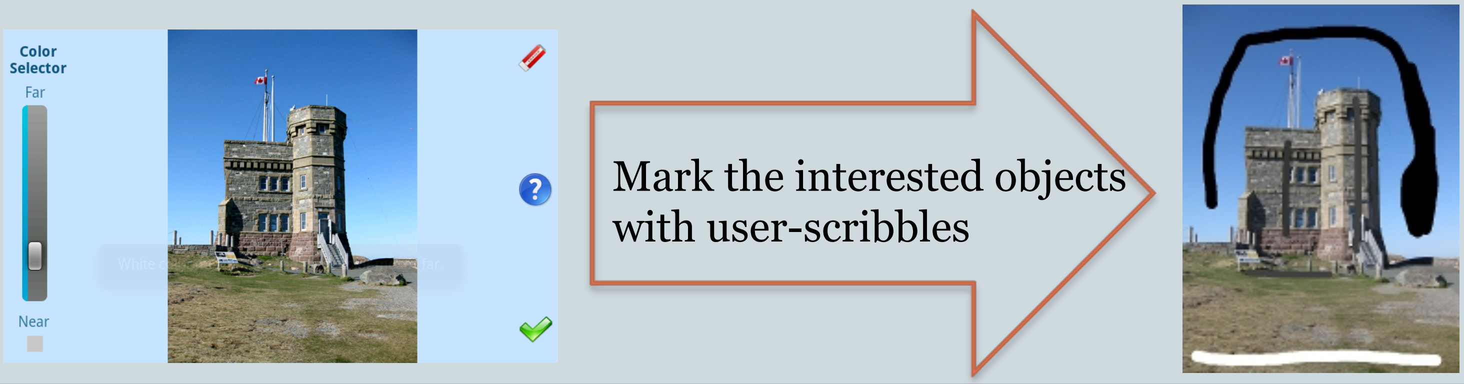
- The benefit of these conventional depth map estimation methods is that they are all **automated**.

Cons :

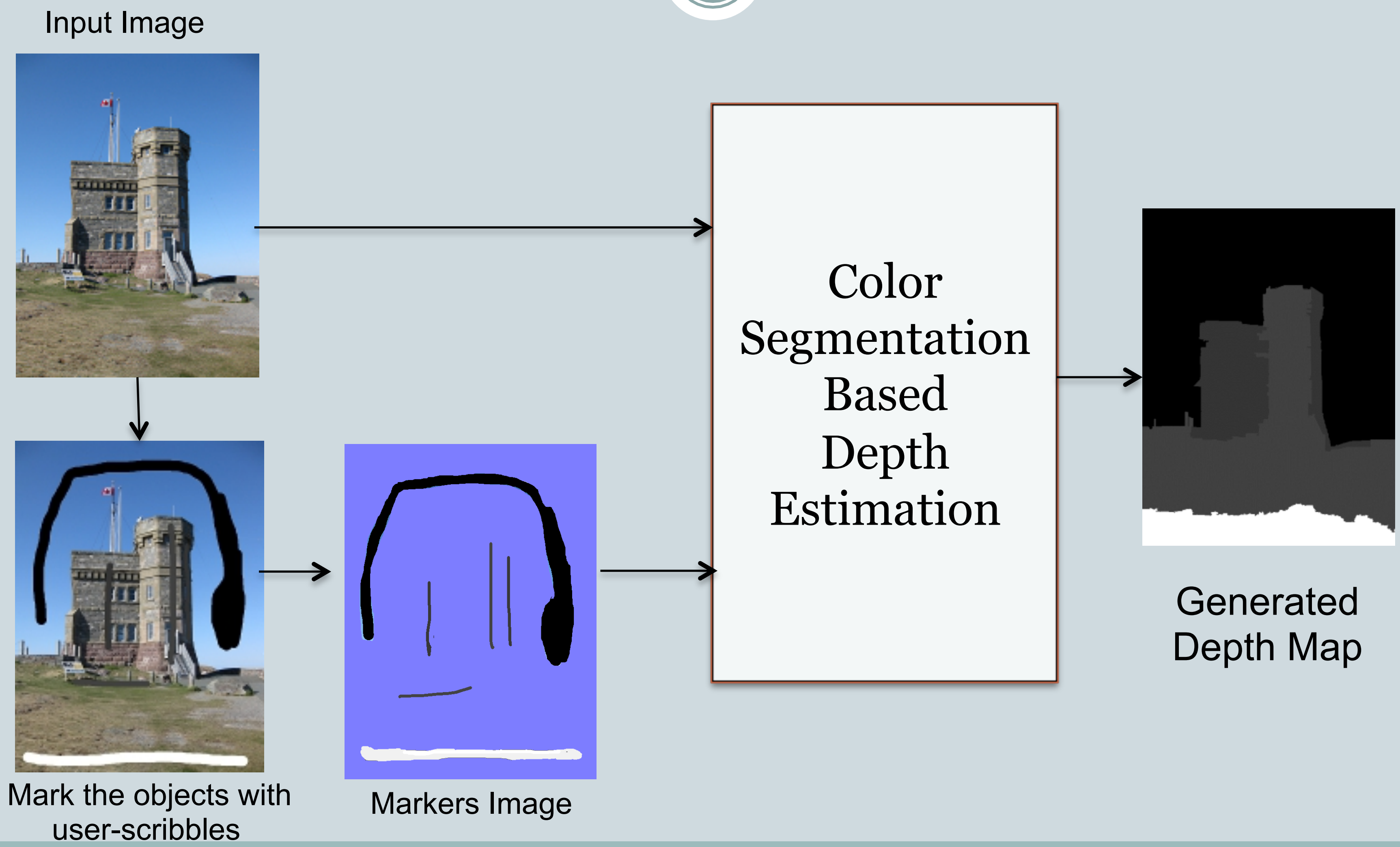
- They do not provide any means of correcting errors where object may appear to be at the wrong depth.
- Very annoying artifacts may appear in the synthesized views
- **Basically, automatic approaches are hard to achieve high quality 2D-to-3D image conversion**

User-Guided Semi-Automated Depth Estimation

- **Semi-Automated Methods are possible to achieve high quality depth estimation as they provide user's input and feedback for corrections**
- In our proposed method, user only needs to mark different objects of interest and assign each a corresponding depth value.
- The interface is very user-friendly for touchscreen devices.
  - A sliding bar is used for selecting depth values
  - Lighter denote closer to the camera and darker denote farther from camera

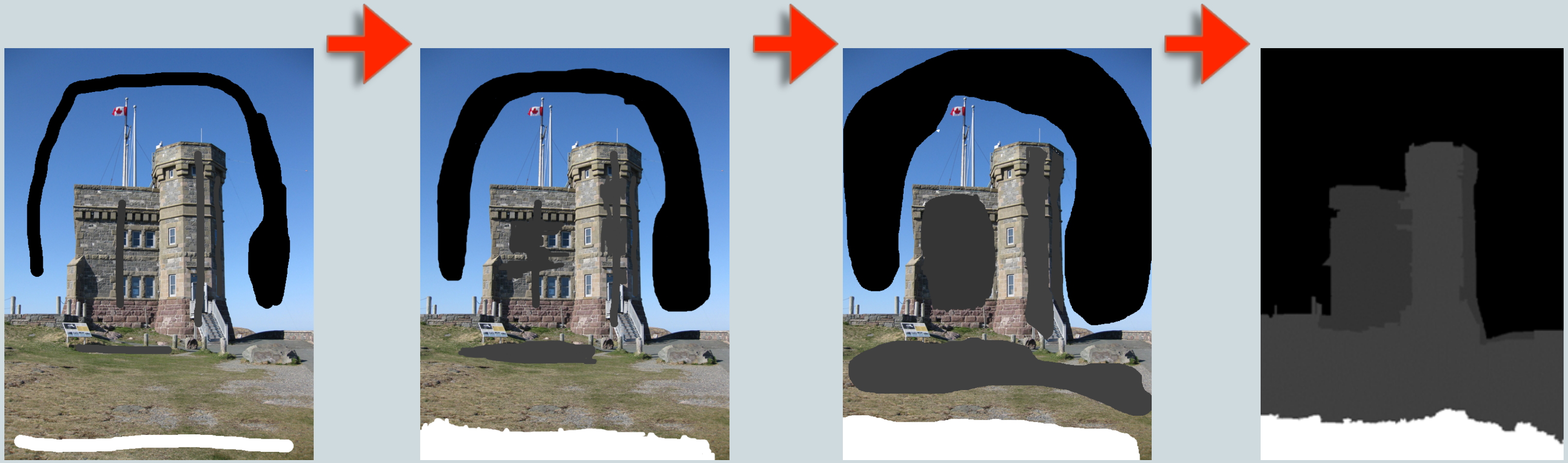


Semi-Automatic Depth Estimation Process



Growing of the Markers

- Rather than considering each label as a separate object in conventional segmentation process, each label is considered as a separate initial *depth region*.
- The color segmentation based depth estimation can be interpreted as a diffusion simulation where the initial depth values are diffused throughout the depth image but it is based on the color image content.



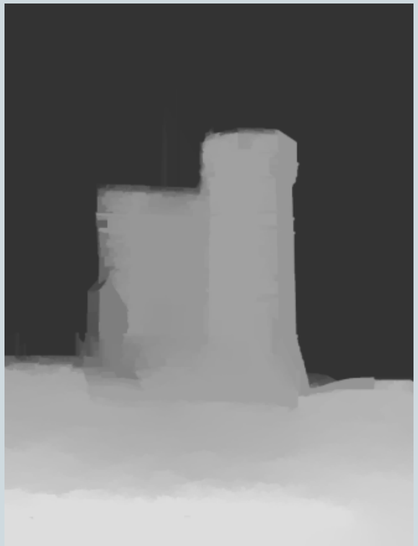
Algorithms for Color Segmentation Based Depth Estimation

- Three conventional color segmentation algorithms can be used for semi-automatic depth estimation:
  - **Random Walks**
  - **Graph Cuts**
  - **Watershed**

Random Walks based Depth Estimation

- Pro : Smoothing properties inside objects
- Con : Smooth boundaries
  - Random Walks is also the solution to a linear system and therefore has problems preserving strong edges

Input Image with marker



Depth map from Random Walks

Graph Cuts based Depth Estimation

- Pro : Hard segmentation/sharp boundary
  - Graph Cuts does extremely well in preserving strong edges
- Con : Flat depth values inside objects
  - The hard segmentation with Graph Cuts does not respect smooth gradients or fine detail.

Input Image with markers

Depth map from Graphs Cuts

Watershed based Depth Estimation

- Hard segmentation/sharp boundary
  - Watershed is as well as Graph Cuts in preserving strong edges
- Fast
  - The computational requirement of Watershed based depth estimation is much lower than Graph Cuts
  - Complexity comparison on Intel i7 CPU 2.93GHz 4G RAM

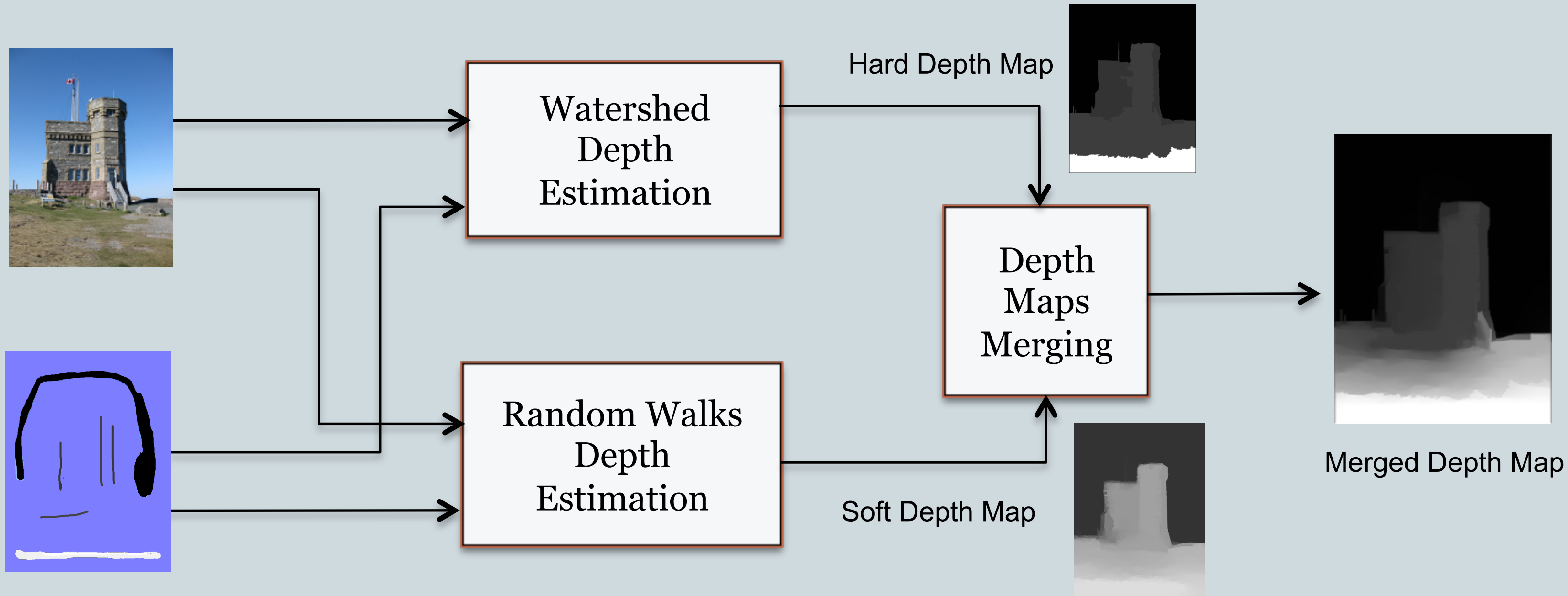
Input Image with markers

Depth map from Graph Cuts

Depth map from Watershed

Merging of Soft and Hard Estimated Depth Maps

- To generate the depth map with sharp boundary and fine depth detail inside the objects, a two stage process utilizing the smoothing properties of Random Walks, and the hard segmentation returned by Watershed.



- By combining the two, we can simultaneously retain strong object boundaries while also allowing for smooth gradients.

Depth Maps Merging Process

- Depth Prior is used to generate the final depth map.
- The depth map from Watershed is appended to the edge weight of the Random Walks. The edges are calculated

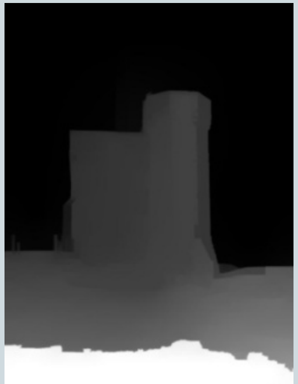
$$d(\vec{c_i}, \vec{c_j}, d_i, d_j | \alpha) = \sqrt{d(\vec{c_i}, \vec{c_j})^2 + (\alpha(d_i - d_j))^2}$$

- Where *d(ci,cj)* is color distance, *di* and *dj* are the normalized depth labels of pixels i and j from Watershed. α is chosen to equal to 0.5.
- By adding the depth map of Watershed as the edge weighting in the Scale-Space Random Walks, the edges values are increased only at the object boundary region.
- Thus the probabilities walk through the edge will decrease and the edges of depth can be preserved more clearly after the Random Walks.

Experimental Results

Input Image

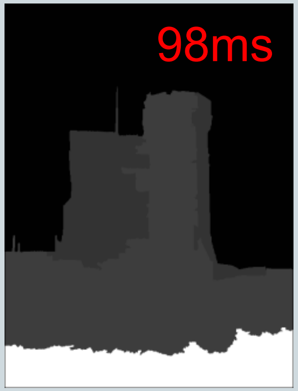
Depth map from Graph Cuts



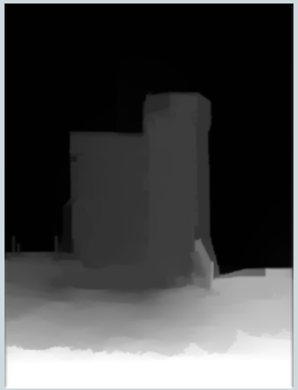
Final depth map from Graph Cuts and Random Walks

Generated anaglyph 3D Image

Input Image with markers



Depth map from Watershed



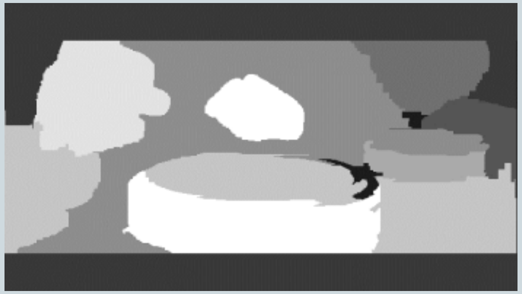
Final depth map from Watershed and Random Walks

Generated anaglyph 3D Image

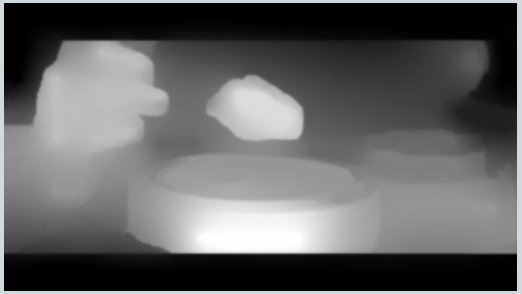
Intel(R) i7 CPU 293GHz
4GB RAM.

Experimental Results

Input Image



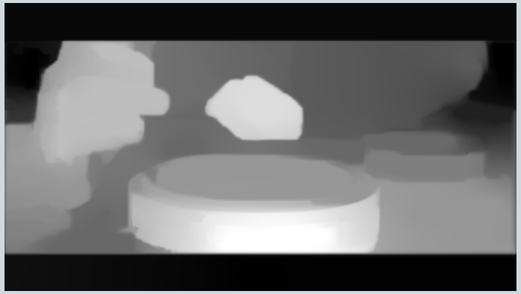
Depth map from Graph Cuts



Final depth map from Graph Cuts and Random Walks

Input Image with markers

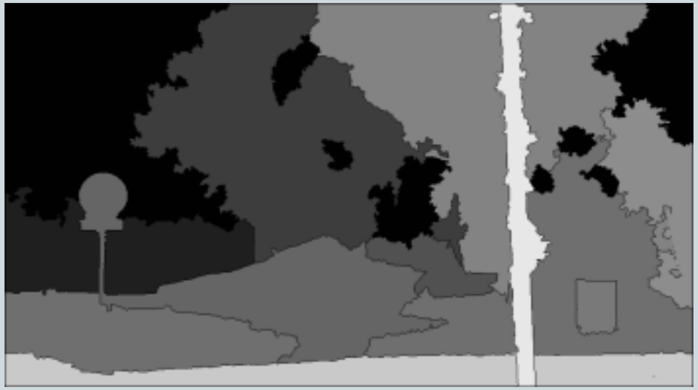
Depth map from Watershed



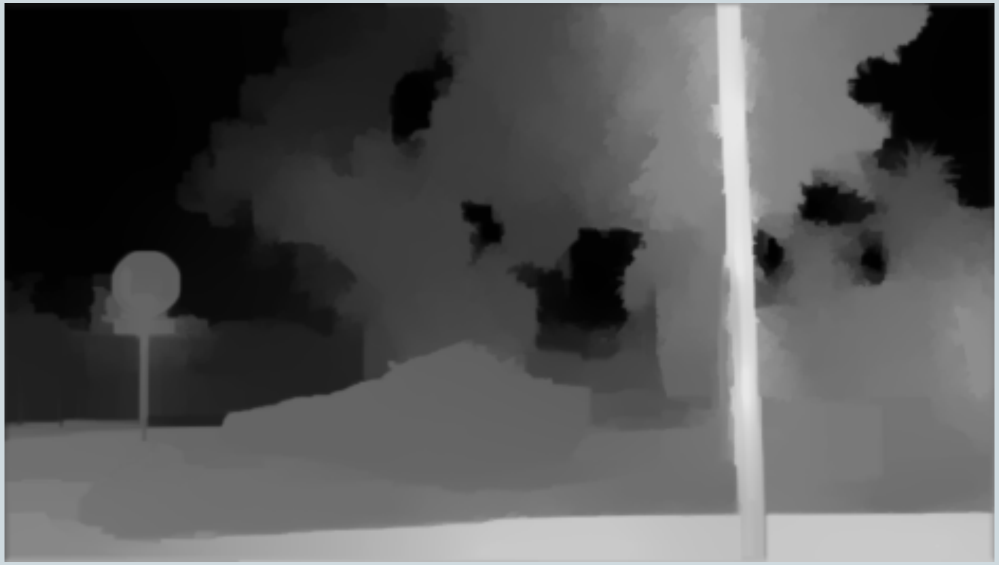
Final depth map from Watershed and Random Walks

Experimental Results

Input Image with markers

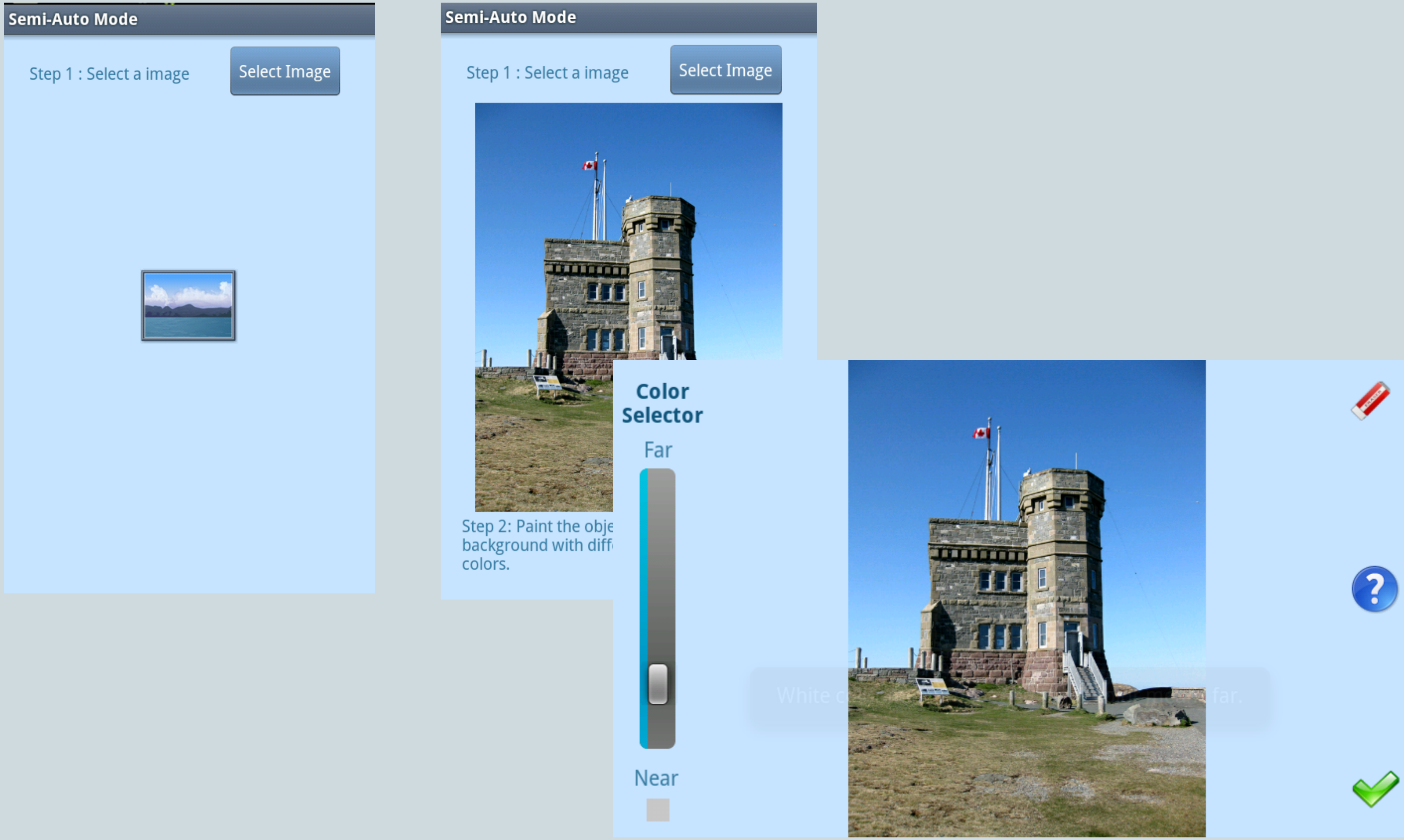


Depth map from Watershed



Final depth map from Watershed and Random Walks

2D-to-3D Android App on LG 3D Smartphone



2D-to-3D Image Conversion App Demo

Conclusions

- A semi-automatic 2D-to-3D image conversion algorithm that based on Random Walks and Watershed segmentations algorithm is developed.
- The proposed method is very suitable for applications on touchscreen devices with 3D autostereoscopic display.

- Problems still exist in this semi-automatic 2D-to-3D image conversion approach:
  - Different initial labels may results various output depth maps
  - For complicated image, more labels are required to generate better results.

Thank You for your attention!!