Technical Challenges of 3-D Video Processing

July 11th, 2011

Prof. Yo-Sung Ho

Gwangju Institute of Science and Technology

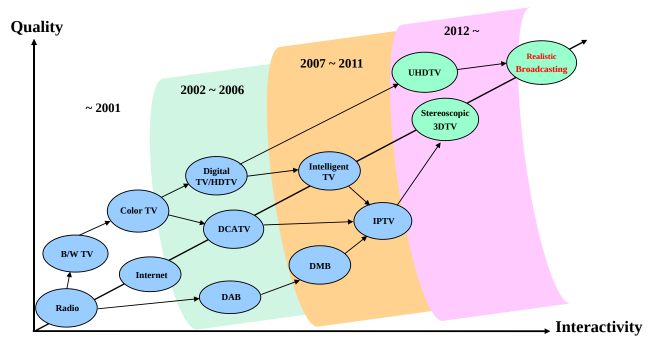
Yo-Sung Ho: Biographical Sketch

- 1977~1983 Seoul National University [BS, MS]
- 1984~1989 University of California, SB [Ph.D.]
- 1983~1995 Electronics & Telecomm. Research Institute (ETRI)
- 1990~1993 North America Philips Labs, NY, USA (NAPL)
- 1995~Now Gwangju Institute of Science & Technology (GIST)
- 2003~2011 Director of Realistic Broadcasting Research Center
- 1981~Now Senior Member of IEEE, SPIE, IEEK, KICS, KSBE
- 2004~2011 Associate Editor of IEEE Trans. on CSVT
- 1991~Now Delegate to MPEG/JVT Meetings
- 1999~Now Tutorial Lecturer at Various International Conferences

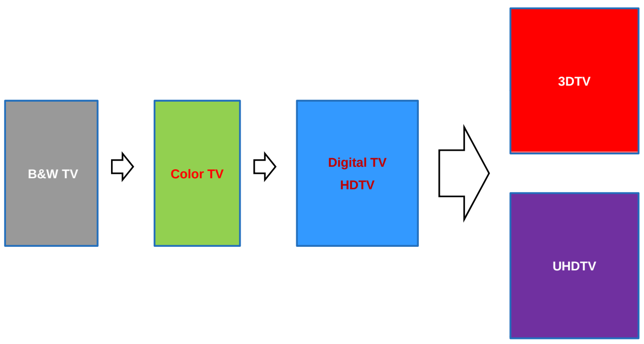
Outline

- Trend of Broadcasting Technologies
- Basics of 3DTV System
- MPEG Activities for 3D Video Coding
- Multi-view and Free-viewpoint Video
- Multi-view Camera System (MCS)
- Hybrid Camera Systems (HCS)
- Conclusions

Trend of Broadcasting Technologies

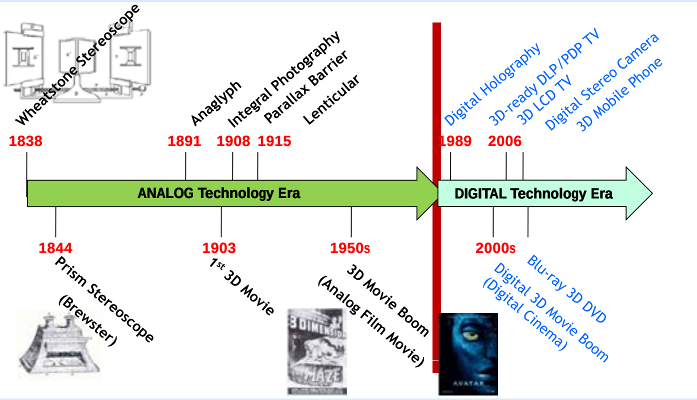


Evolution of TV Technologies

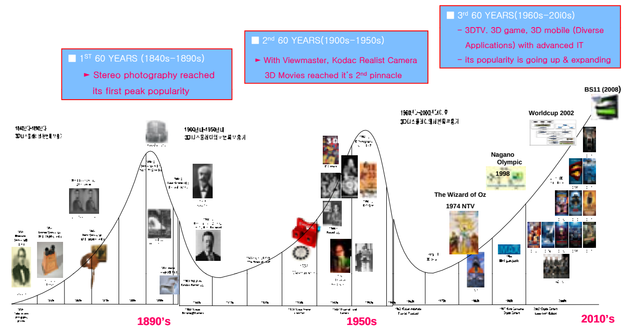


Hit of 3D Movies

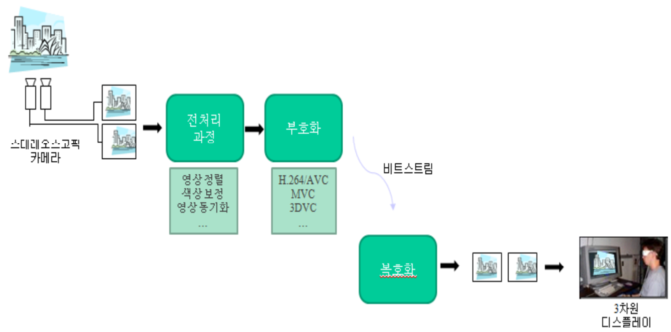
History of 3D Technologies



Development of 3D Technologies

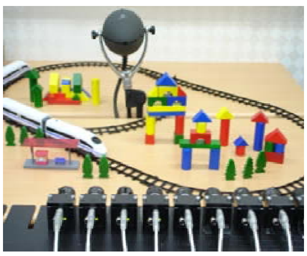


3DTV System



Stereoscopic Cameras

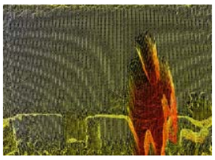
Multi-view Cameras



Multi-view Camera System

- Various arrangements of multiple cameras
- Classified by the shape and dimension
  - Parallel vs. convergent camera arrays
  - 1-D vs. 2-D camera arrays

Stereoscopic 2D + Depth Cameras



Multi-view 2D + Depth Cameras

3D Display Monitors

Anaglyph

Polarization

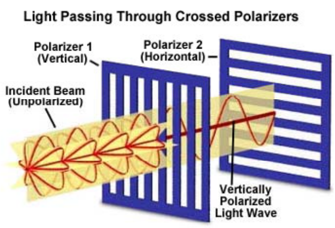
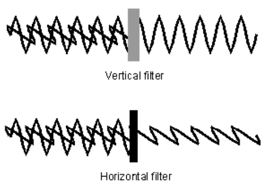
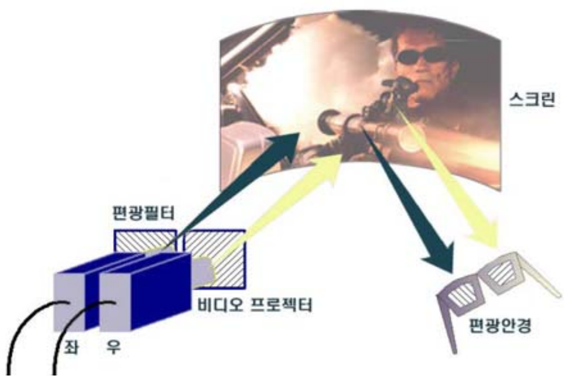


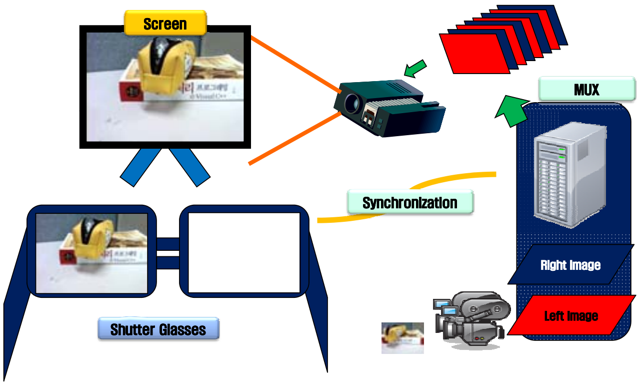
Figure 1



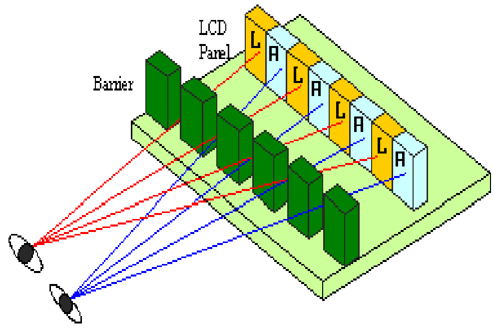
Polarized Glasses



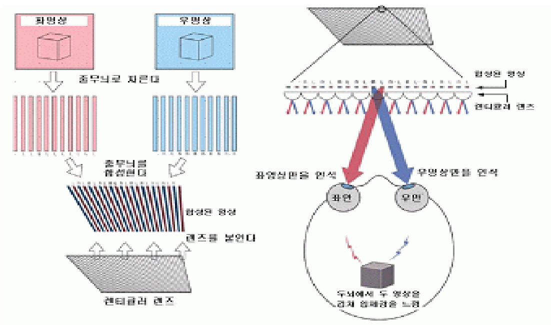
Shutter Glasses



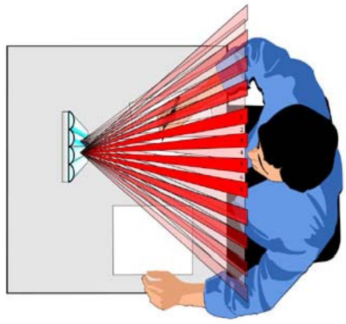
Parallax Barrier



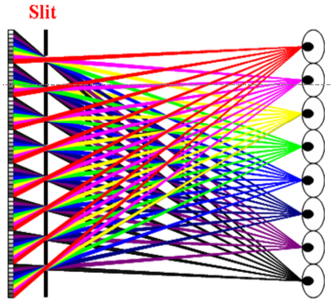
Lenticular Sheet



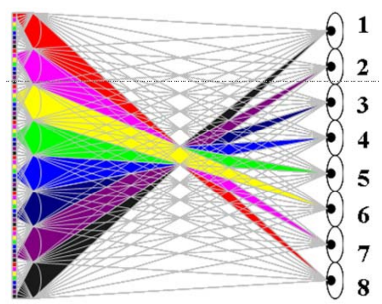
Auto-stereoscopic 3D Display



Multi-view Parallax Barrier

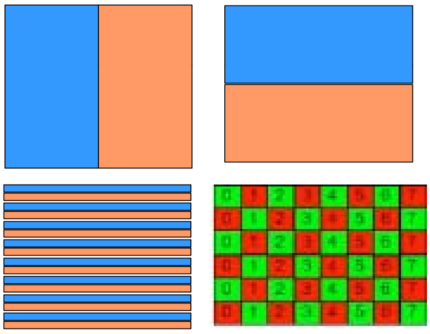


Multi-view Lenticular Sheet

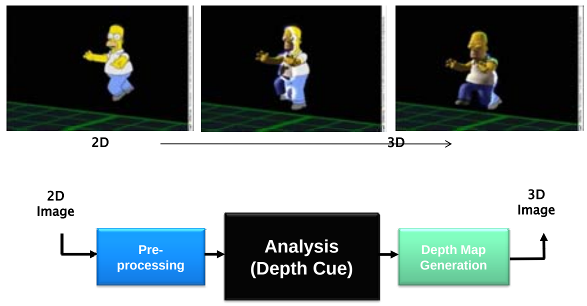


3D Image Formats

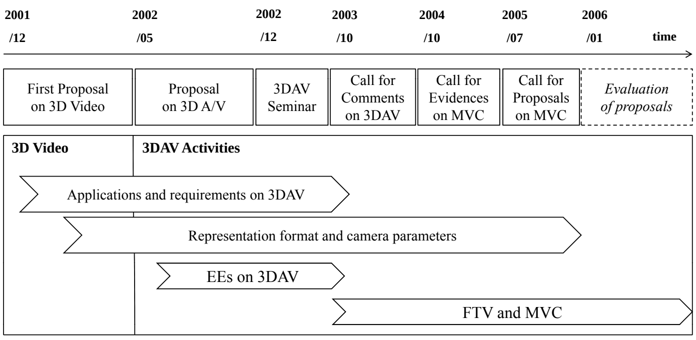
- Full resolution L+R
- Side-by-side
- Top-down
- Interlaced
- Checker-board



2D-to-3D Conversion



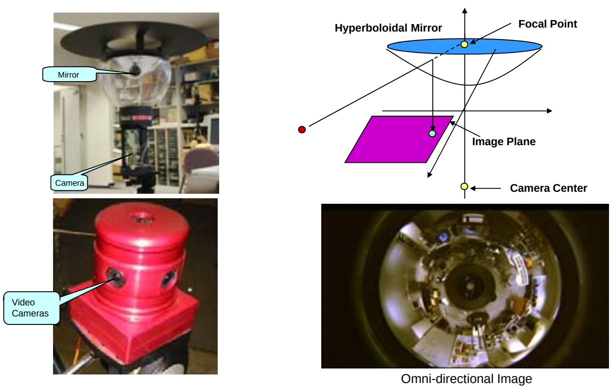
History of 3DAV/MVC



3DAV Exploration Experiments

- EE1
  - Omni-directional Video
- EE2
  - Free viewpoint Video
- EE3
  - Stereoscopic video coding using MAC
- EE4
  - Depth/disparity coding for 3DTV and intermediate view interpolation

EE1: Omni-directional Video



Omni-directional Image

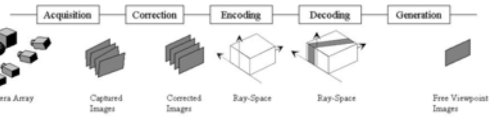
Omni-directional Video: Example

- Concentric Mosaic
  - Extension of the planar 2D image plane to a spherical or cylindrical image plane

Panoramic Video

EE2: Free Viewpoint TV

❖ FTV System

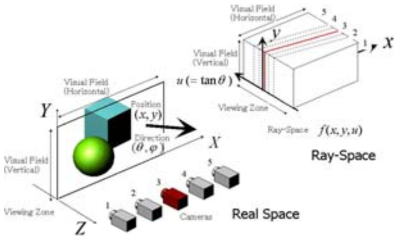


❖ Experimental System

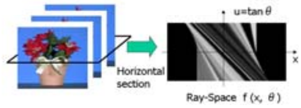
Generated Multi-view Images

Free Viewpoint Video: Ray Space

❖ Acquisition

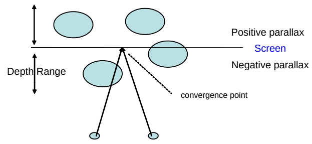


❖ Example



EE3: Stereoscopic Video

Left-view Image



Right-view Image

Stereoscopic Video Coding

❖ Two viewpoint images, one for each eye, are provided to produce 3-D impression to the viewer

❖ Various stereoscopic video coding methods were proposed:
- Method 1: Left-view Image + Right-view Image
- Method 2: Left-view Image + Residual Texture
- Method 3: Left-view Image + Disparity Map
- Method 4: Temporal/Spatial Scalability

EE4: Depth/Disparity Coding

❖ Color and Depth Information

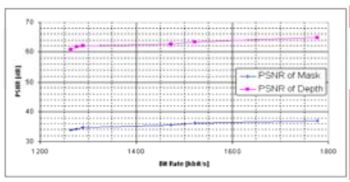
Color Information

Depth Information

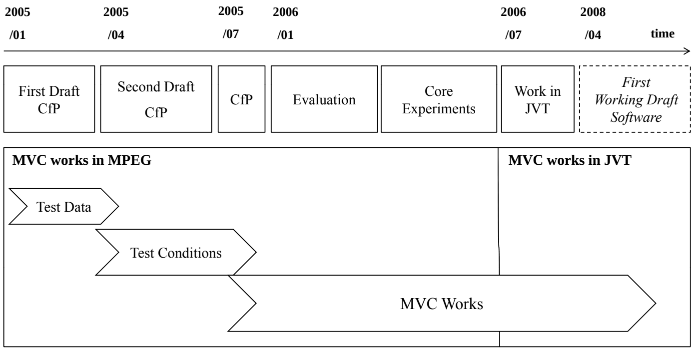
Depth/Disparity Coding

❖ Performance of Depth/Disparity Coding using MPEG-4

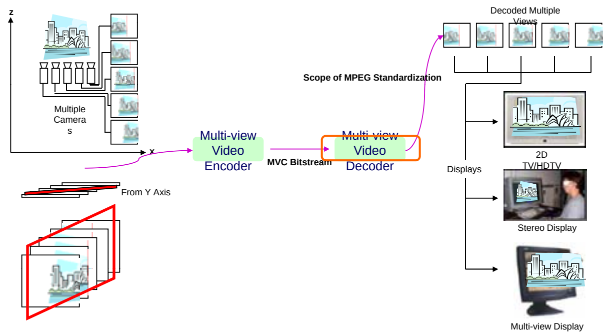
Color image Mask image Depth image



History of MVC

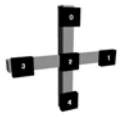


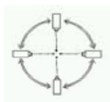
Multi-view Video Coding



Camera Arrangements

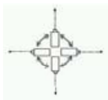

1D Parallel with 8 Cameras


2D Parallel with 5 Cameras


Convergent 4 cameras

1D Arc with 8 Cameras


2D Array with 128 Cameras


Divergent 4 cameras

Requirements for MVC

"shall" (mandatory)	▪ Compression efficiency ▪ View scalability/Free viewpoint scalability ▪ Backward compatibility ▪ Low delay ▪ Resolution, bit depth, chroma sampling format ▪ Temporal random access/View random access ▪ Resource management ▪ Parallel processing
"should" (desirable)	▪ Spatial/Temporal/SNR scalability ▪ Resource consumption ▪ Robustness ▪ Picture quality among views ▪ Spatial random access

Test Data for MVC

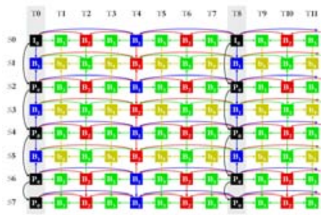
Data Set	Sequence	Image Property	Camera Arrangement
MERL	Ballroom and Exit	640x480, 25fps	8 cameras with 20cm spacing 1D/parallel (rectified)
HHI	Uli	1024x768, 25fps	8 cameras with 20cm spacing 1D/parallel convergent (non-rectified)
KDDI	Race 1	640x480, 30fps	8 cameras with 20cm spacing 1D parallel (non-rectified)
	Flamenco2	640x480, 30fps	5 cameras with 20cm spacing 2D parallel (non-rectified)
Microsoft Research	Breakdancers	1024x768, 15fps	8 cameras with 20cm spacing 1D/arc (non-rectified)
Nagoya University	Rena	640x480, 30fps	100 cameras with 5cm spacing 1D/parallel (rectified)
	Akko&Kayo	640x480, 30fps	100 cameras with 5cm horizontal and 20 cm vertical spacing; 2D array (non-rectified)

Download Test Data

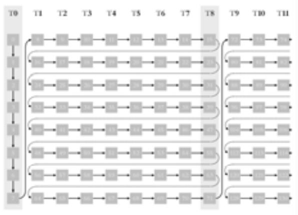
- MERL
  - ftp://ftp.merl.com/pub/avetro/mvc-testseq
- HHI
  - https://www.3dtv-research.org/3dav_CfP_FhG_HHI
- KDDI
  - ftp://ftp.ne.jp/KDDI/multiview
- Microsoft Research
  - http://www.research.microsoft.com/vision/ImageBasedRealities/3DVideoDownload
- Tanimoto Lab
  - http://www.tanimoto.nuee.nagoya-u.ac.jp/

MVC Reference Model

- ❖ MVC based on H.264/AVC
  - Is fully compatible to H.264/MPEG-4 AVC
  - Uses hierarchical-B pictures combined in inter-view and temporal dimensions
  - Reorganizes input images into a single stream prior to encoding



Inter-view-temporal prediction structure based on AVC, using hierarchical B pictures



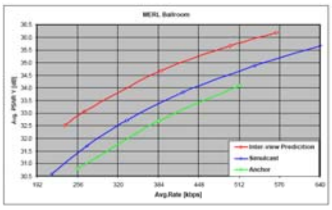
Associated reordering of multi-view input for compression with AVC

MVC Experiments

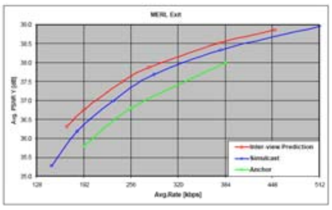
- ❖ Anchor (green curve)
  - AVC anchor coding results, as described in CfP
- ❖ Simulcast (blue curve)
  - Results using hierarchical B pictures only
- ❖ Inter-view Prediction (red curve)
  - Results of the MVC reference model combining inter-view prediction with hierarchical B pictures

Results: Ballroom/Exit

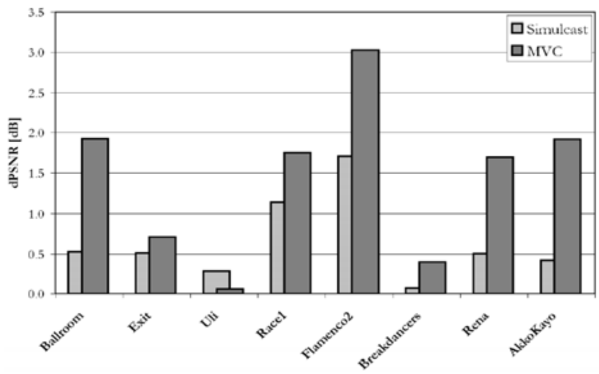
Ballroom



Exit

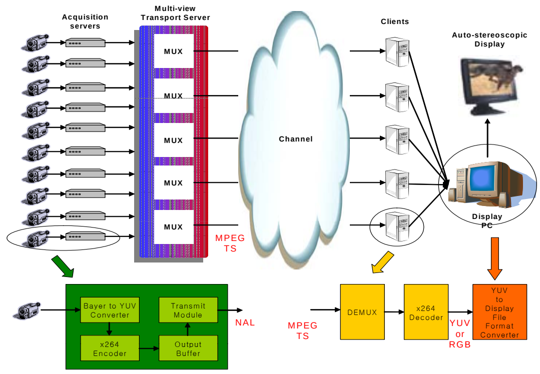


Summary: Coding Results



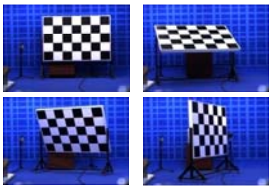
Multi-view Camera System

Multi-view Codec System



Multi-view Video Acquisition

Camera Calibration

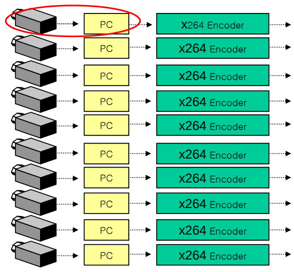


Extraction of Camera Parameters

Shot of 3D Broadcasting Content

Using Multi-view camera system

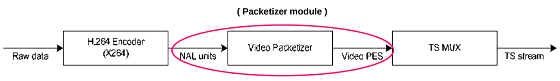
Coding of Multi-view Video



- Simulcast multi-view video coding
- Converting Bayer format to YUV format
- x264 coder for real-time compression
- Coding 10 frames/sec. (1024x768)

Transmission of Multi-view Video

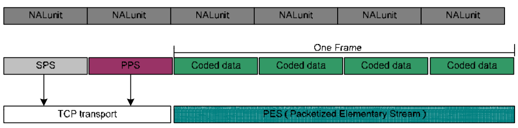
H.264/AVC over MPEG-2 system



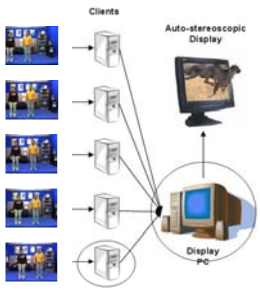
▪Packetizer module

(1) Parses the encoded H.264 data to the NAL units.

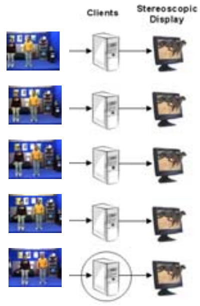
(2) Divides these based on their types. (Encode data, SPS, and PPS)



Display of Multi-view Video



Multi-view Auto-Stereoscopic Monitor



N Stereoscopic Monitors

Technical Problems

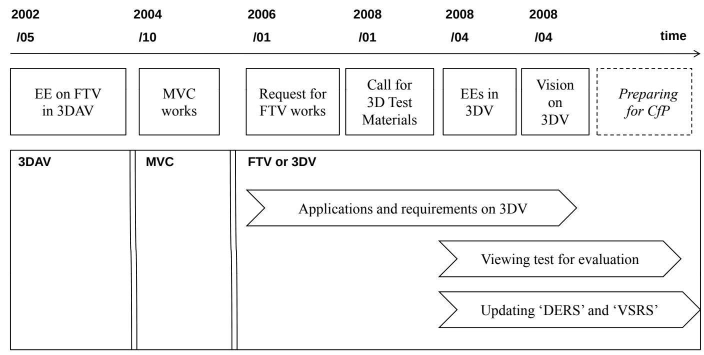
- Data acquisition
  - N cameras (N=2-128 or more)
  - How to calibrate multiple cameras?
  - How to translate and rotate each camera?
    - Elaborate camera control system is required (Hardware)
  - All the camera parameters should be stored
- Data size
  - A huge amount of data
  - Raw data rate with no compression (example)
    - VGA color video, 8 views, 30 fps, 10 sec.
    - 1024 x 768 x 3 bytes (R, G, and B) x 8 views x 30 fps x 10 sec. = 5,662 Gbytes
    - For 1 multi-view video, 8 CDs are required
  - If the image resolution is HD (1920 x 1080)
    - Over 14 Gbytes for a single 10 sec. multi-view video
    - More than 3 DVDs for storage only

Technical Problems

- Color/luminance inconsistency among multiple views

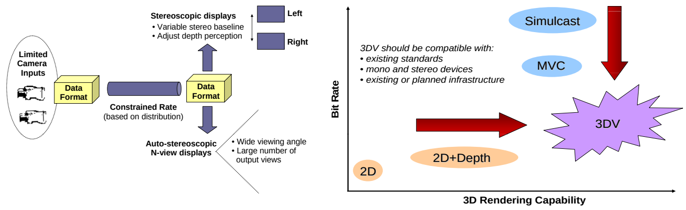
- Synchronization among multiple cameras (capture, display)
- Transmission of a huge amount of data
  - Is it possible to transmit raw multi-view videos in the current network?
- Real-time rendering
  - Is it possible to render multi-views/free-views in real-time?
  - Partial decoding and view /temporal/spatial random access are important
- Multi-view 3-D display devices
  - No glasses, simultaneous multiple view displays

History of 3DV/FTV



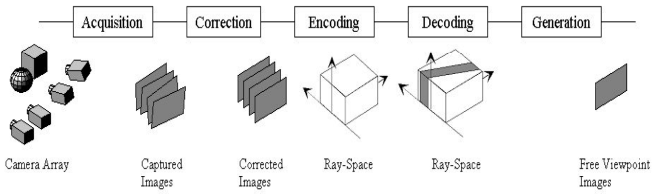
Vision on 3D Video

- Vision on 3D Video
  - To develop a new 3D video format
  - To support stereo or auto-stereoscopic displays
  - To make a standard for a new 3D vodeo codec within next two years.

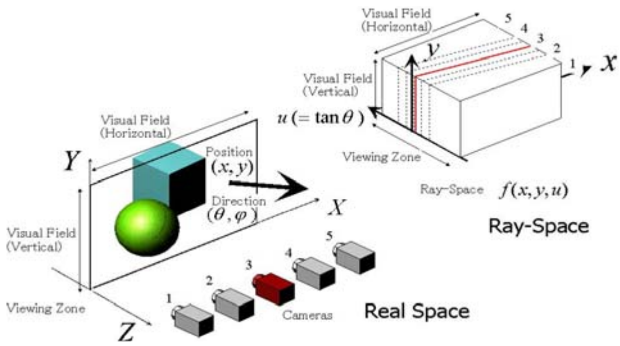


Free-viewpoint TV

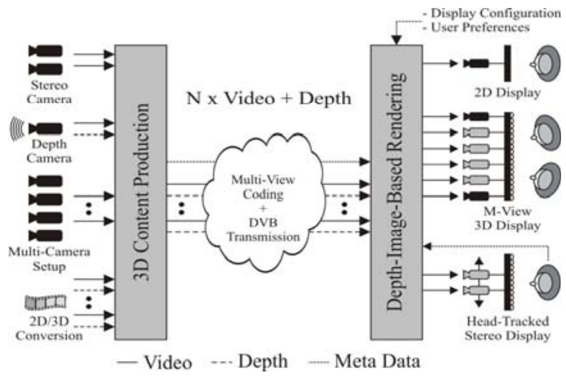
- FTV
  - Free viewpoint functionality
  - View generation for auto-stereoscopic displays
- FTV System



Ray-Space Representation

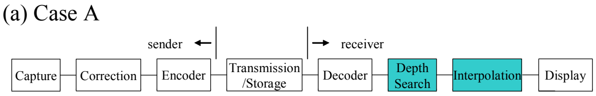


N Video + Depth

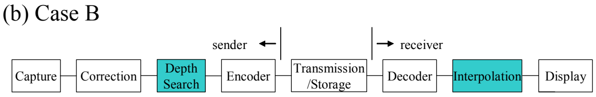


Depth Search and Interpolation

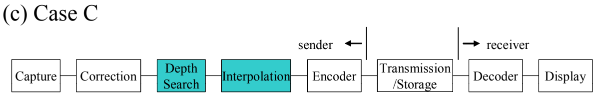
(a) Case A



(b) Case B



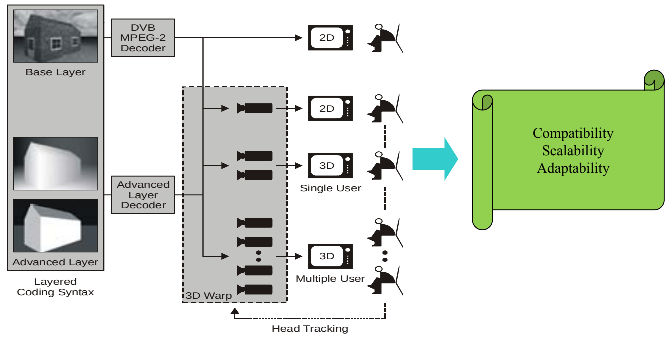
(c) Case C



Challenging Issues

- ❖ Depth/Disparity Estimation
  - Sub-pixel accuracy
  - Temporal enhancement to reduce flickering effects
  - Depth map refinement for distorted depth map
- ❖ Coding of Multi-view Video + Depth Map
  - Coding structure
  - Depth map coding scheme
  - Bit allocation for depth map coding
- ❖ Intermediate View Synthesis
  - View synthesis method for depth map distortion
  - Filtering along object boundaries

Broadcast 3-D Video

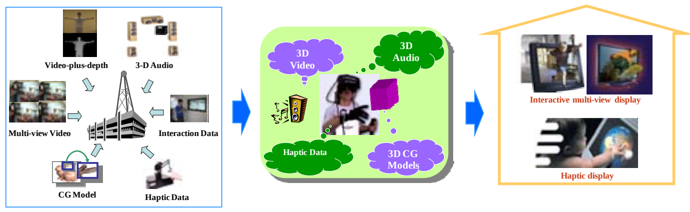


3-D Television System

Capture → Correction/ Conversion → Encoder → Transmission/ Storage → Decoder → Interpolation → Display

- ❖ 3-D video data format
- ❖ Capture module (camera calibration and rectification)
- ❖ Correction module (illumination and color correction)
- ❖ Encoder module
- ❖ Transmission module
- ❖ Decoder module
- ❖ Interpolation module
- ❖ Display module

Realistic Broadcasting Services



Immersive Multimedia | Immersive Media Processing | Interactive Services

Deliver natural, realistic and immersive experiences to users

Immersive Multimedia

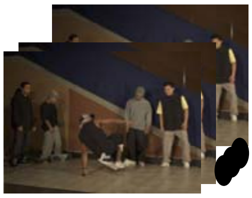
- ❖ Conventional 2-D audio-visual information
  - Not enough to give us vivid feeling through our five senses
- ❖ Multi-modal immersive media
  - Data overcoming the spatial-temporal limits
  - 3-D video, multi-channel audio, computer graphic data, haptic data
- ❖ 3-D video is an essential part of future multimedia services

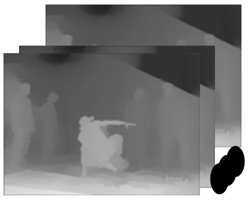
Requirements for Realism

- ❖ Multi-modal information
  - 3D video with depth information at multiple viewpoints
  - Multi-channel audio
  - CG model
  - Haptic data
- ❖ High-quality data
- ❖ User-friendly interaction

3D Depth Video | Multi-channel Audio | CG Models | Haptic Data

2-D Video + Depth


Color image sequence

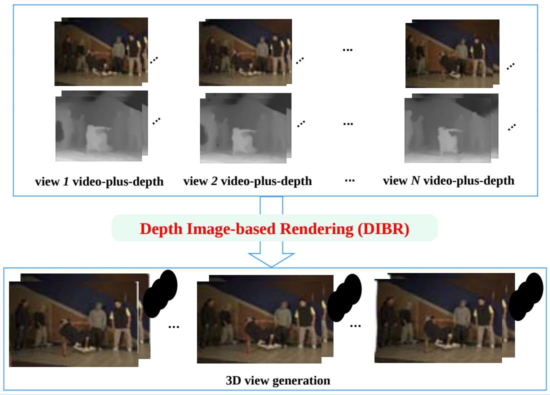

Depth map sequence

※ 'Breakdancers' sequence from MSR

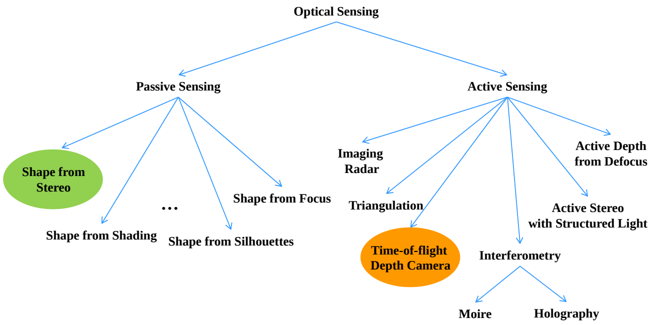
Monoscopic video with per-pixel depth information for 3-D video representation

Backward-compatibility and easy adaptability to the current 2-D digital system

Multi-view Video-plus-depth

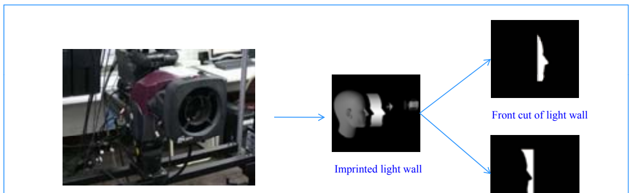


Taxonomy of Depth Estimation



※SIGGRAPH 2000: Course on 3D Photography

Time-of-flight Depth Camera

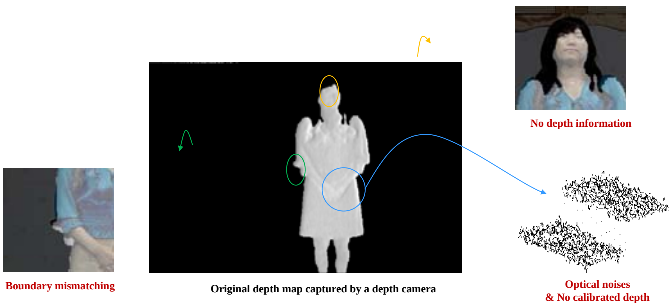


Z-Cam developed by 3DV System, Ltd

※ G.J. Iddan and G. Yahav, "3D Imaging in the Studio and Elsewhere…," *Proc. of SPUE Videometrics and Optical Methods for 3D Shape Measurements*, 2001.

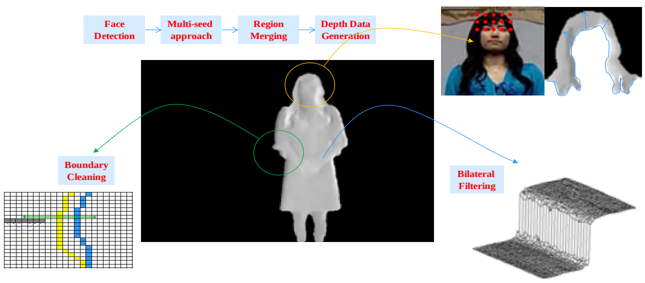
Real-time capturing of synchronized color and depth information

Problems in Depth Map



Original depth map captured by a depth camera

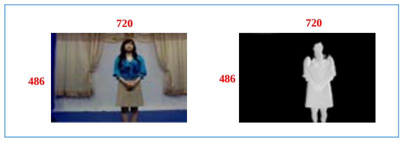
Depth Map Enhancement



Limitation in Depth Measurement

Low-resolution Depth Map


Z-Cam



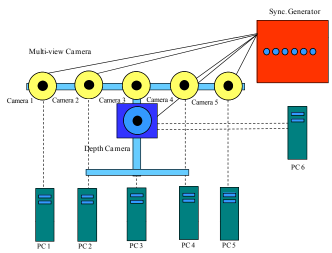

NHK Axi-vision HDTV Camera

※ M. Kawakita, T. Kurita, H. Kikuchi, and S. Inoue, "HDTV Axi-vision Camera," Proc. of International Broadcasting Conference, 2002

Hybrid Camera System (HCS)

Combination of both active and passive depth sensing

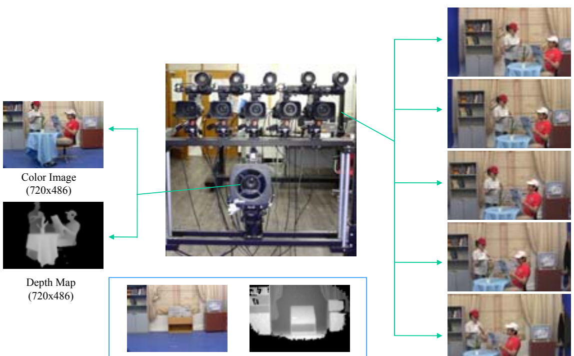
Specifications of HCS



Device	Specification	Details
Multi-view camera (Cannon XL-H1)	Output Format	NTSC or PAL (16:9 ratio, High Definition)
Depth Camera (Z-Cam)	Depth Range	0.5 to 7.0m
	Field of View	40 degrees
	Output Format	NTSC or PAL (4:3 ratio, Standard Definition)
Sync. Generator (LT 443D)	Output Signal	SD/HD Video Generation

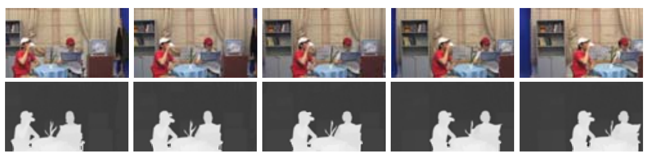
Baseline distance between two cameras: 20cm

Captured Images by HCS



Multi-view Image (1920x1080)

Output Images from HCS



Multi-view video-plus-depths (Image resolution 1920x1080)

Camera Calibration

- Finding out camera parameters
  - Relationship between 3-D object point and its 2-D image projection
  - Form a 3x4 projection matrix **P**
  - Homogeneous coordinate representation of points
  - Camera parameters
    - Intrinsic parameters: matrix **A**
    - Extrinsic parameters: matrix **R** and vector **t**

$$\tilde{\mathbf{m}} \equiv \mathbf{P}\tilde{\mathbf{M}} = \mathbf{A}[\mathbf{R}\,|\,\mathbf{t}]\tilde{\mathbf{M}}$$

Camera Intrinsic Parameters

- Represented by a 3x3 matrix **A**

$$\mathbf{A} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

α, β : Focal lengths in horizontal and vertical pixels

- Focal length: distance between the camera center and the image plane

$(u0,v0)$: Coordinate of the principal point

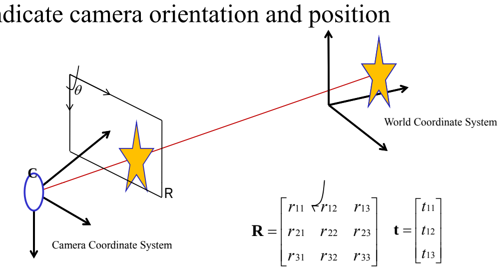
- Principal point: intersection between the Z-axis (principal axis) of the camera coordinate system and the image plane

γ : Skew parameter

- Non-orthogonality between u and v axes

Camera Extrinsic Parameters

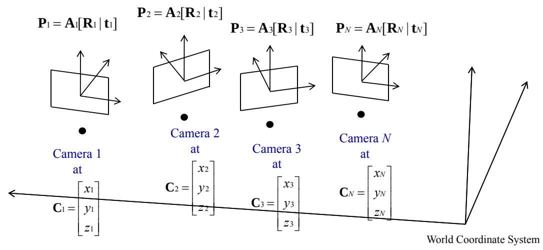
- Represented by
  - 3x3 rotation matrix **R**
  - Translation vector **t**
- Indicate camera orientation and position



$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad \mathbf{t} = \begin{bmatrix} t_{11} \\ t_{12} \\ t_{13} \end{bmatrix}$$

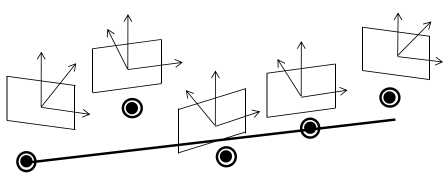
Camera Parameter Extraction

- Extraction of camera parameters from 2-D pattern images at each view



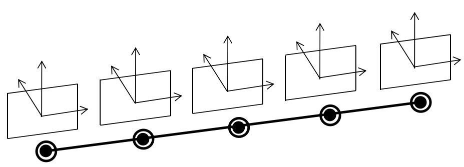
Problems of 1-D Parallel Camera Array

- Cameras are located around a certain line (Baseline)
- Unequal camera distances
- Different camera rotations
- Non-coplanar image planes



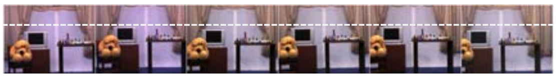
Ideal 1-D Parallel Camera Array

- Cameras are located on a common baseline
- Equal camera distances
- Coplanar image planes
- Perpendicular optical axes to the baseline

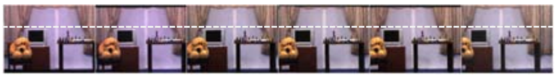


Multi-view Image Rectification

- Why Multi-view Image Rectification?
  - To compensate for non-ideal conditions
  - Non-ideal conditions are due to
    - Manual adjustment of multiple cameras
    - Hard to use mechanical instruments for camera alignment
  - Non-ideal conditions cause
    - High complexity in finding pixel correspondence or matching
    - Unclear viewpoints and viewpoint change



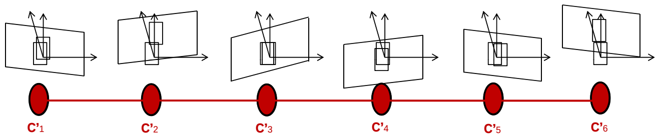
Before



After

Rectification Matrix

- Find rectification transformations
  - By resolving the relationship between the original and rectified camera parameters (or camera projection matrices)
  - Rectified camera parameters
    - Relocate camera centers
    - Rotate all the image planes
    - Adjust and equalize the intrinsic parameters



Original (Newspaper)

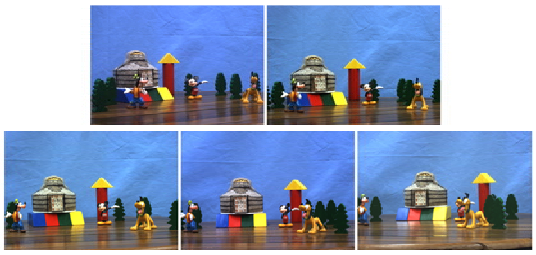
Rectified (Newspaper)

Multi-view Image Rectification

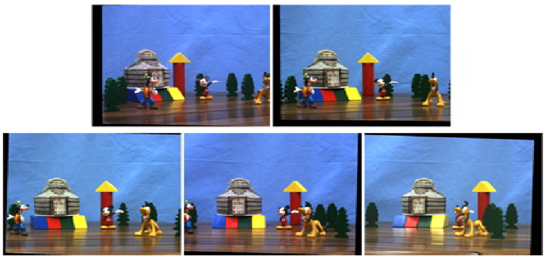
Before Rectification

After Rectification

Original (Toy)



Rectified (Toy)

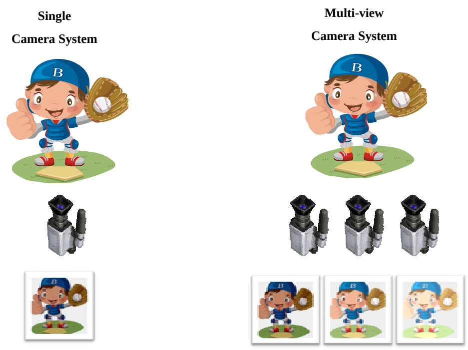


Multi-view Image Rectification

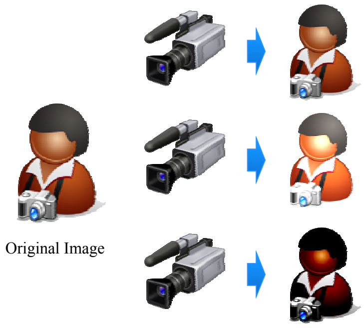
Before Rectification

After Rectification

Color Inconsistency Problem



Color Mismatch Problem



Camera Properties : Gain, Offset, Gamma, White Balance

Example of Color Mismatch

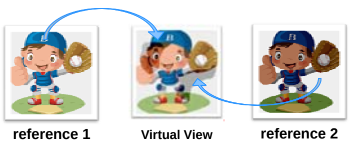
Race

Uli

Related Issues

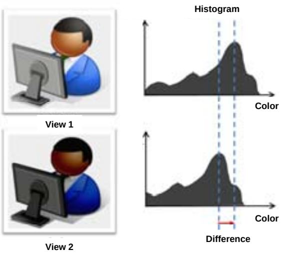
- Correspondence(compression, stereo matching)
  - Compare color values

- View Synthesis
  - Refer to neighboring images
  - Cause of stain

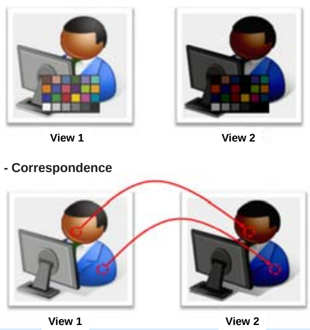


Correcting Methods

1. Global Property
   - Average brightness, histogram

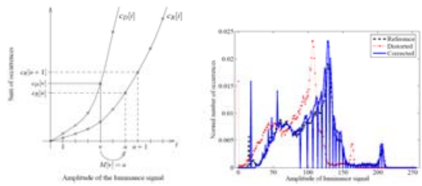


2. Corresponding Property
   - Color Chart



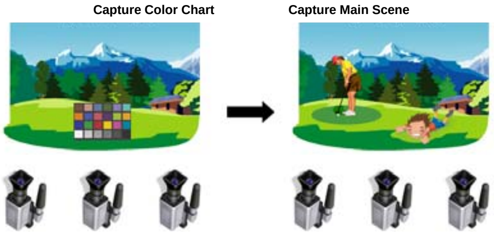
Histogram Matching

$M[v] = u \quad \text{with} \quad c_R[u] < c_D[v] \le c_R[u+1]$



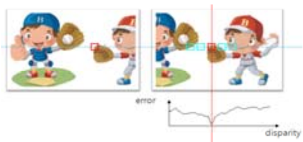
Color Chart

❖ Capturing the Standard Color Chart



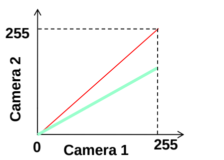
Correspondence

❖ Block-based Method
- Dense correspondence
- Low accuracy

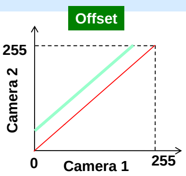


❖ Feature-based Method
- Sparse correspondence
- High accuracy

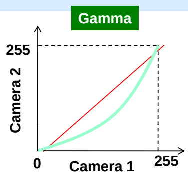
Camera Characteristic Curve



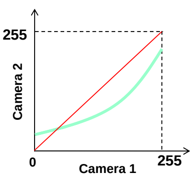
$P_{ref} = C_{gain} \times P_{tar}$



$P_{ref} = P_{tar} + C_{offset}$



$P_{ref} = \{P_{tar} / (2^{bitdepth} - 1)\}^{C_{gamma}} \times (2^{bitdepth} - 1)$



$$P_{ref} = \underbrace{C_{gain}}_{\text{Gain}} \{P_{tar} / (2^{bitdepth} - 1)\}^{\overbrace{C_{gamma}}^{\text{Gamma}}} \times (2^{bitdepth} - 1) + \underbrace{C_{offset}}_{\text{Offset}}$$

Corresponding Points

❖ Using SIFT Algorithm

View 3

R	G	B
147	146	155
116	107	72
131	134	149
138	140	146
115	114	129
65	44	39
178	177	180
59	71	93
59	71	93
91	92	92
...	...	...

View 4

R	G	B
152	155	161
169	124	88
141	148	159
152	155	161
113	116	127
31	27	25
193	193	192
52	63	84
52	63	84
99	100	99
...	...	...

Finding Coefficients

- Non-linear Regression
  - Define Error

$$e_i(\overline{\beta}) = y_i - \beta_0\{x_i/(2^{bitdepth}-1)\}^{\beta_2} \times (2^{bitdepth}-1) + \beta_1$$

$$E(\overline{\beta}) = \sum_m e_i^2(\overline{\beta})$$

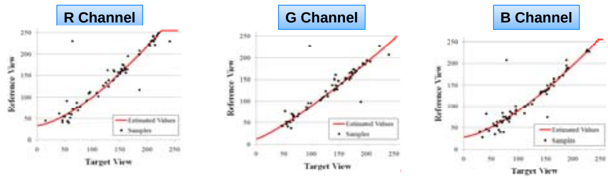
  - Iterate

$$\overline{\beta}^{s+1} = \overline{\beta}^{s} + \delta\overline{\beta}$$

$$(\overline{J}_{\overline{e}}^T \overline{J}_{\overline{e}})\delta\overline{\beta} = -\overline{J}_{\overline{e}}^T \overline{e}$$

  - Minimize Error

Results of Regression



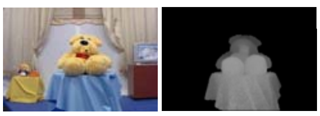
$$P_{ref} = C_{gain}\{P_{tar}/(2^{bitdepth}-1)\}^{C_{gamma}} \times (2^{bitdepth}-1) + C_{offset}$$

Coefficient	Red	Green	Blue
Gain	0.5456	0.1136	0.3702
Offset	13.9680	0.46	0.37
Gamma	0.42	0.39	0.38

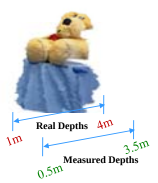
Experimental Result

Experimental Result

Why Depth Calibration?

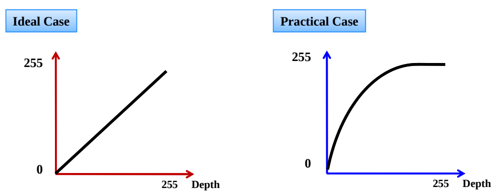


Color image and its depth map captured by Z-Cam

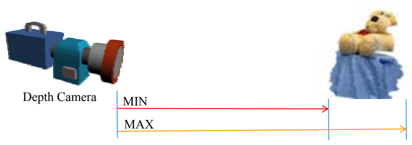


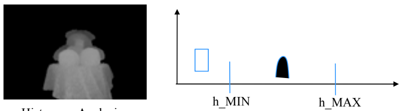
Mismatch between measured depths and real depths

Depth Calibration



Depth Calibration (Min, Max)



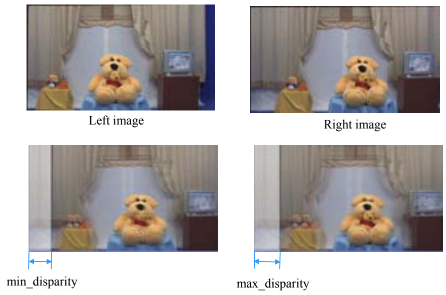


Histogram Analysis

d_MAX = 255- h_MIN

d_MIN = 255- h_MAX

Depth Map Calibration



$$Camera\_MAX = \frac{f \times B}{\min\_disparity}, Camera\_MIN = \frac{f \times B}{\max\_disparity}$$

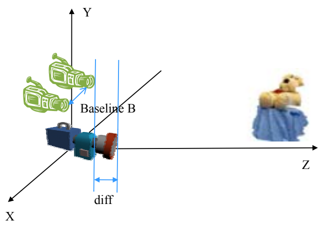
Depth Data Calibration

$$MAX = Camera\_MAX - diff, MIN = Camera\_MIN - diff$$

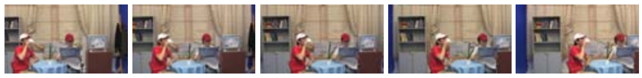
$$histo\_range = d\_MAX - d\_MIN$$

$$depth\_range = MAX - MIN$$

$$Z(i,j) = MIN + ((255 - d(i,j)) - d\_MIN) \times \frac{depth\_range}{histo\_range}$$



Color Segmentation



Rectified Views

Color Segmentation of Rectified Views

※D. Comaniciu and P. Meer, "Mean shift: A Robust Approach toward Feature Space Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002.

Relative Camera Calibration

Camera Calibration

$$P_s = K_s[R_s \mid t_s]$$
$$P_1 = K_1[R_1 \mid t_1]$$
$$\cdots$$
$$P_n = K_n[R_n \mid t_n]$$

Multi-view Image Rectification

$$\tilde{P}_1 = K_1'[R_1' \mid t_1]$$
$$\cdots$$
$$\tilde{P}_n = K_n'[R_n' \mid t_n]$$

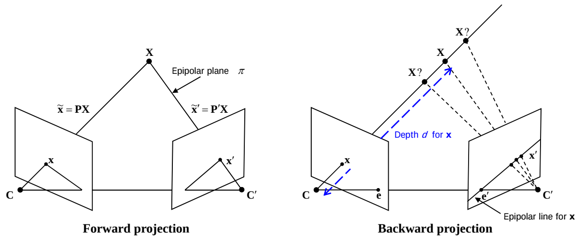
Relative Camera Calibration

$$P_s' = K_s[I \mid O]$$
$$\tilde{P}_1' = K_1'[\tilde{R}_1' \mid \tilde{t}_1']$$
$$\cdots$$
$$\tilde{P}_n' = K_n'[\tilde{R}_n' \mid \tilde{t}_n']$$

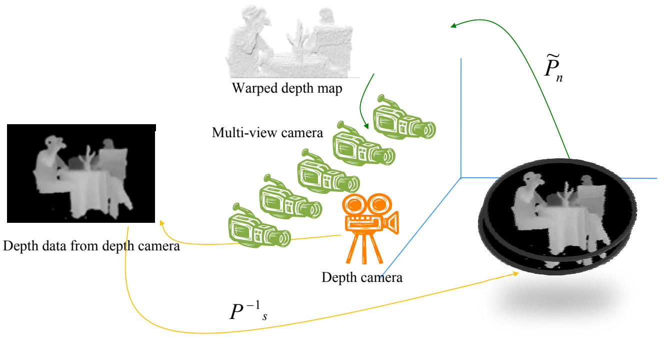
$$\tilde{R}_1' = R_1' R_s^{-1}, \tilde{t}_1' = t_1 - t_s$$
$$\cdots$$
$$\tilde{R}_n' = R_n' R_s^{-1}, \tilde{t}_n' = t_n - t_s$$

3-D Warping

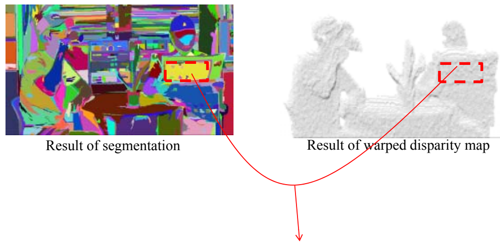
- Forward Projection
  - World -> Camera: $\tilde{\mathbf{x}} = \mathbf{PX}$
- Backward Projection
  - Camera -> World: $\mathbf{X} = \mathbf{R}^{-1} \cdot \mathbf{A}^{-1} \cdot \tilde{\mathbf{x}} \cdot \mathbf{D(x,y)} - \mathbf{R}^{-1} \cdot \mathbf{A}^{-1} \cdot \mathbf{T}$



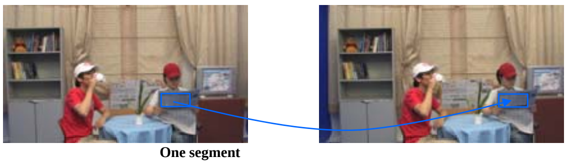
3-D Warping in HCS



Initial Disparity for Each Segment



Segment-based Stereo Matching

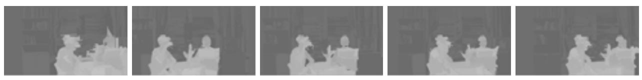


- Segmentation-based stereo matching using SAD
- InitialSAD from 3-D warping results
- MinDispSAD in a small search range

  Search range is set from (*initial disparity*-5) to (*initial disparity* +5)

Boundary Mismatch Problem

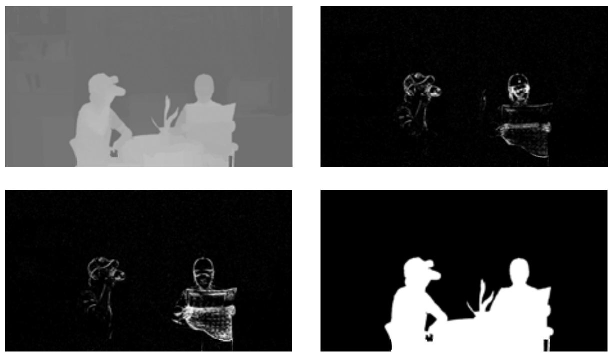
- In the boundary region of the foreground object
  - Boundary mismatch between the color image and its warped disparity map
  - Region separation: foreground, background, and unknown regions



Foreground/Background Separation

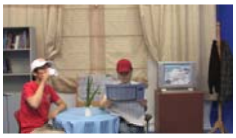
- Moving object detection
  - Color difference between frames
  - Modification of foreground boundaries

Result of Region Separation



3D Warping with Initial Disparity Map

- To reduce mismatched regions
  - Occlusion/disocclusion detection
  - Warping from the current camera to the reference camera with initial depth map
  - Re-warping from the reference camera to the target camera with the warped image

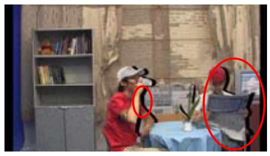

Camera 2


Warping from Camera 2 to Camera 3

Occlusion/Disocclusion Detection


Camera 3


Occlusion regions


Disocclusion regions

Temporal Consistency

- Temporal Consistency for Dynamic Scenes
  - Foreground depth
    - Recomputation of the disparity in a small search range
    - $f_d(x_i, y_i)$: Minimum SAD with the foreground initial depth
    - $w_d = 1, w_t = 0$
  - Background depth
    - Median disparity in the temporal domain
    - $f_t(x_i, y_i)$: Minimum SAD with the background initial depth
    - $w_d = 0.5, w_t = 0.5$

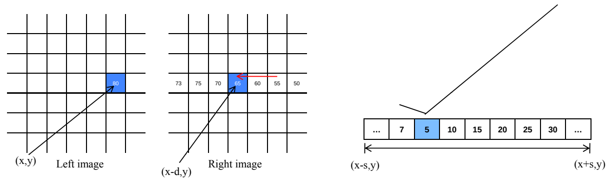
$$d = w_d f_d(x_i, y_i) + w_t f_t(x_i, y_i)$$

Refinement of Disparity Map

- Modified Belief Propagation (BP)
  - Modified BP
    - Search range restriction using initial depth information
  - Data cost energy function

$$d = \lambda min(|I_l(x,y) - I_r(x-d,y)|, \tau) \; , -s \le d \le +s$$

    - λ: scaling factor, τ: truncation value (threshold of cost)
    - s: search range



Final Disparity Map


Final disparity map of view 3

Final Multi-view Depth Map

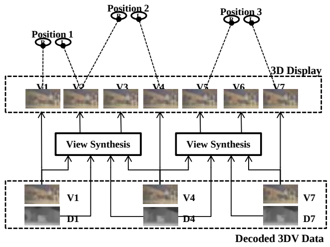

Final multi-view depth map from view 1 to view 5

Intermediate View Synthesis

- View Synthesis by 3-D Warping
- Disocclusion Problem and Hole Filling
- Intermediate View Generation

3D Rendering using View Synthesis

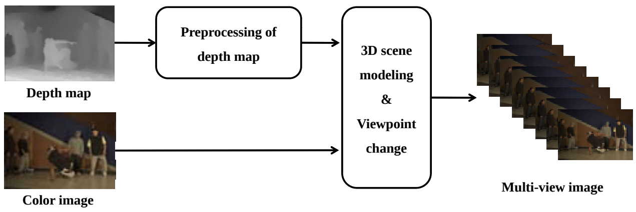
- Render Novel Views using View Synthesis
  - Generate virtual viewpoint images
  - Depth data is necessary for view synthesis



Depth Image-based Rendering

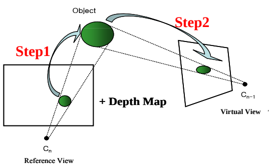
- Construct consecutive 3D video using texture images and depth maps
- Assumption: we have the depth information in advance

General Block Diagram of DIBR



3D Image Warping

- 3D Image Warping Using Depth Maps
- Two Operations
  - Project the pixels of the reference view into world coordinates
  - Re-project the image in the world coordinates into the virtual view



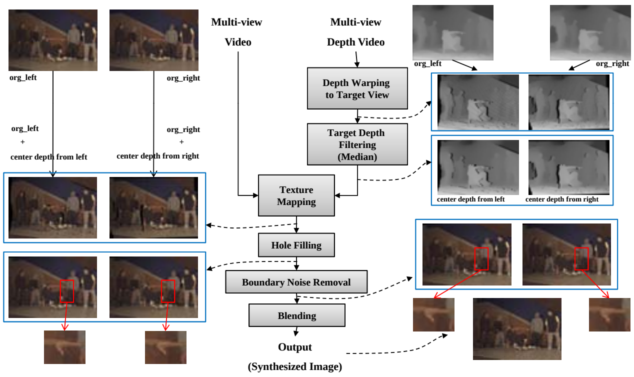
Step 1: Backward projection from image to world

$$M_r = R_r^{-1}\cdot A_r^{-1}\cdot m_r\cdot d(m_r) - R_r^{-1}\cdot t_r$$

Step 2: Re-projection from world to target world

$$\tilde{m}_t = A_t[R_t \mid t_t]\tilde{M}_r$$

Procedure of View Synthesis



Warped Image

❖ Depth-based 3D Warping
- Use warped depth image in virtual view

❖ Hole Area
- Newly Exposed Area (white)

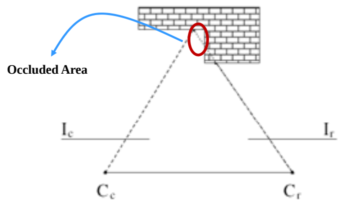

< Warped Images: 3D image warping left to right >

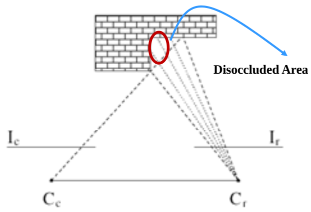

< Warped Images: 3D image warping right to left >

Occlusion and Disocclusion

❖ Occlusion
- Disappeared region, overlapped region

❖ Disocclusion
- Revealed region, exposed region





Small Hole Problem

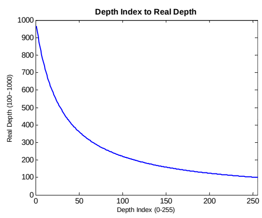
Real depths may not be perfectly represented by the depth map.

❖ Small holes are generated due to
- Truncation errors of converting floating point (real depth) to integer value
- Quantization errors of representing real depth with intensities in the depth map

Floating Point

Integer Value

$$Z = \frac{1.0}{\frac{d}{255.0}\cdot\left(\frac{1.0}{Z_{near}} - \frac{1.0}{Z_{far}}\right) + \frac{1}{Z_{far}}}$$



Hole Filling

❖ Holes can be filled with texture information from other views

❖ Replace the corresponding data
- Fill the holes with the region colored in white


< Corresponding Texture for the Right Side Hole >


< Corresponding Texture for the Left Side Hole >

Blending

❖ Blend Two Synthesized Images to one final image


I_L :Left view

View Synthesis


I_R :Right view

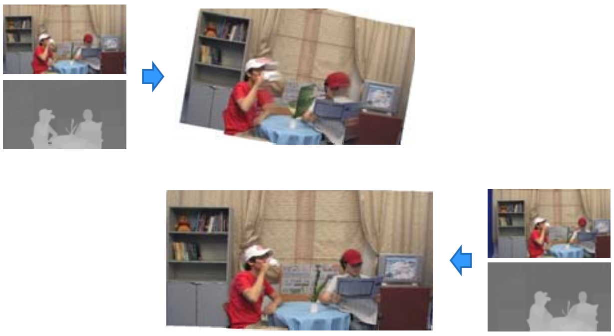
$$A_V = \alpha\cdot A_L + (1-\alpha)\cdot A_R$$
$$R_V = \alpha\cdot R_L + (1-\alpha)\cdot R_R$$
$$t_V = \alpha\cdot t_L + (1-\alpha)\cdot t_R$$

$$\begin{cases} 0 \le \alpha \le 1 \\ \alpha = \dfrac{B_R}{B_L + B_R} \end{cases} \quad \begin{cases} B_R = |t_V - t_R| \\ B_L = |t_V - t_L| \end{cases}$$

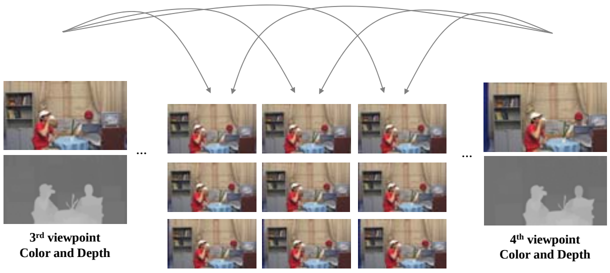
Experimental Result

❖ Demo: 6 reference views and 45 additional views

3-D Scene Rendering



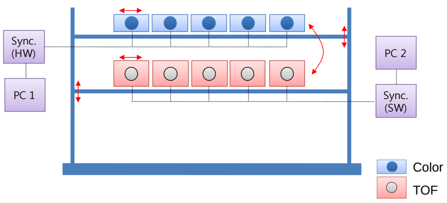
Generated Intermediate Views



Intermediate images at virtual viewpoints

Multiple Depth Camera System

- System Configuration
  - 5 HD cameras + 5 TOF cameras
  - Height: 50cm to 160cm
  - Width: 120cm
  - Min. distance between two adjacent cameras: 6.5cm



Multiple Depth Camera System

- Camera Specification

Basler Pioneer piA I1900-32gm/gc

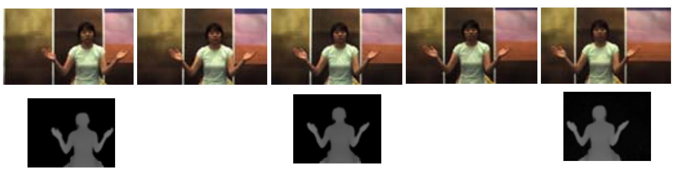
Resolution: 1920x1080 (HD)
Frame rate: up to 32
Width & height: 4.5cm x 3cm

Mesa Imaging SR4000

Resolution: 176x144 (QCIF)
Frame rate: up to 54
Operating range: 0.3 ~ 5.0 meters
Depth & intensity images
Width & height: 6.5cm x 6.5cm

Captured Images

- 5 Color images
- 3 Depth images
  - Viewpoint #1, #3, #5
  - We cannot capture more than three views simultaneously due to the available TOF frequencies (29/30/31 MHz).

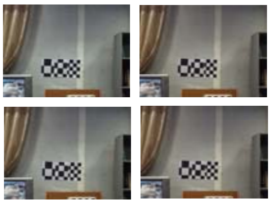


Synchronization Problem

- Multi-camera Synchronization
  - HD cameras: Synchronized with the pulse generator
  - TOF cameras: Synchronized with the capturing software
  - No synchronization instrument or software available for HD cameras and TOF depth cameras

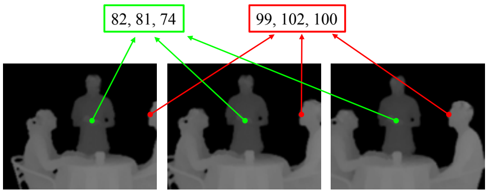
Illumination Flickering

❖ Illumination Flickering
- Caught by Basler Pylon HD camera
- Affect color correction and stereo matching operations
- Deteriorate visual quality of processed images
- Require electrically more stable illumination conditions



Inconsistency of Depth Values

❖ Different depth values for the same object
❖ Capturing using 3 different modulation frequencies
- 29MHz (Maximum range: 5.17m)
- 30MHz (Maximum range: 5.00m)
- 31MHz (Maximum range: 4.84m)

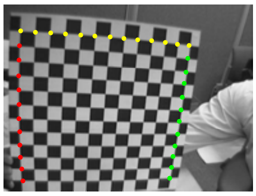


Radial Distortion

❖ Radial Distortion of TOF Camera
- Intrinsic camera problem
- Affect feature-based image processing, such as camera calibration and depth calibration
- Shape mismatch between color and depth images

Correction of Radial Distortion

❖ Radial Distortion Correction
- Find line components in the image
- Determine the distortion center
- Estimate distortion parameters
- Correct the radial distortion in the image



A. Wang, et al. "A Simple Method of Radial Distortion Correction with Centre of Distortion Estimation," Journal of Mathematical Imaging and Computer Vision, Jul. 2009.

Radial Distortion Correction

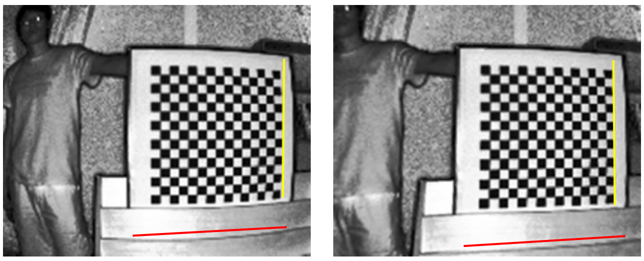


Image Resolution Difference

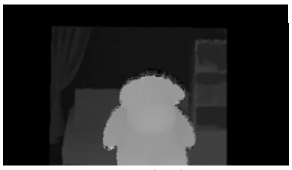
❖ HD vs. QCIF
- Depth image is warped to the color image position
- Depth image cannot cover the whole scene

Depth Image Warping

❖ Depth Image Warping

- 3D warping from TOF camera position to HD camera position
- Perform interpolation in the hole region

Warped depth



Interpolated

Image Resolution Adjustment

❖ Result of Depth Image Warping

- Mismatch between the color and warped depth images
  - Errors in camera parameters
  - Nonlinear depth values
  - Shape mismatch due to the capturing position

Conclusions

❖ MPEG Activities on 3D Video Coding

- 3D Audio Visual (3DAV)
- Multi-view Video Coding (MVC)
- 3D Video Coding (3DV/FTV)

❖ Challenging Issues for 3D Video

- Capturing of Multi-view Images
  - Camera Calibration and Rectification
  - Illumination Compensation and Color Correction
  - Synchronization and Registration
- Depth/Disparity Estimation
- Coding of Multi-view Video + Depth Map
- Intermediate Virtual View Synthesis

Acknowledgments

❖ MPEG/JVT Contributions

❖ GIST RBRC Members

❖ GIST VCL Members

Thank you for your attention!

Any questions?

Prof. Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST)

Realistic Broadcasting Research Center (RBRC)

+82-62-970-2211

hoyo@gist.ac.kr

http://rbrc.gist.ac.kr/

Contact Information

Prof. Yo-Sung Ho

- +82-62-970-2211, +82-10-3162-3669
- hoyo@gist.ac.kr
- http://vclab.gist.ac.kr/

Realistic Broadcasting Research Center (RBRC)

- +82-62-970-3131
- http://rbrc.gist.ac.kr/

Gwangju Institute of Science and Technology (GIST)

- http://www.gist.ac.kr/