3D Video Processing Techniques for AR/VR Contents Generation

December 12th, 2017

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 IEEE Fellow, SPIE, IEEK, KICS, KSBE
- 2004~2011 Associate Editor of IEEE Trans. on CSVT
- 1991~Now Delegate to MPEG/JCT-3V Meetings
- 1999~Now Tutorial Lecturer at Various International Conferences

Contents

3D Video for AR/VR Systems

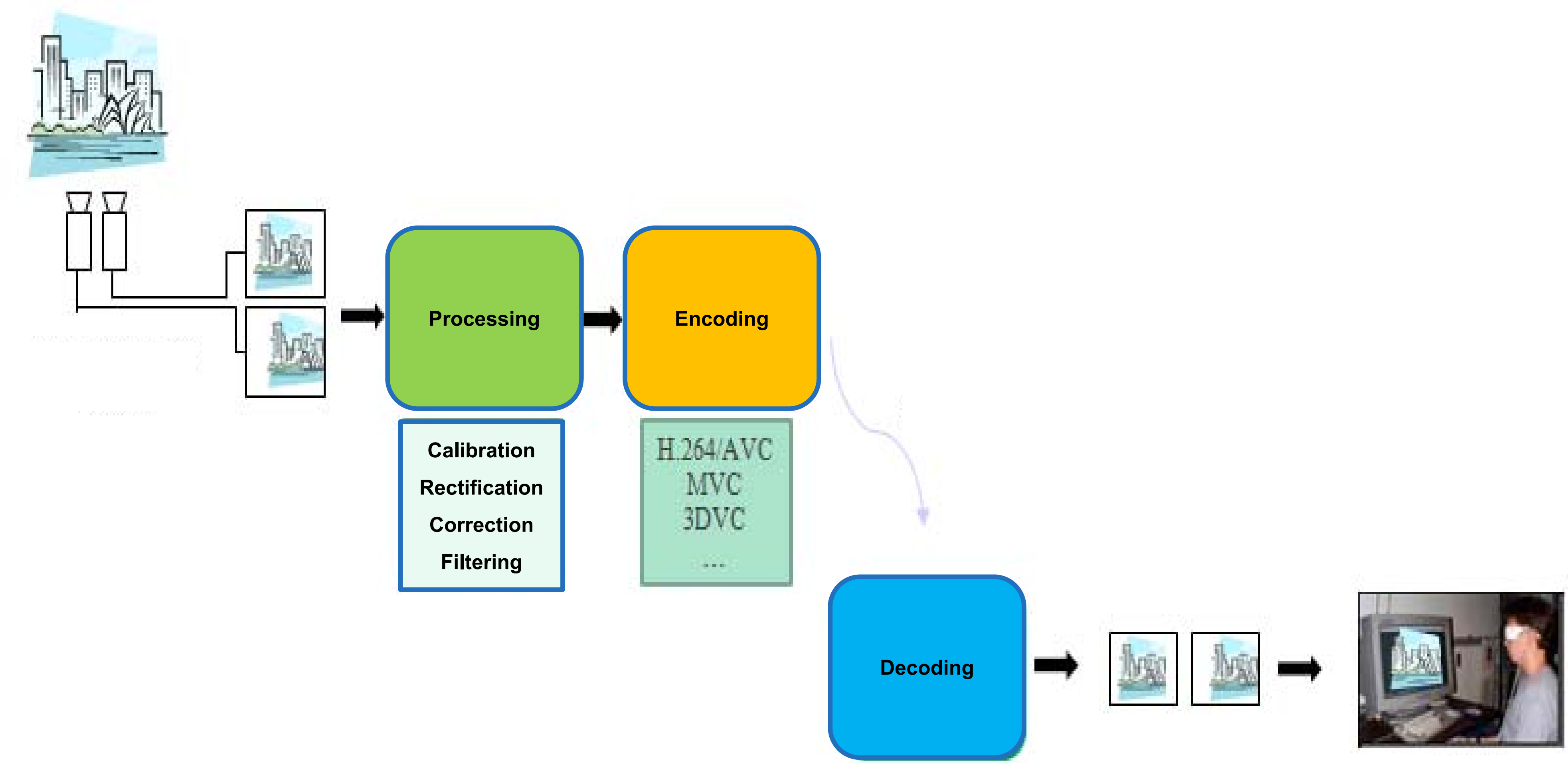
❖ Realistic 3D Contents

- Sports, game, dance, education
- Architecture, construction, military
- Healthcare, entertainment, fashion
- Heritage, business, engineering, film
- Scientific visualization, media

❖ Various AR/VR Devices

- Head mounted display (HMD)
- Multi-view 3D display
- Pseudo holographic display

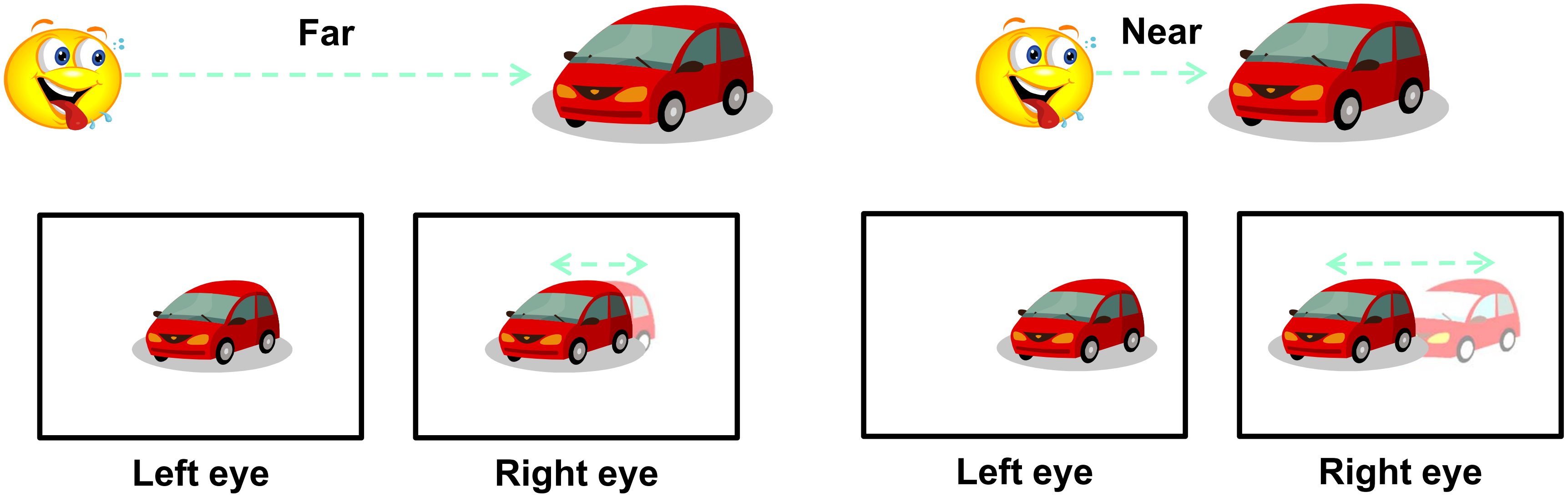
3DTV System


Processing
Encoding
Calibration
Rectification
Correction
Filtering
H.264/AVC
MVC
3DVC
...
Decoding

Stereoscopic Cameras

Binocular Cue

- **Binocular cue (disparity)**
  - Far object => small disparity
  - Near object => large disparity

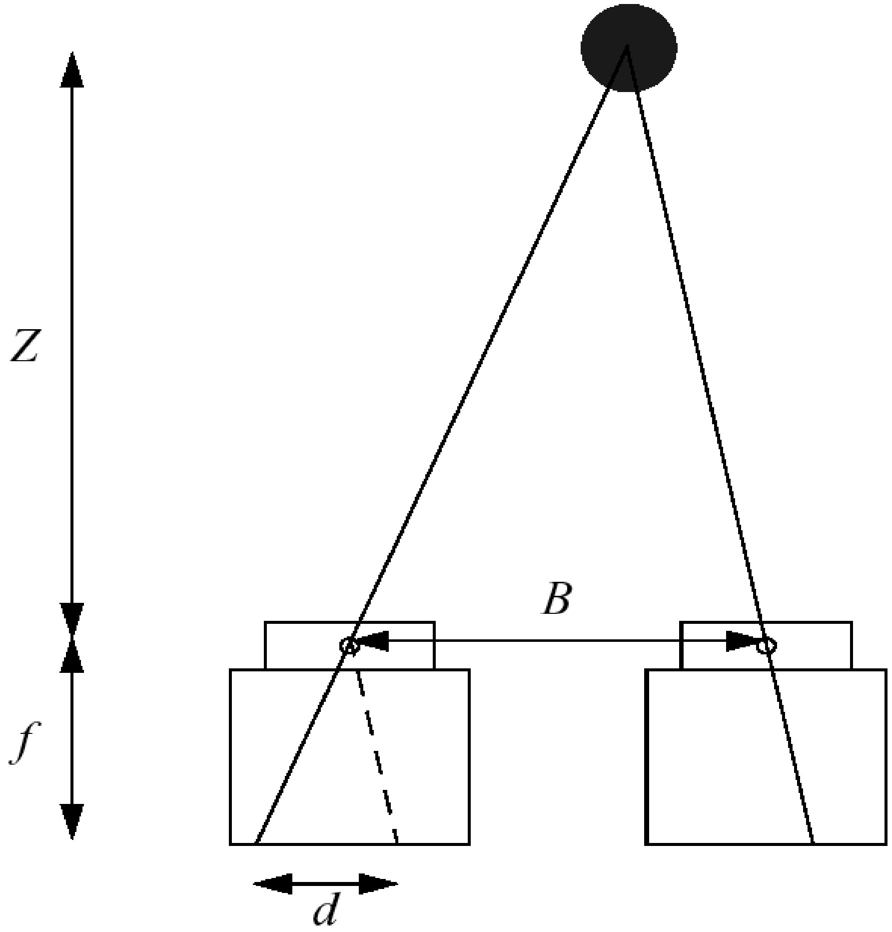


Disparity and Depth

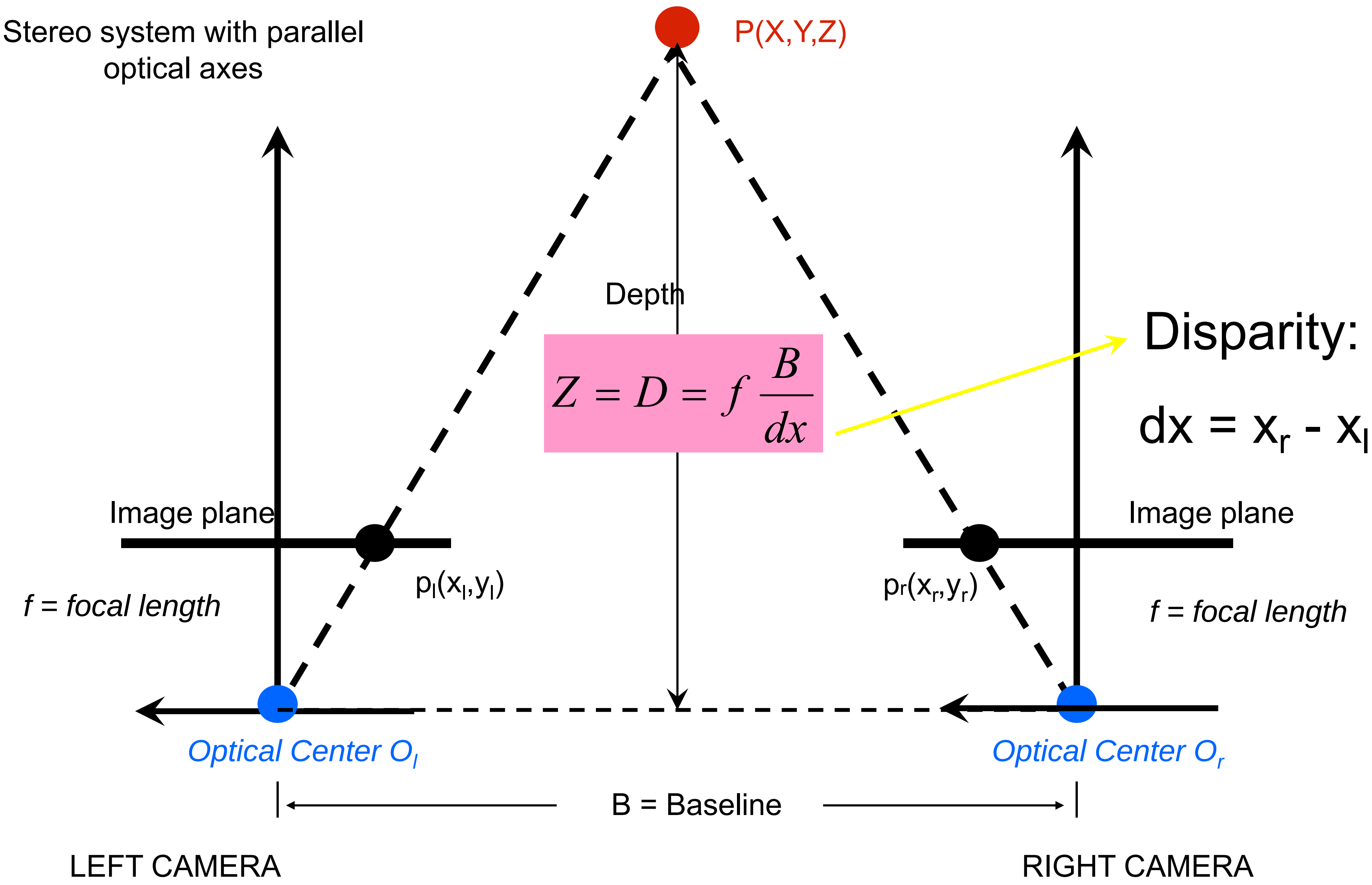
- If the cameras are pointing in the same direction

  (simple geometry)
  - *B*: *baseline* of the camera system
  - *Z*: *depth* of the object
  - *d*: disparity (left *x* minus right *x*)
  - *f*: focal length of the cameras
- Unknown depth is given by

$$Z = \frac{f\,b}{d}$$



Disparity Equation



Stereo Matching

- **Stereo matchers start from some assumptions**
  - Corresponding image regions are similar
  - A point in one image may only match a single point in the other image
  - If two matched features are close together in the images, then in most cases their disparities will be similar, because the environment is made of continuous surfaces separated by boundaries.
- **Many matching methods exist**
  - Feature-based methods which start from image structure extracted by preprocessing; and
  - Correlation-based methods which start from individual grey-levels

Stereo Constraints

- **Color constancy**
  - Color of any world points remains constant from image to image
  - This assumption is true under *Lambertian* Model
  - In practice, given photometric camera calibration and typical scenes, color constancy holds well enough for most stereo algorithms.

- **Uniqueness and continuity**
  - Each item from each image may be assigned at most one disparity value
  - Disparity varies smoothly almost everywhere

Stereo Matching Methods

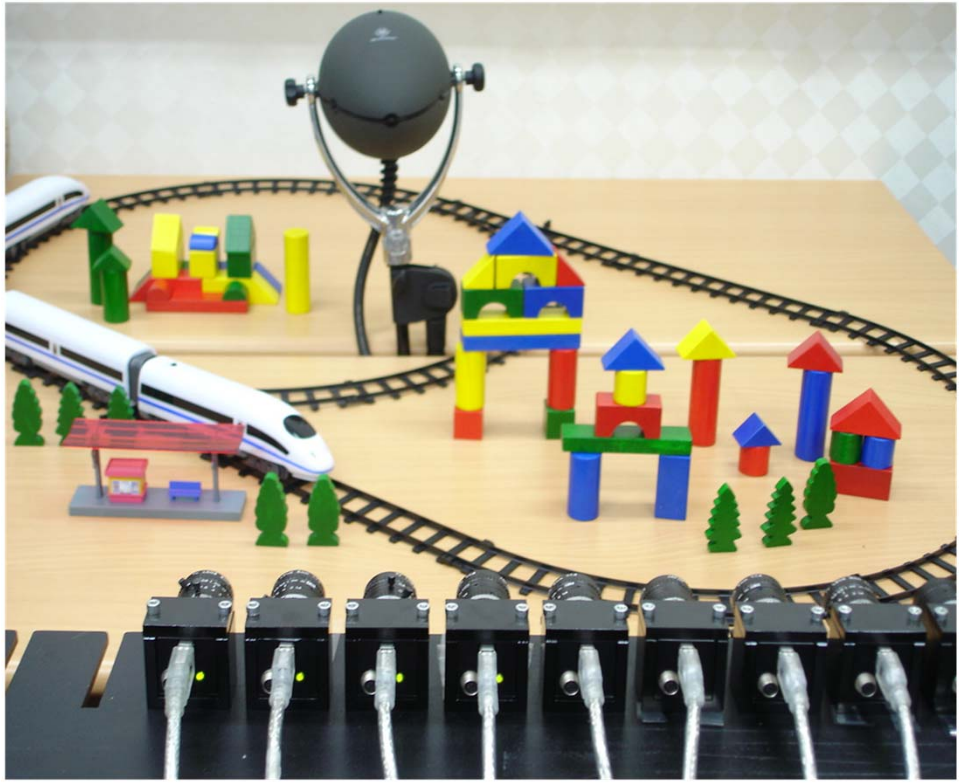
❖ Correlation approach

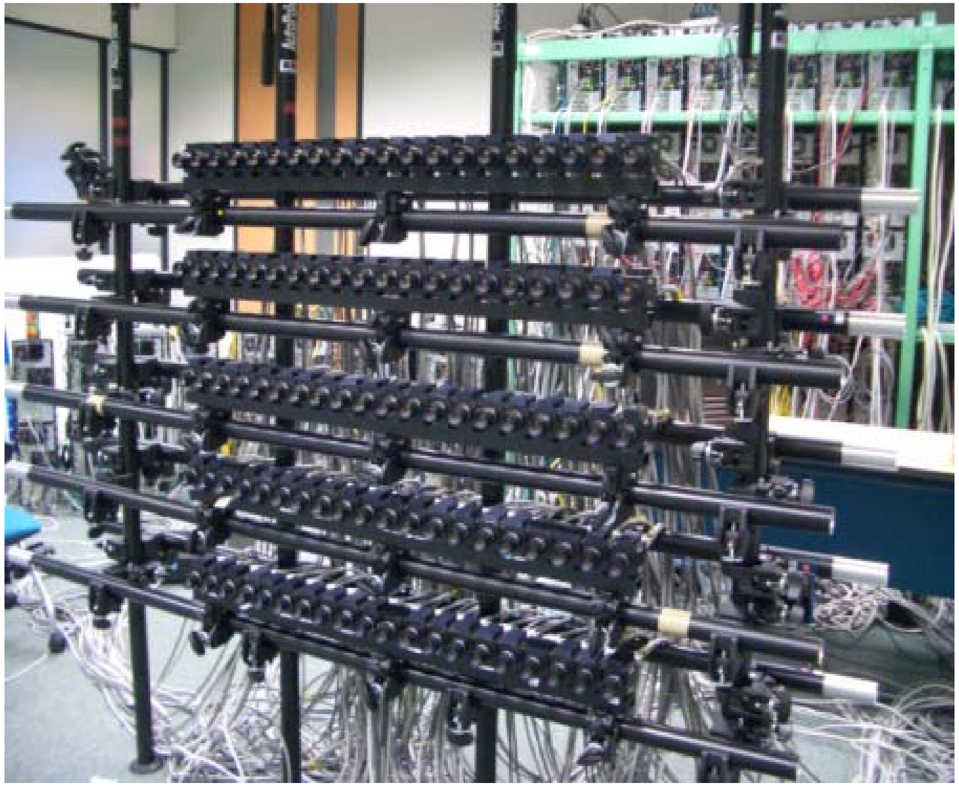
- A measure of similarity between windows in two images
- The corresponding element is given by window that maximizes the similarity criterion within a search region

❖ Feature-based approach

- Extract features in the stereo pair
- Define similarity measure
- Search correspondences using similarity measure and the epipolar geometry

Multi-view Cameras

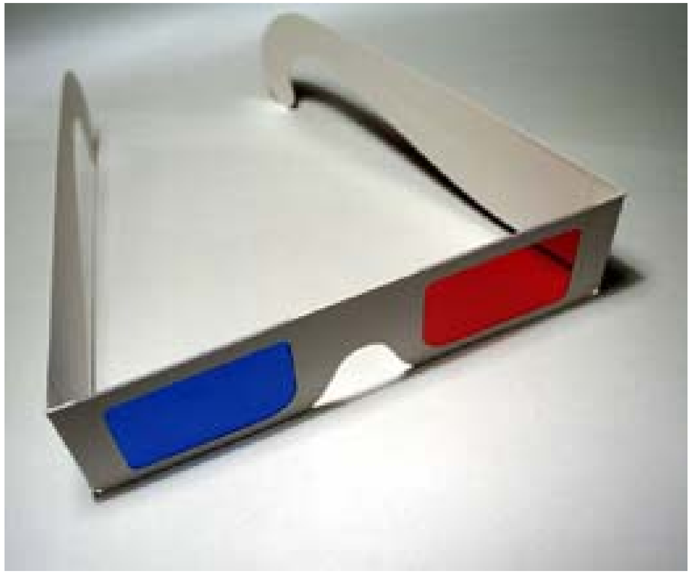




Multi-view 2D + Depth Cameras

3D Display Monitors

Anaglyph



Polarization

Light Passing Through Crossed Polarizers

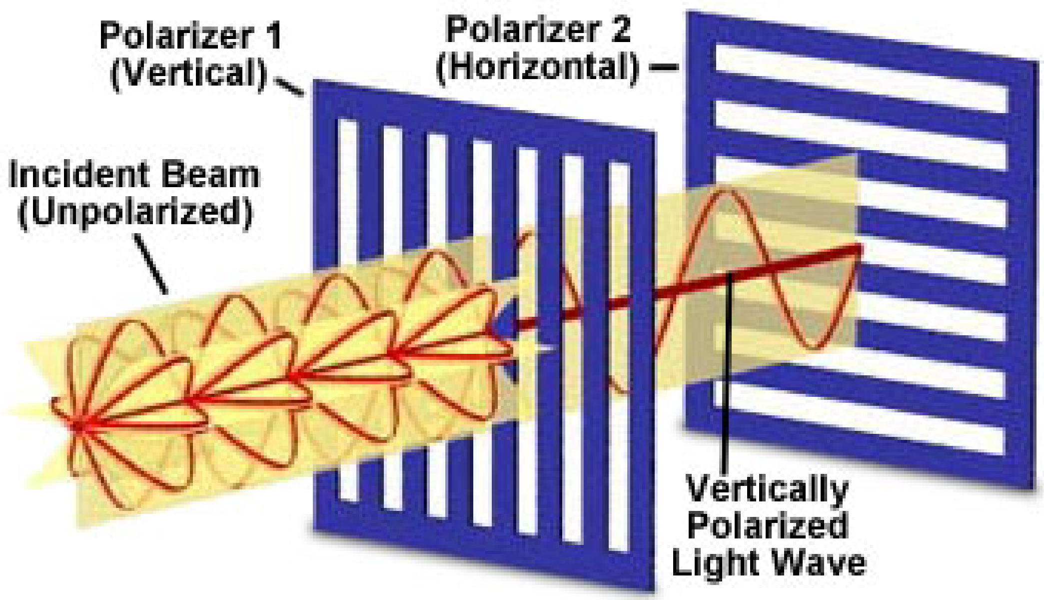
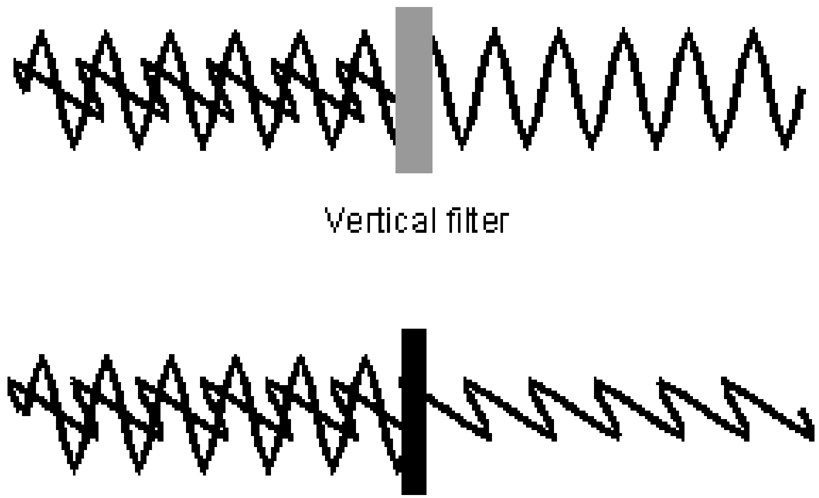
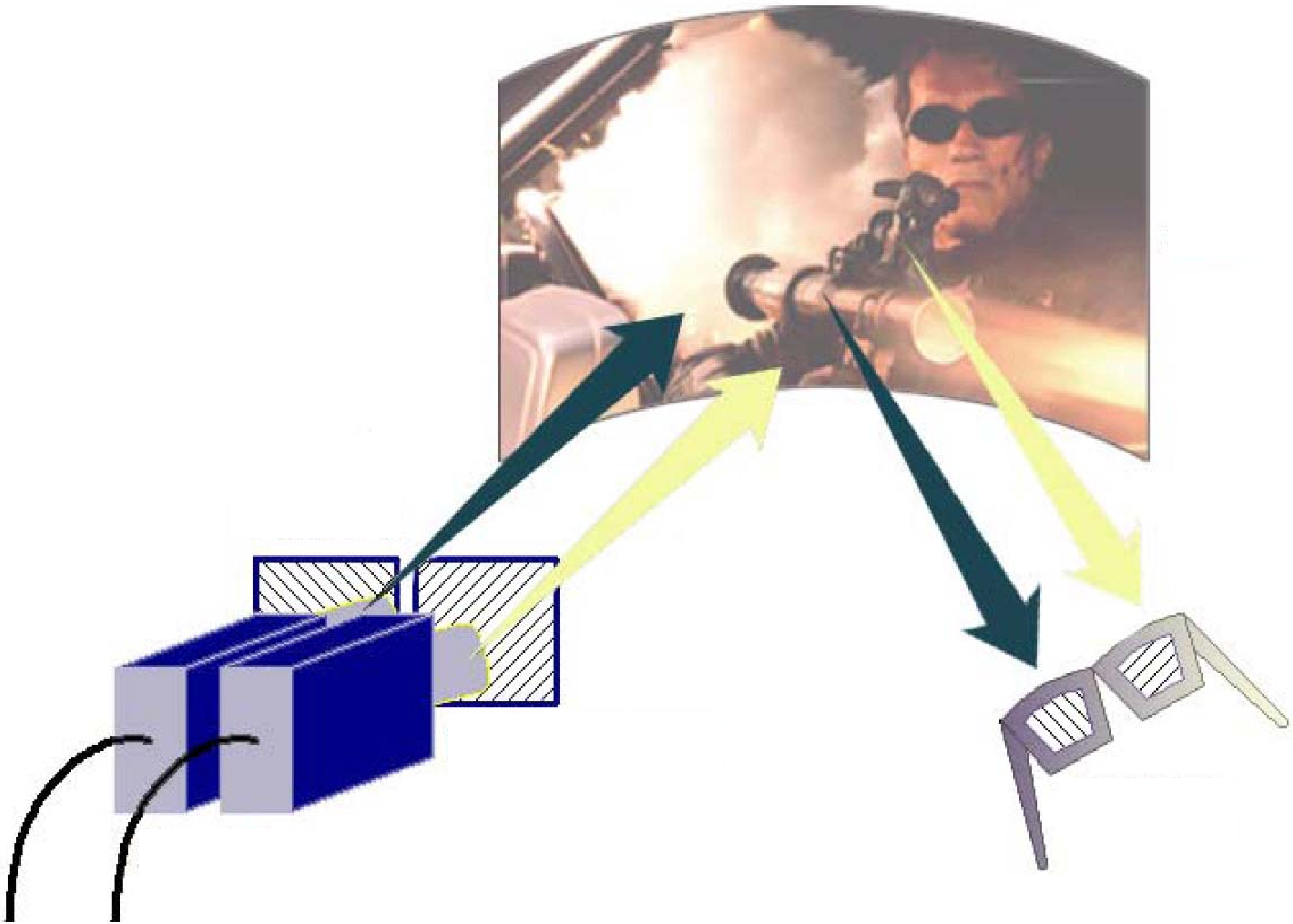


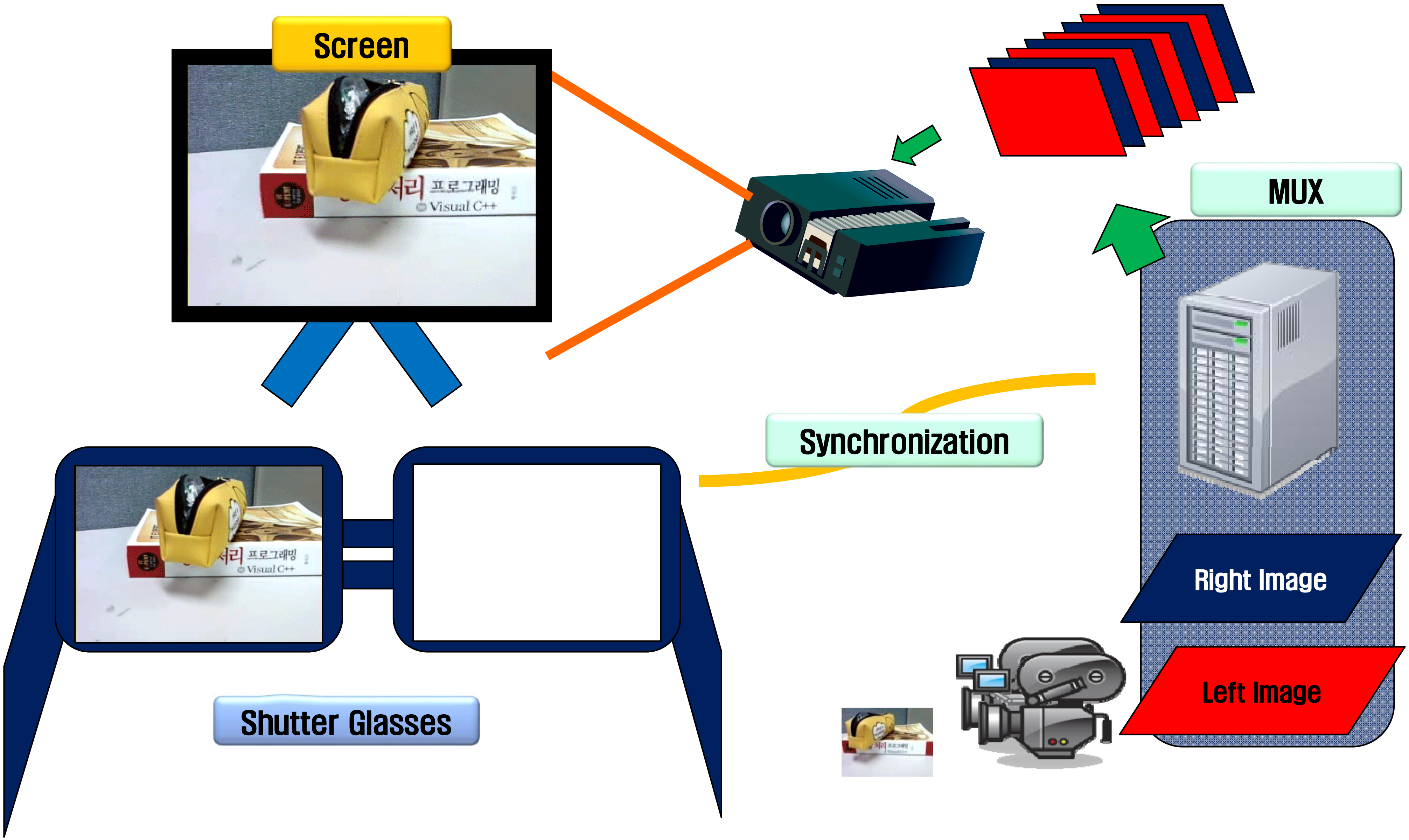
Figure 1



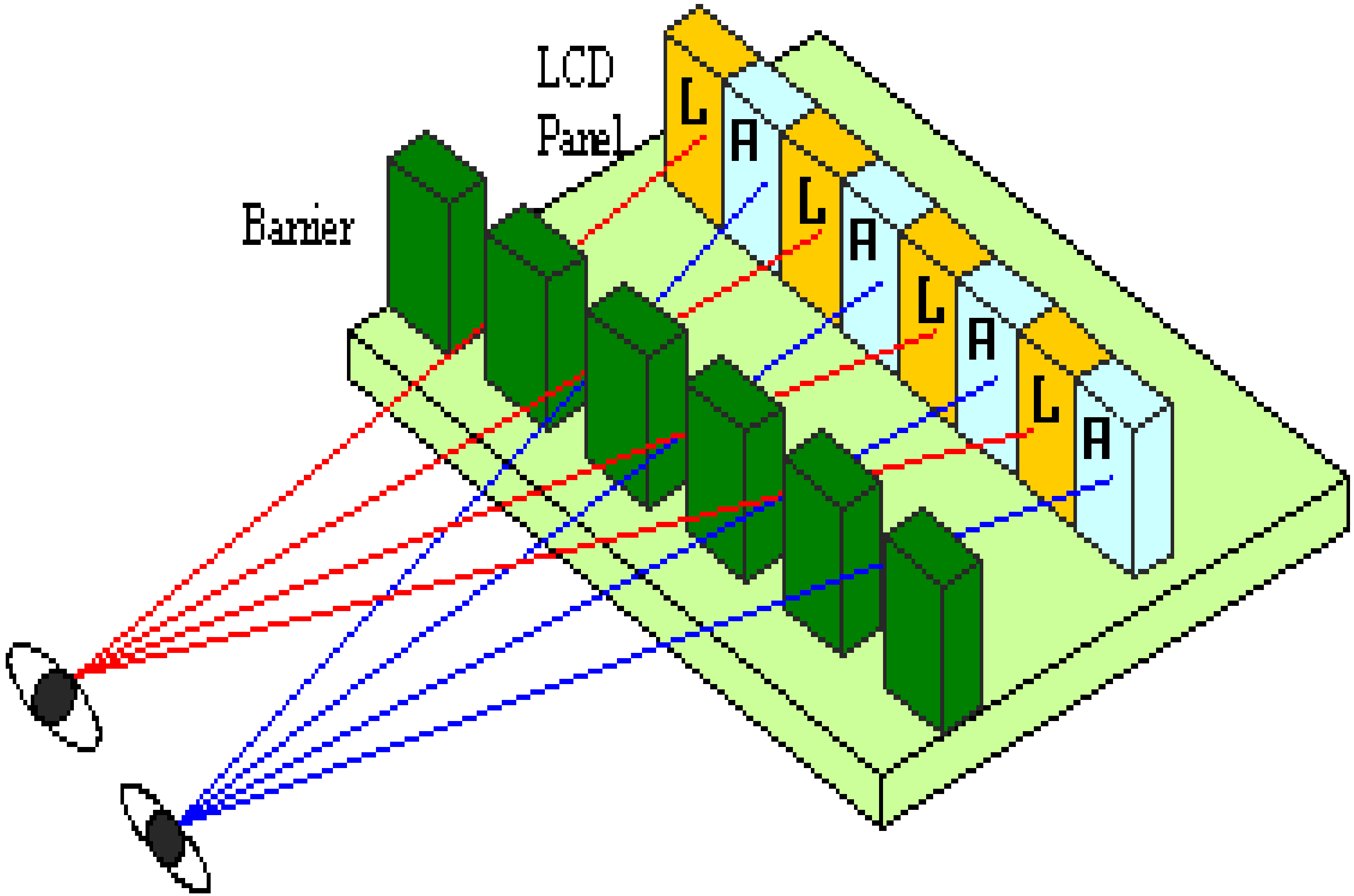
Polarized Glasses



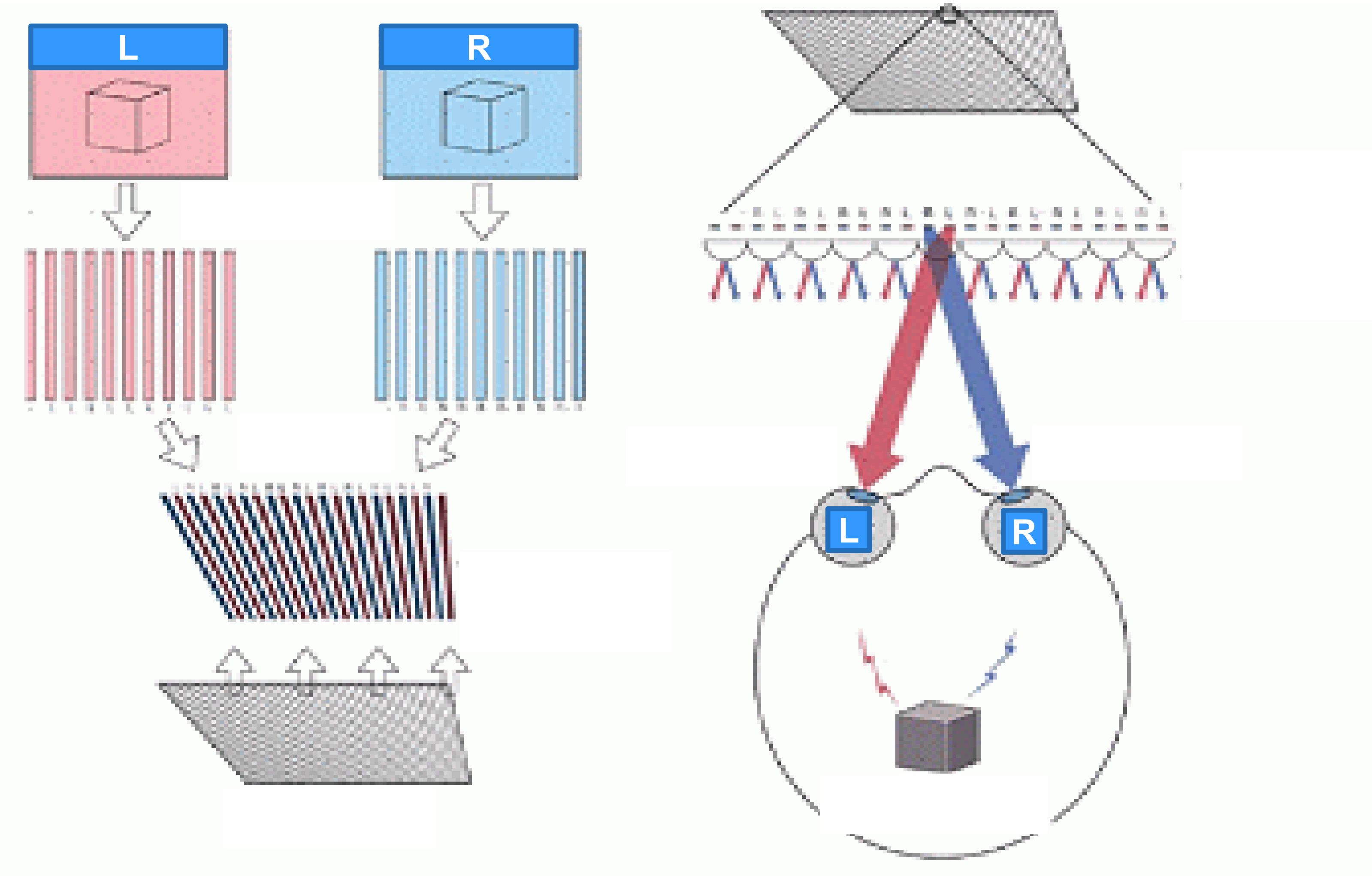
Shutter Glasses



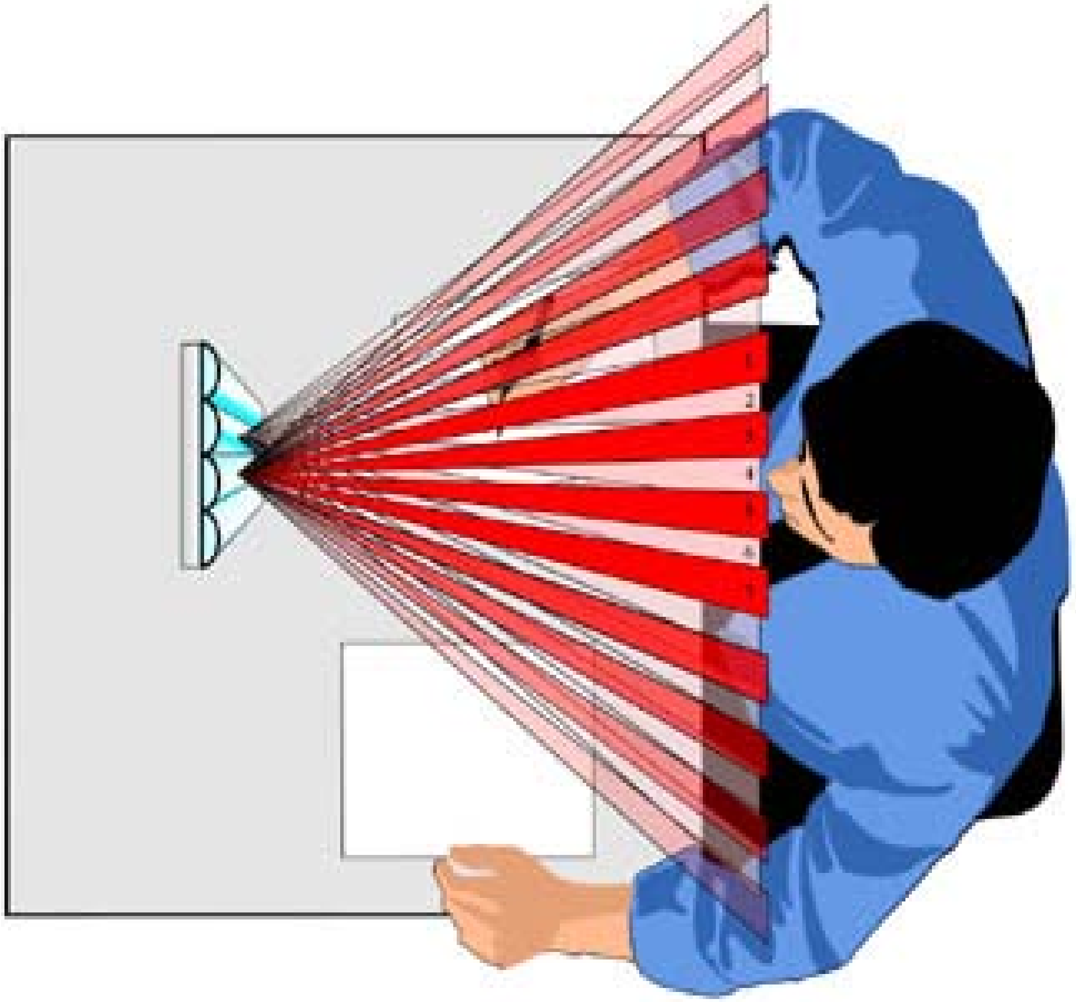
Parallax Barrier



Lenticular Sheet



Auto-stereoscopic 3D Display



3D Image Formats

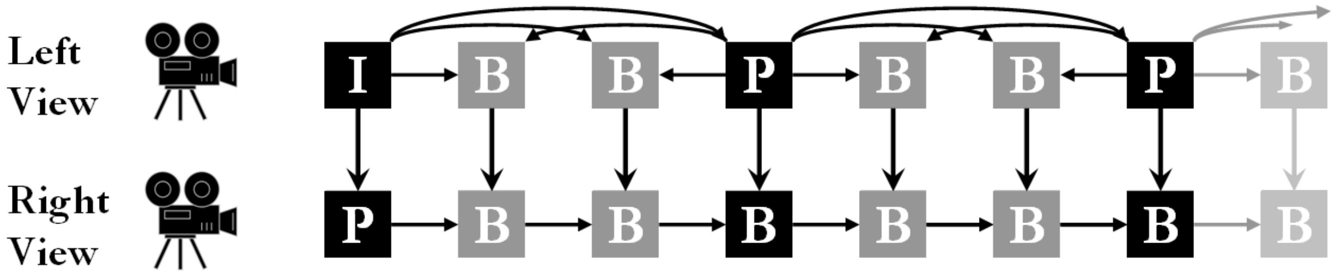
- Stereoscopic or Multi-view Simulcast
- Multi-view Video Coding (MVC)
- Frame-Compatible Stereo Formats
- Depth-based Image Formats
  - 2D Video + Depth (2VD)
  - Multi-view Video + Depth (MVD)

Multi-view Simulcast

- Independent encoding of each view
  - No attempt to exploit redundancy between views
- We can use different video coding schemes:
  - MPEG-2
  - MPEG-4
  - H.264/AVC
  - HEVC/H.265

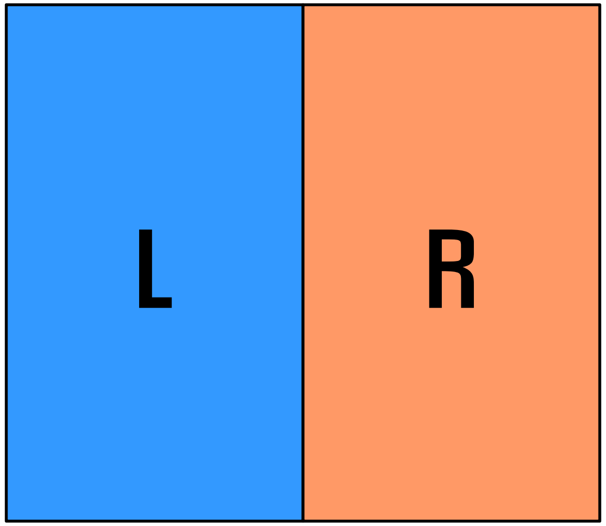
Multi-view Video Coding

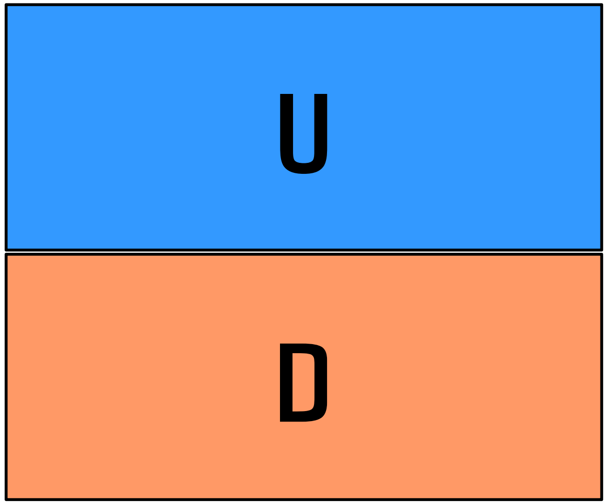
- AVC extension for multi-view video coding
  - Stereo is an important special case
  - Maintain the full resolution for all views
- Technical features
  - AVC-compatible
  - Inter-view prediction
- Adopted as 3D image format for Blu-ray Discs

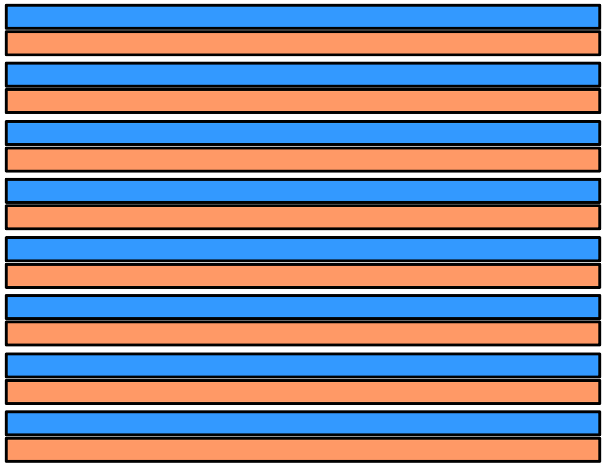


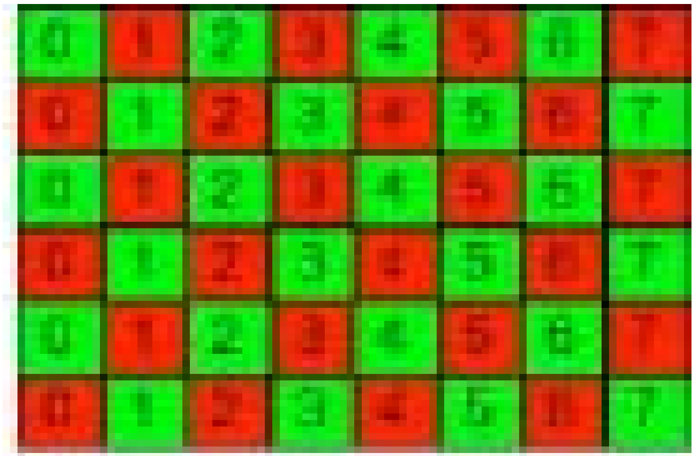
Frame-Compatible Stereo

- Side-by-side
- Top-down
- Interlaced
- Checker-board





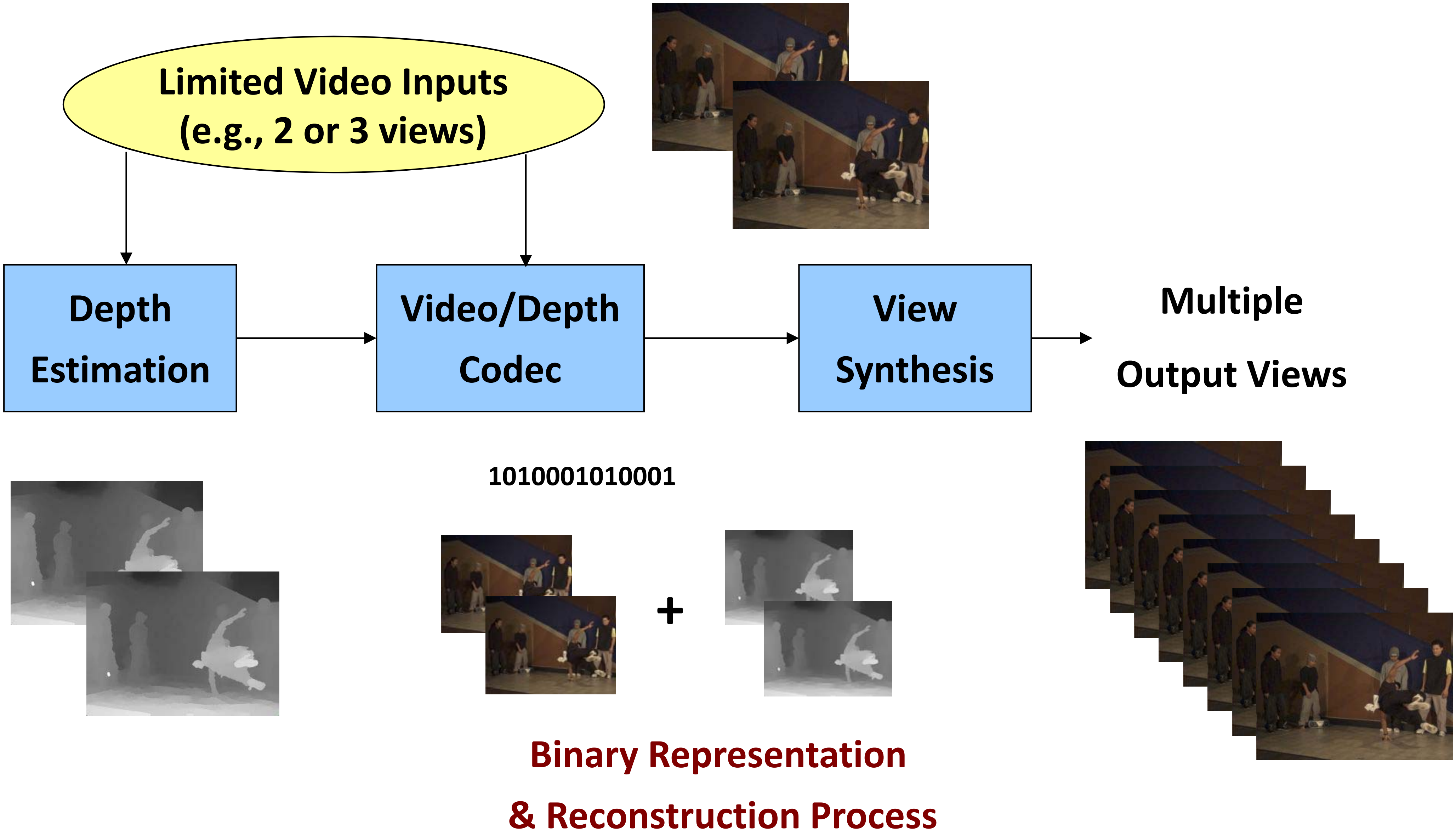




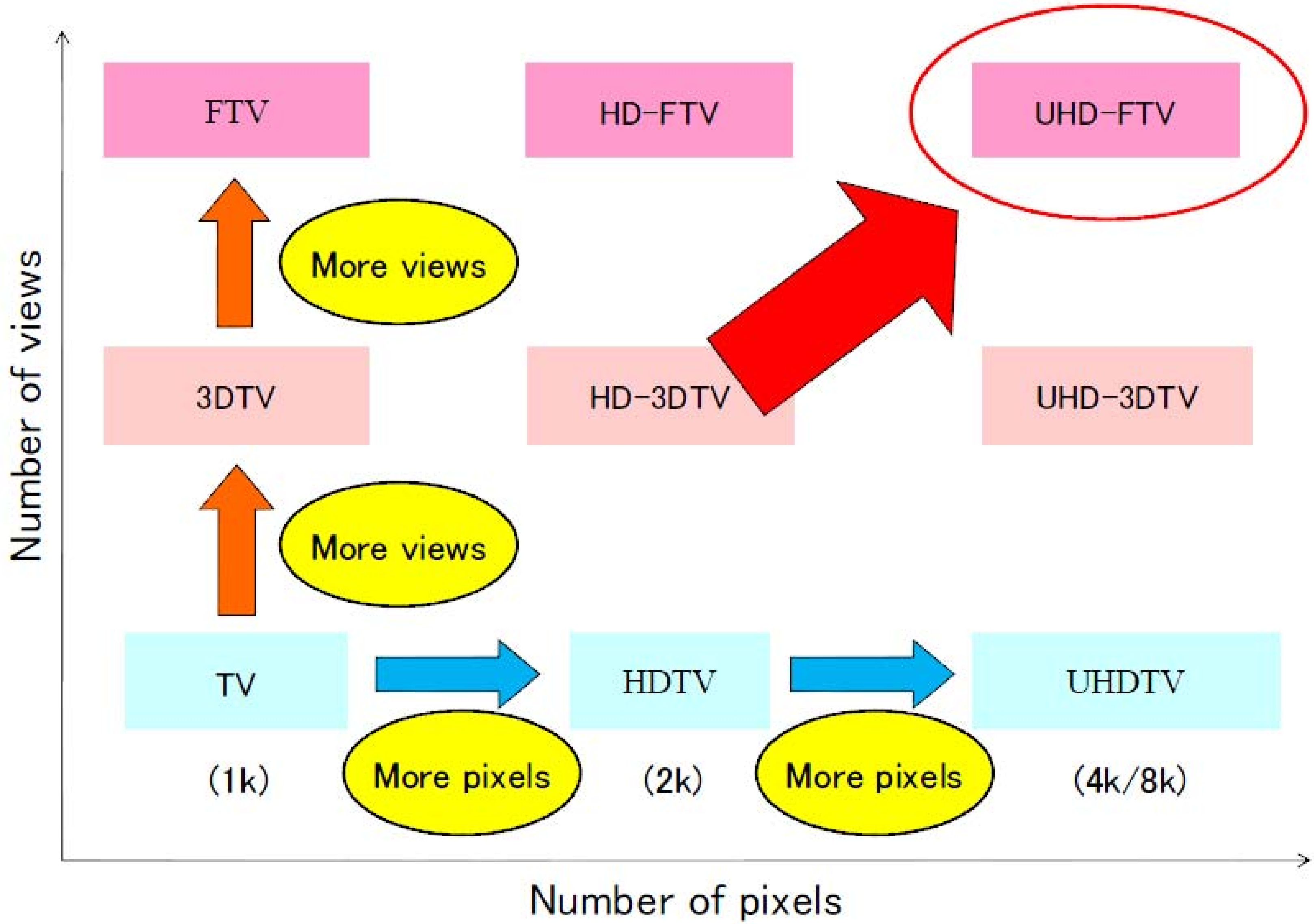
Depth-based Image Formats

- 2D Color Video + Depth
  - Backward compatible to 2D video
  - Limited depth range
  - Low rate overhead
- Multi-view Video + Depth
  - Stereo signal readily accessible
  - Better handling of occlusions
  - Higher rate overhead

3DV Reference Framework



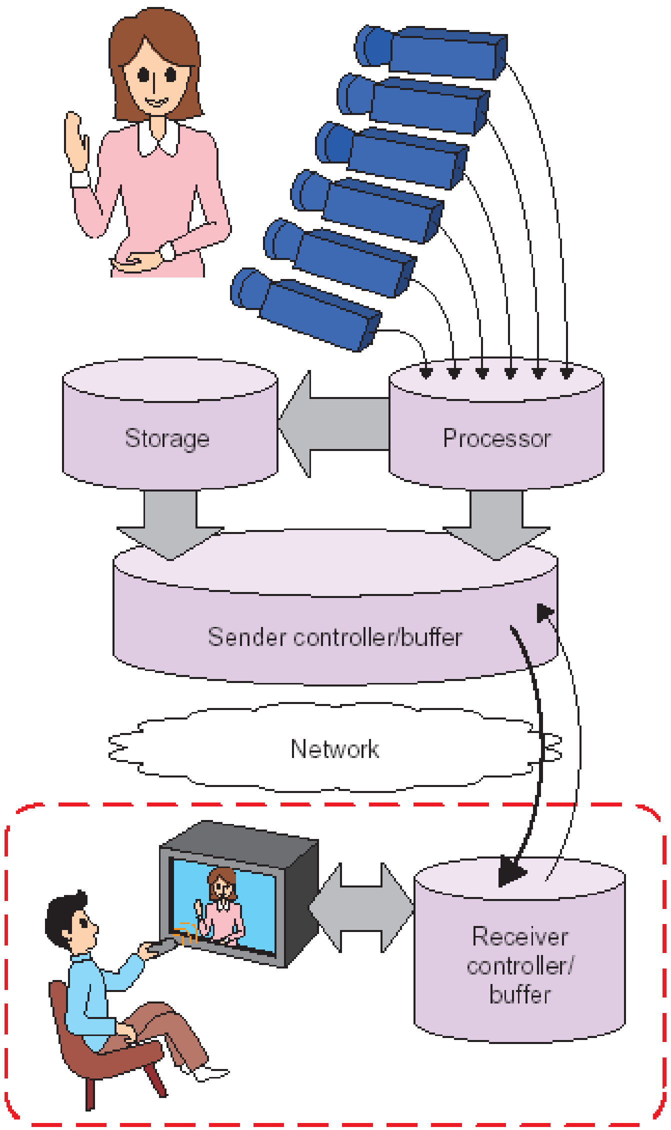
Progress of Visual Media



Number of Views in Visual Media

	2DTV	1
3DTV	Stereoscopic TV	2
	Auto-stereoscopic TV	tens
	FTV	infinite

Free-viewpoint Video



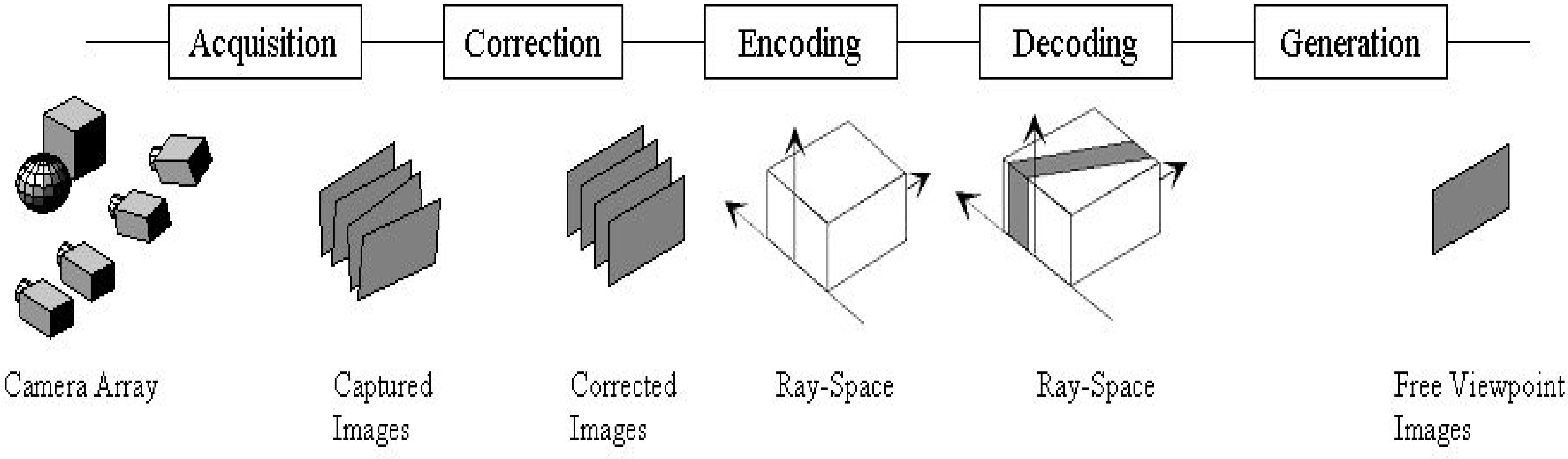
- Provides
  - Ability to change viewpoint freely
- Multiple Views Available
  - Render one view (real or virtual) to legacy 2D display
- Useful for
  - Surveillance, broadcast TV and stored interactive video, games

Free-viewpoint TV (FTV)

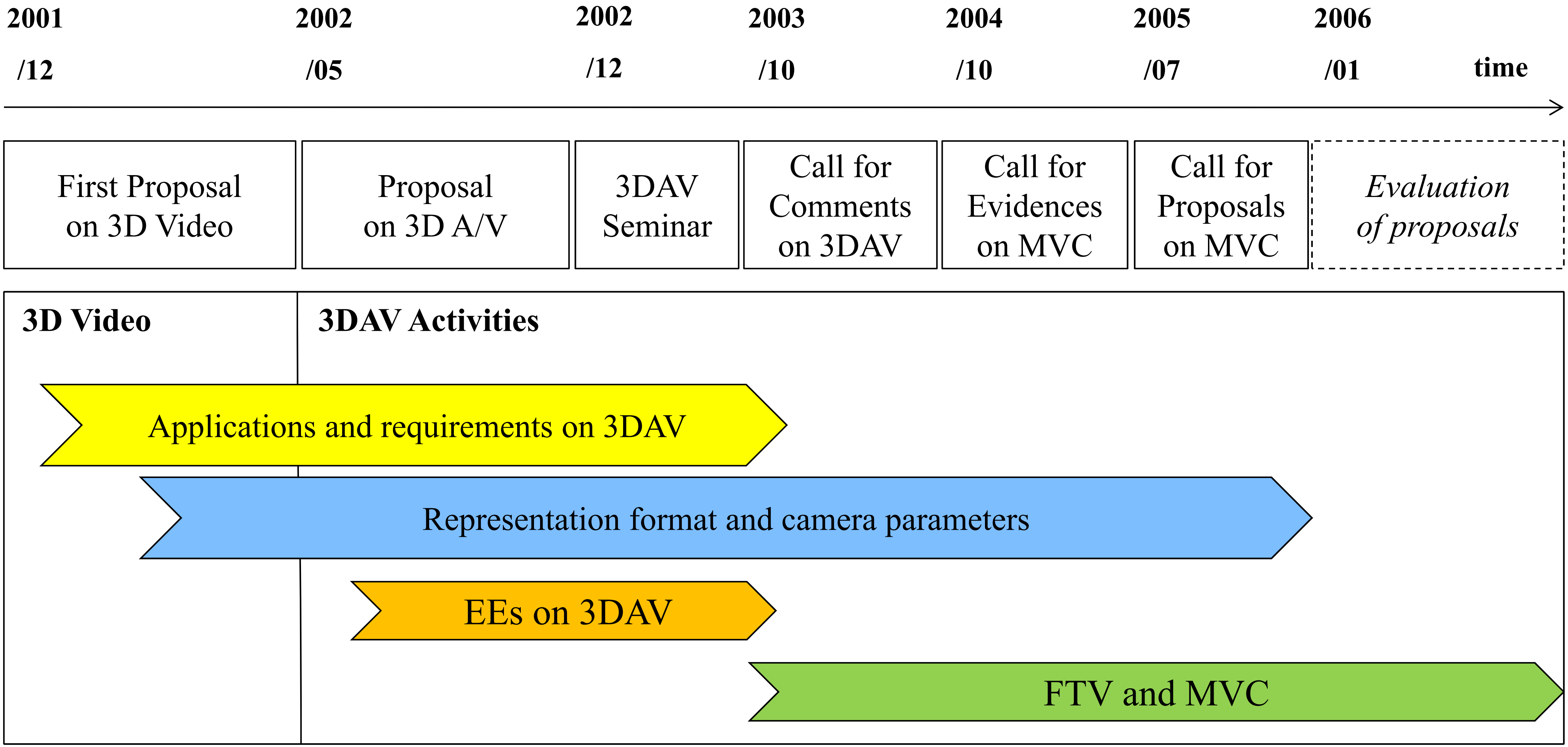
❖ FTV

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

❖ FTV System



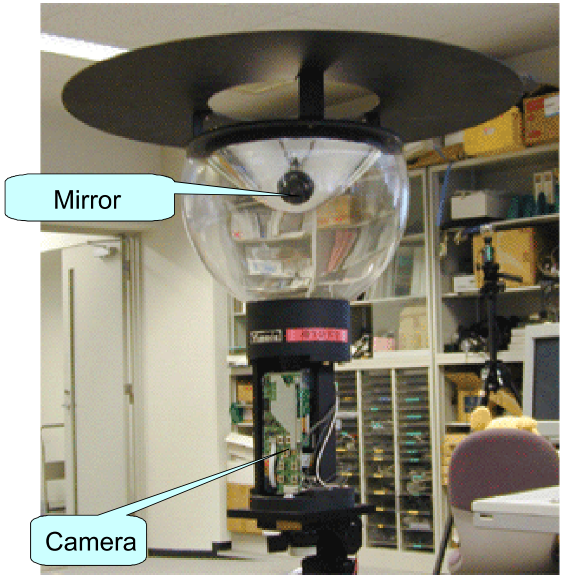
History of 3DAV/MVC

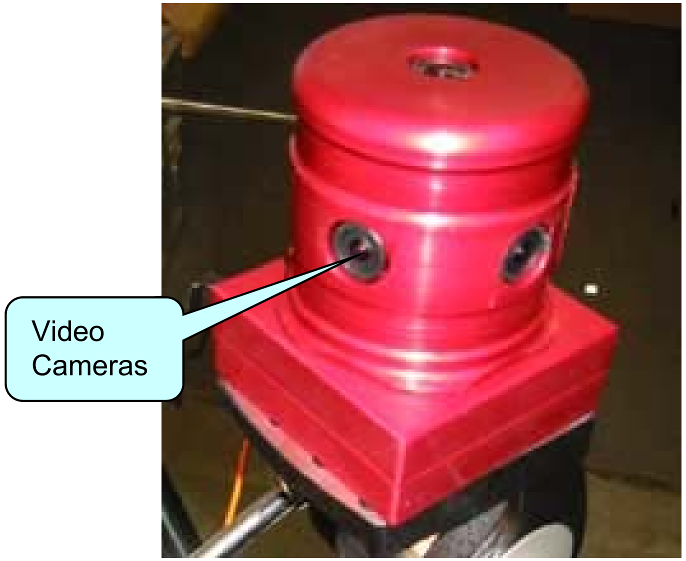


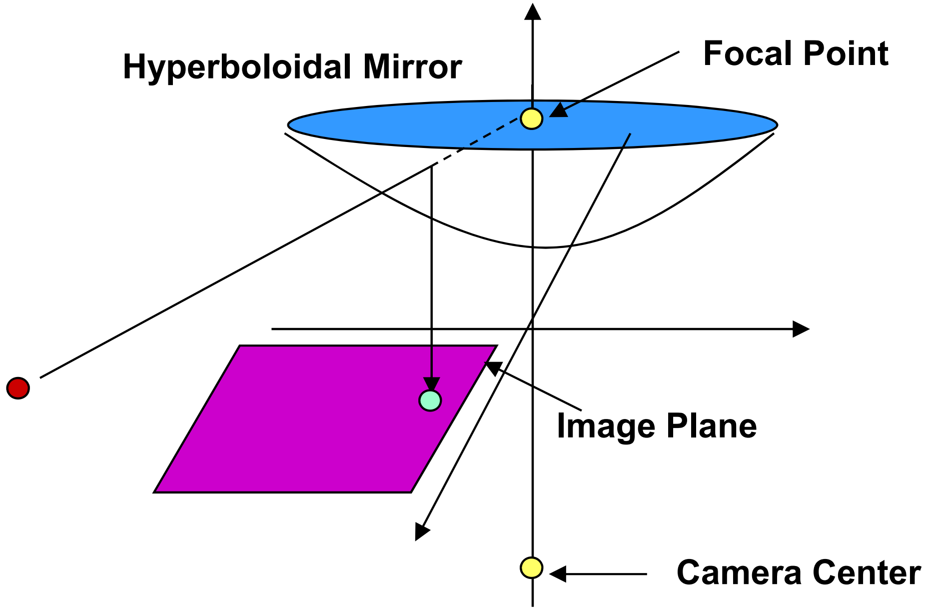
3DAV Exploration Experiments

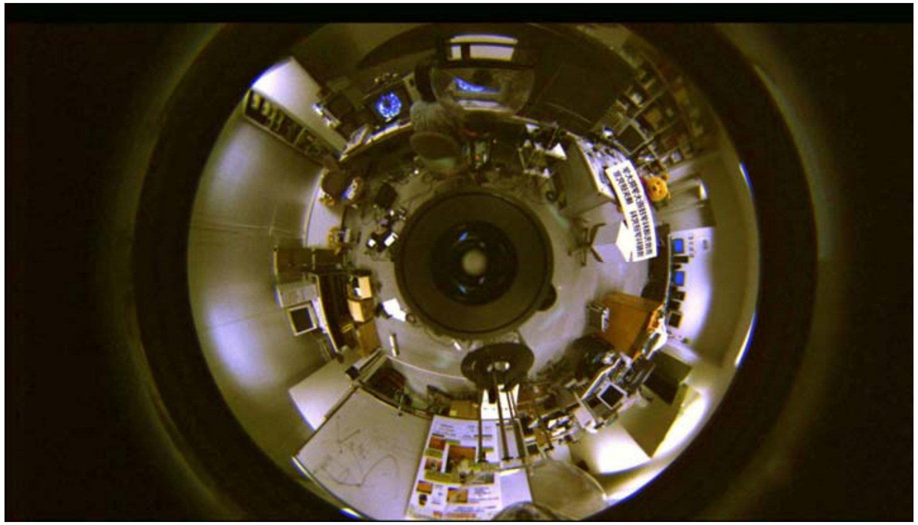
- EE1
  - Omni-directional Video
- EE2
  - Free viewpoint Video
- EE3
  - Stereoscopic video coding using MAC
- EE4
  - Depth/disparity coding for 3DTV and intermediate view interpolation

EE1: Omni-directional Video





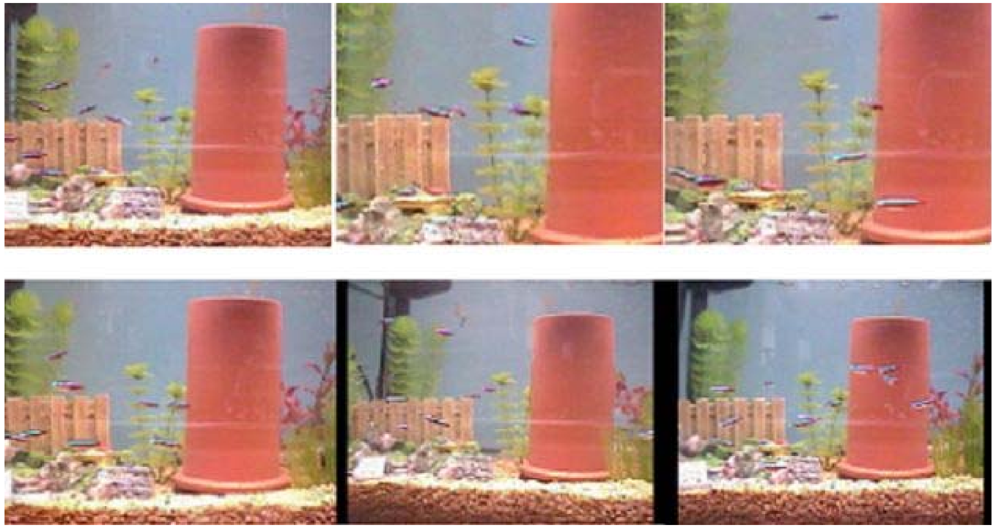




Omni-directional Image

EE2: Free Viewpoint TV

Camera Array

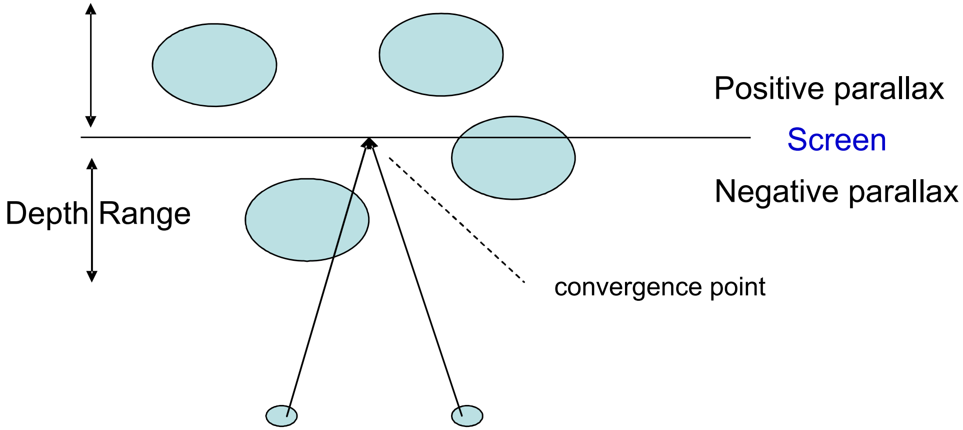


Generated Multi-view Images

EE3: Stereoscopic Video

Left-view Image

Right-view Image



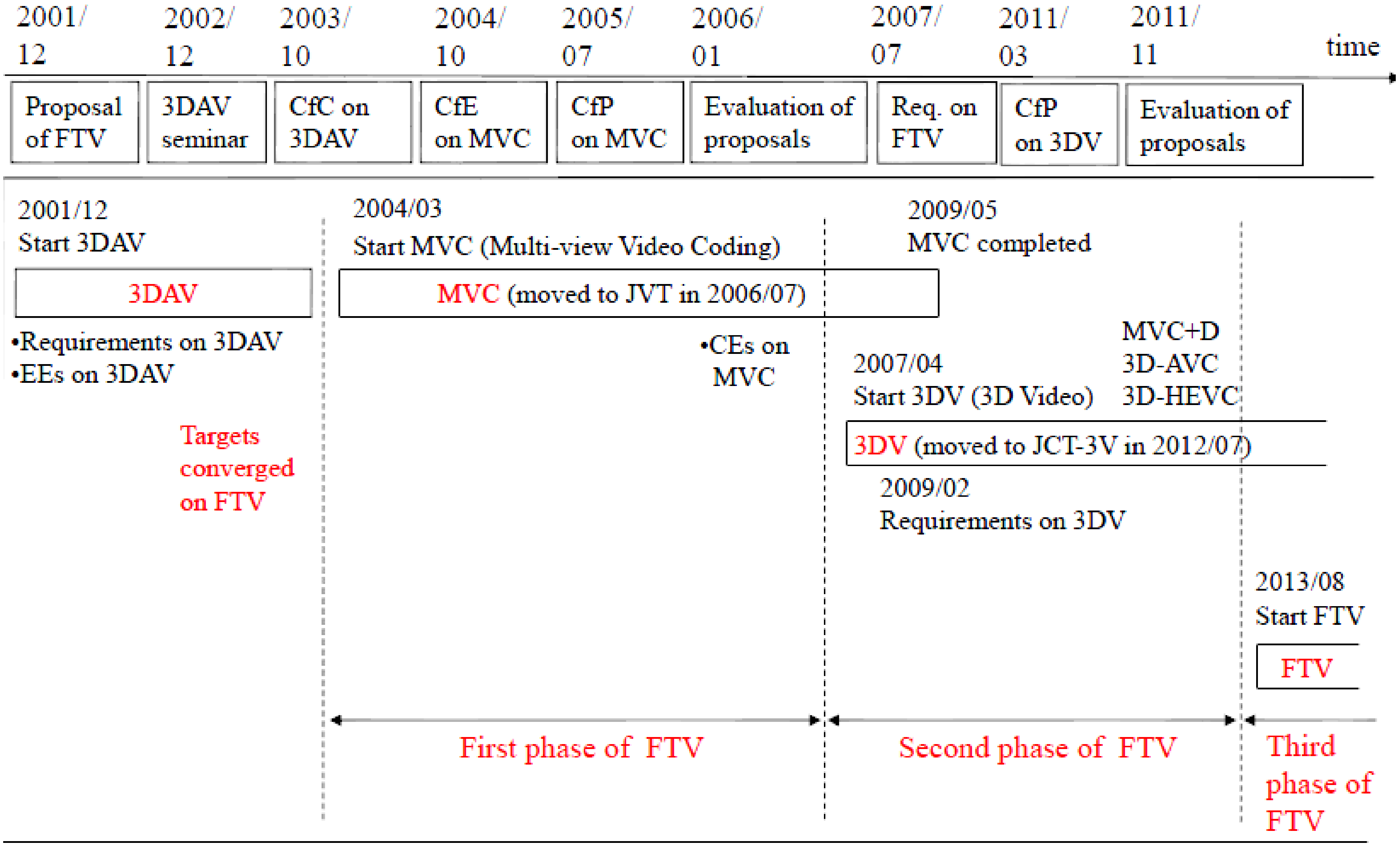
EE4: Depth/Disparity Coding

- Color and Depth Information

Color Information

Depth Information

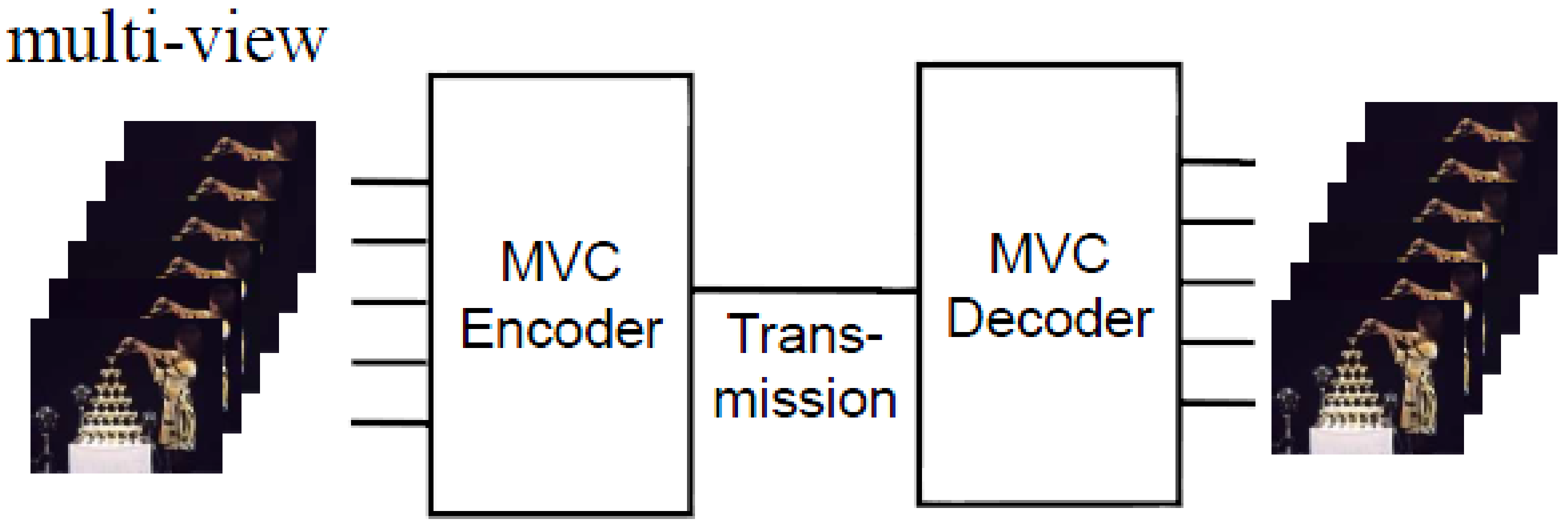
History of MPEG FTV



MPEG FTV

- First phase
  - Multi-view video coding
  - Enables efficient coding of multiple camera views
- Second phase
  - 3D video (3DV)
  - Enables viewing adaptation and display adaption of multi-view displays
- Third phase
  - Started in August 2013
  - Targets super multi-view and free navigation applications
  - To establish a new FTV framework

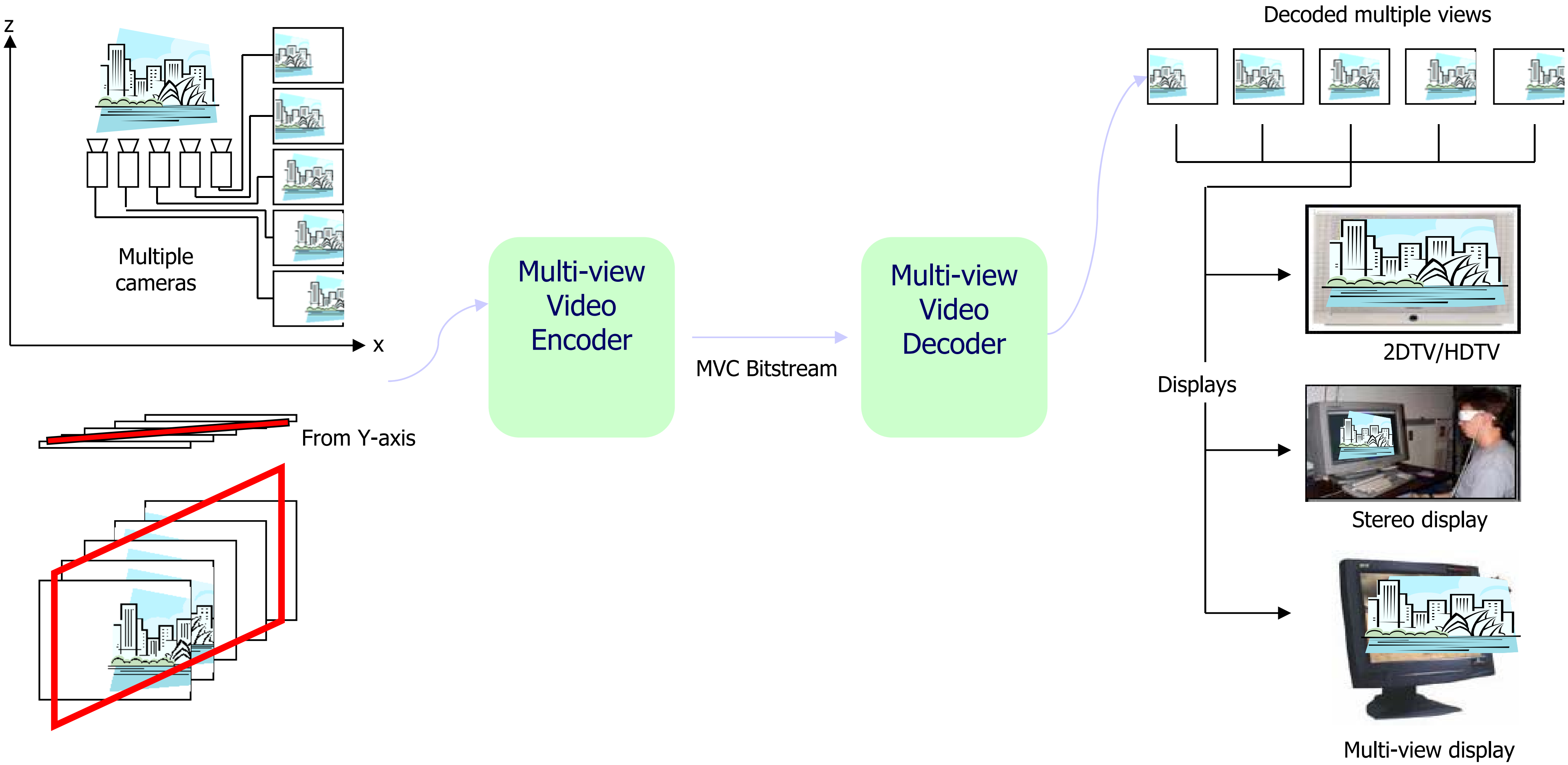
MVC: First Phase of FTV



Coding of multi-camera views

View synthesis is not included

Multi-view Video Coding (MVC)

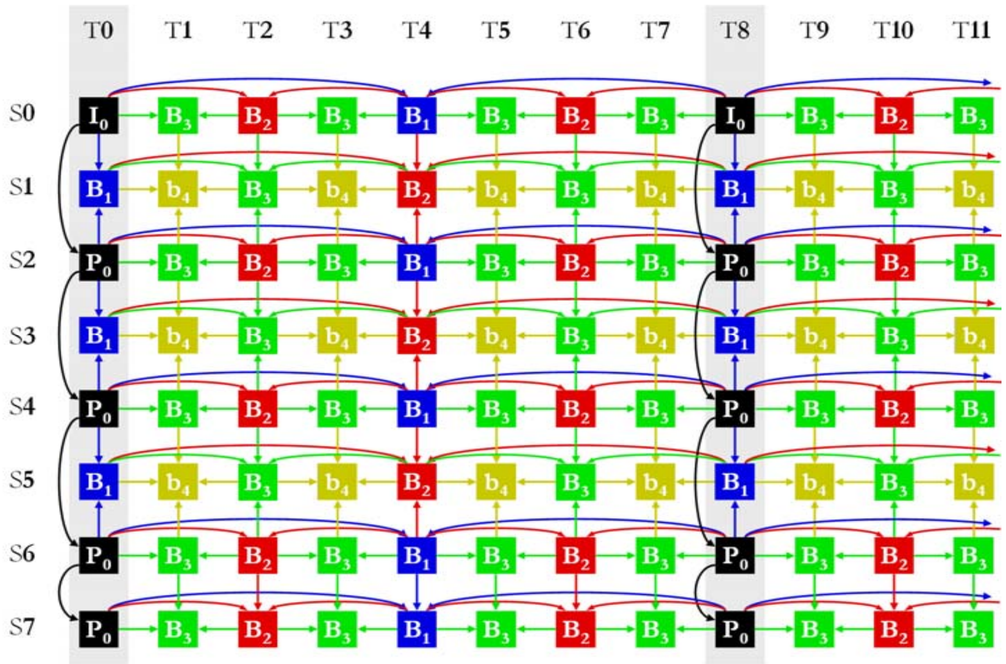


Requirements for MVC

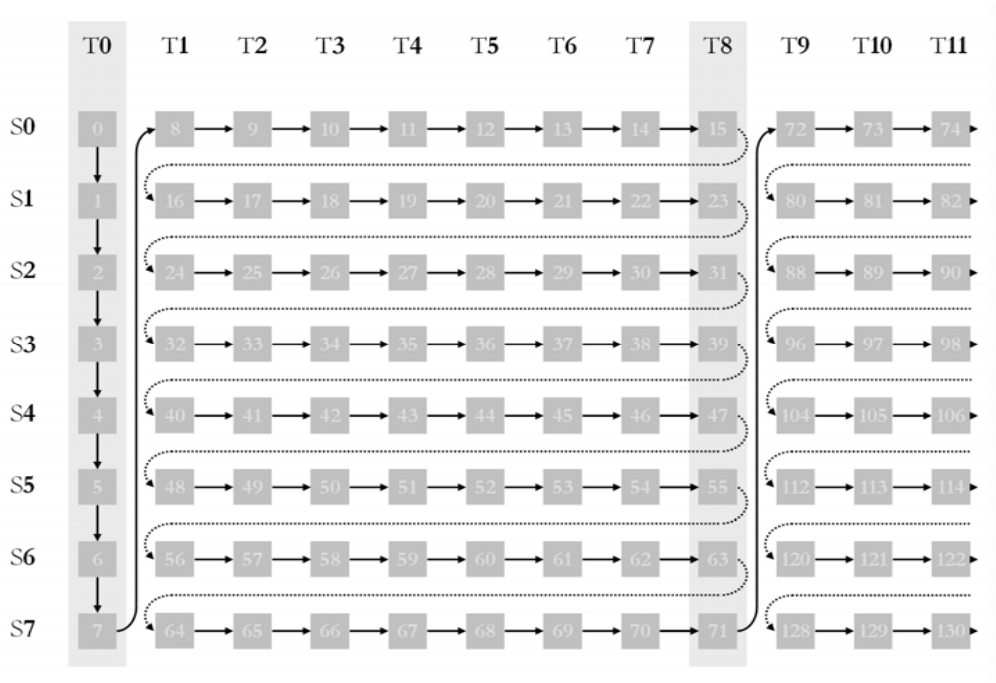
"shall" (mandatory)	▪ Compression efficiency ▪ View scalability/Free viewpoint scalability ▪ Backward compatibility ▪ Low delay ▪ Resolution, bit depth, chroma sampling format ▪ Temporal random access/View random access ▪ Resource management ▪ Parallel processing
"should" (desirable)	▪ Spatial/Temporal/SNR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

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

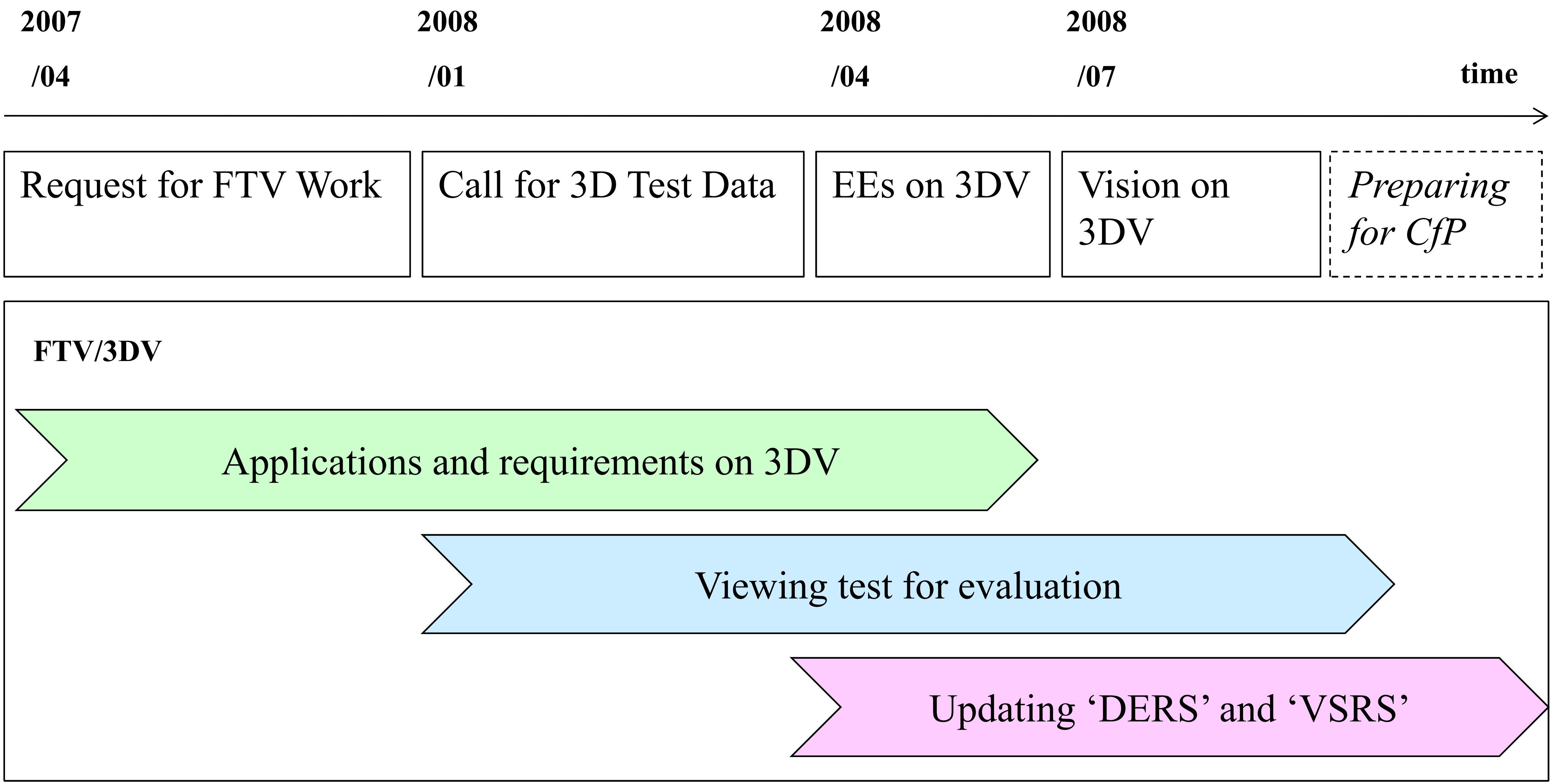


Inter-view-temporal prediction structure based on AVC, using hierarchical B pictures

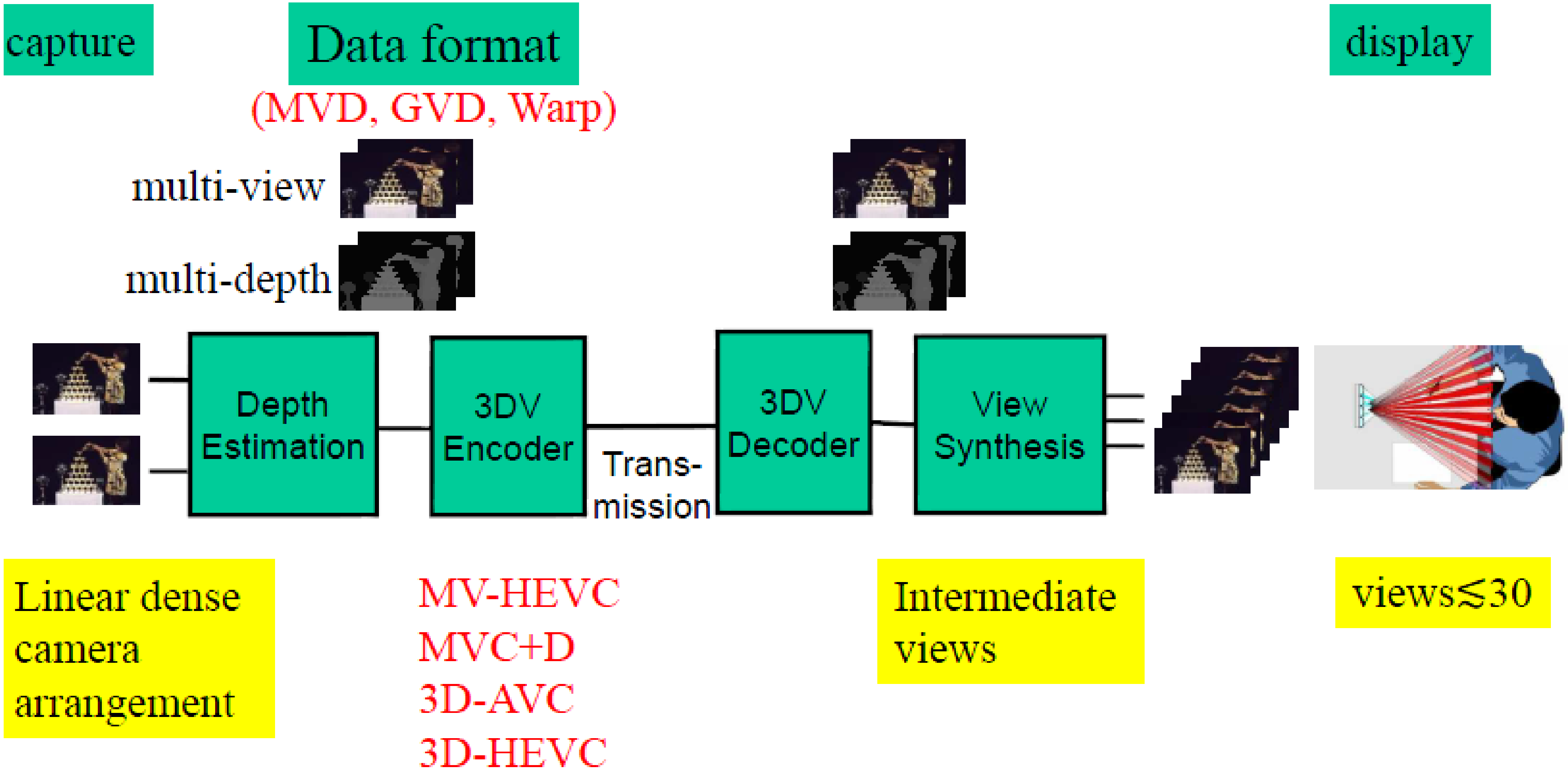


Associated reordering of multi-view input for compression with AVC

History of 3DV/FTV

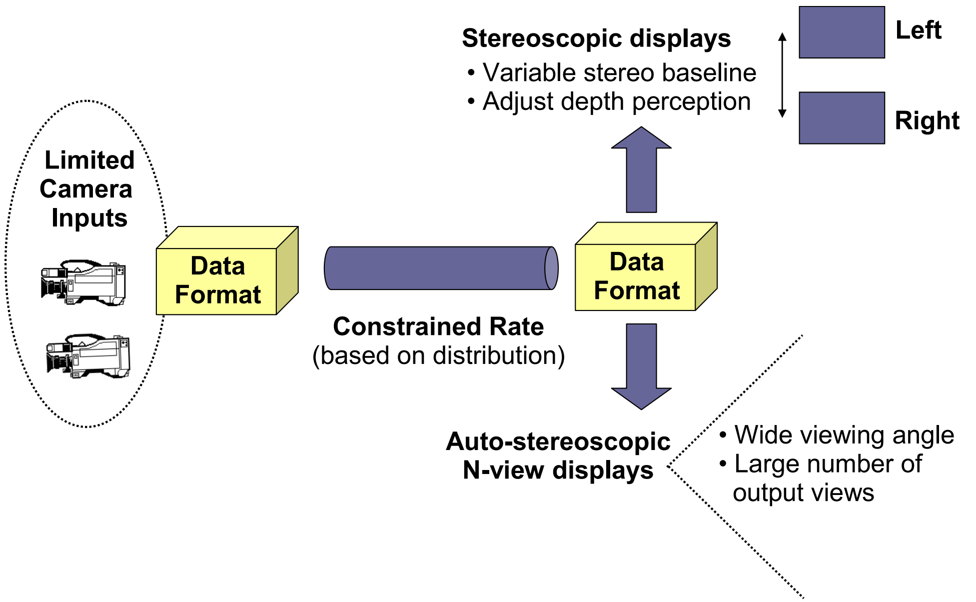


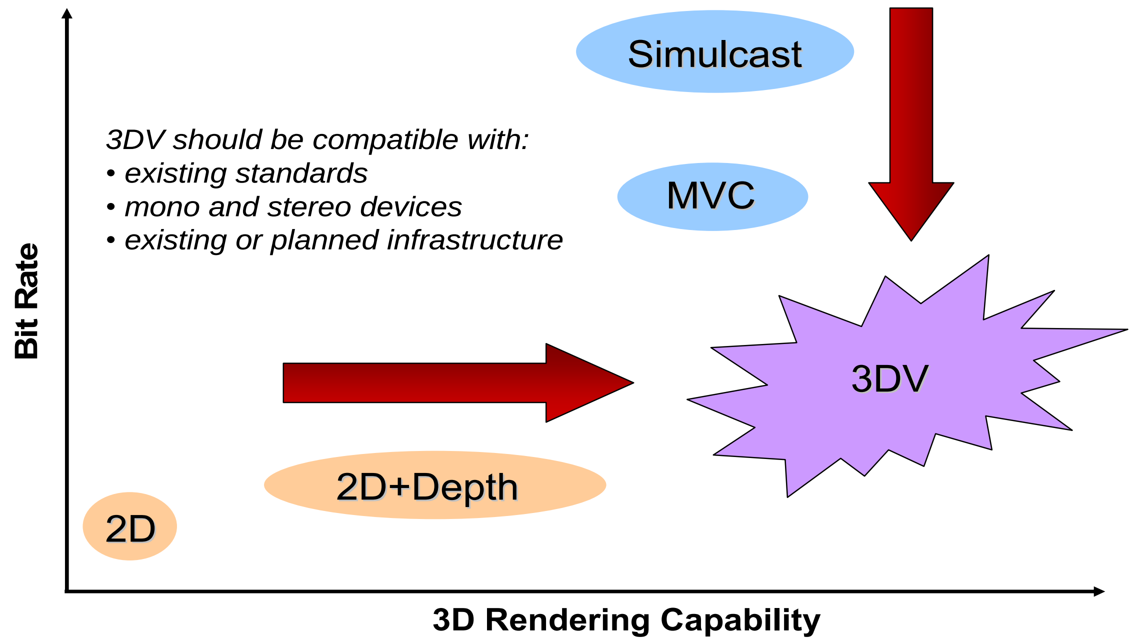
3DV: Second Phase of FTV



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.

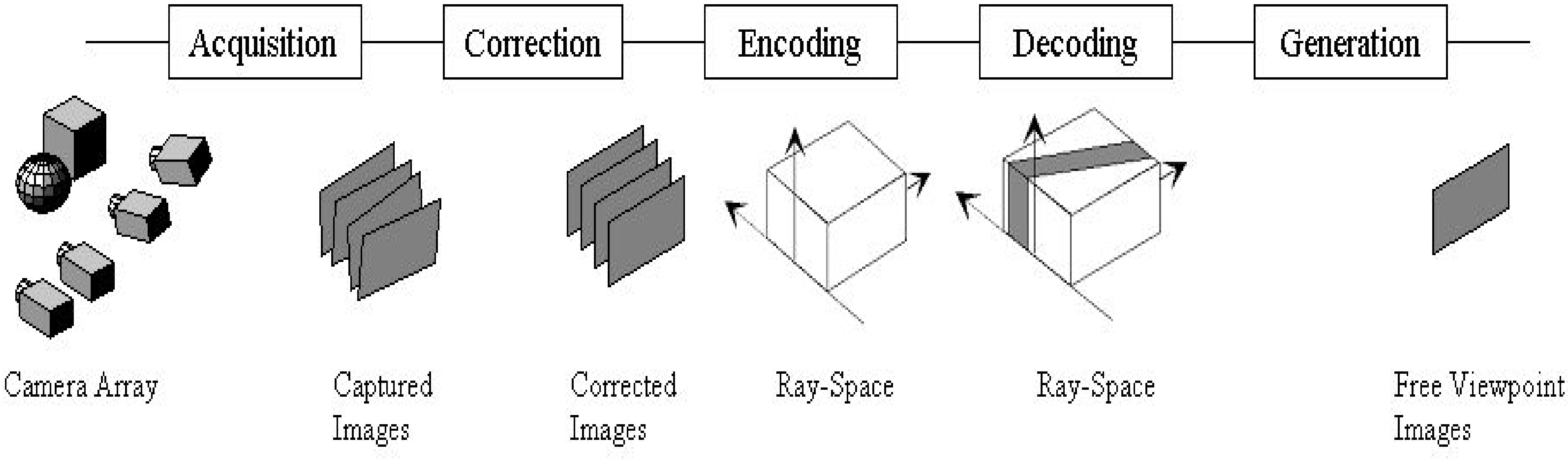




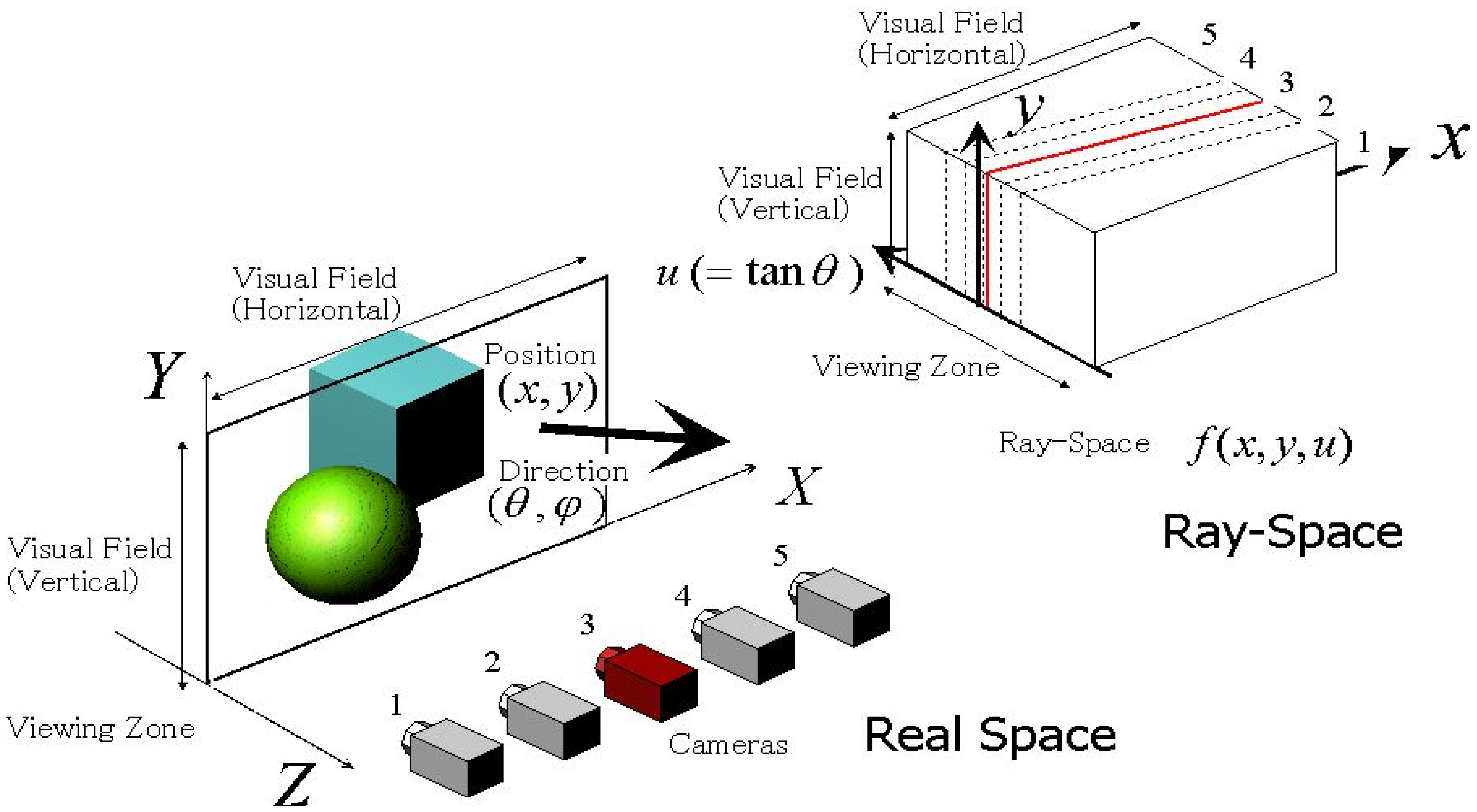
Free-viewpoint TV

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

- FTV System

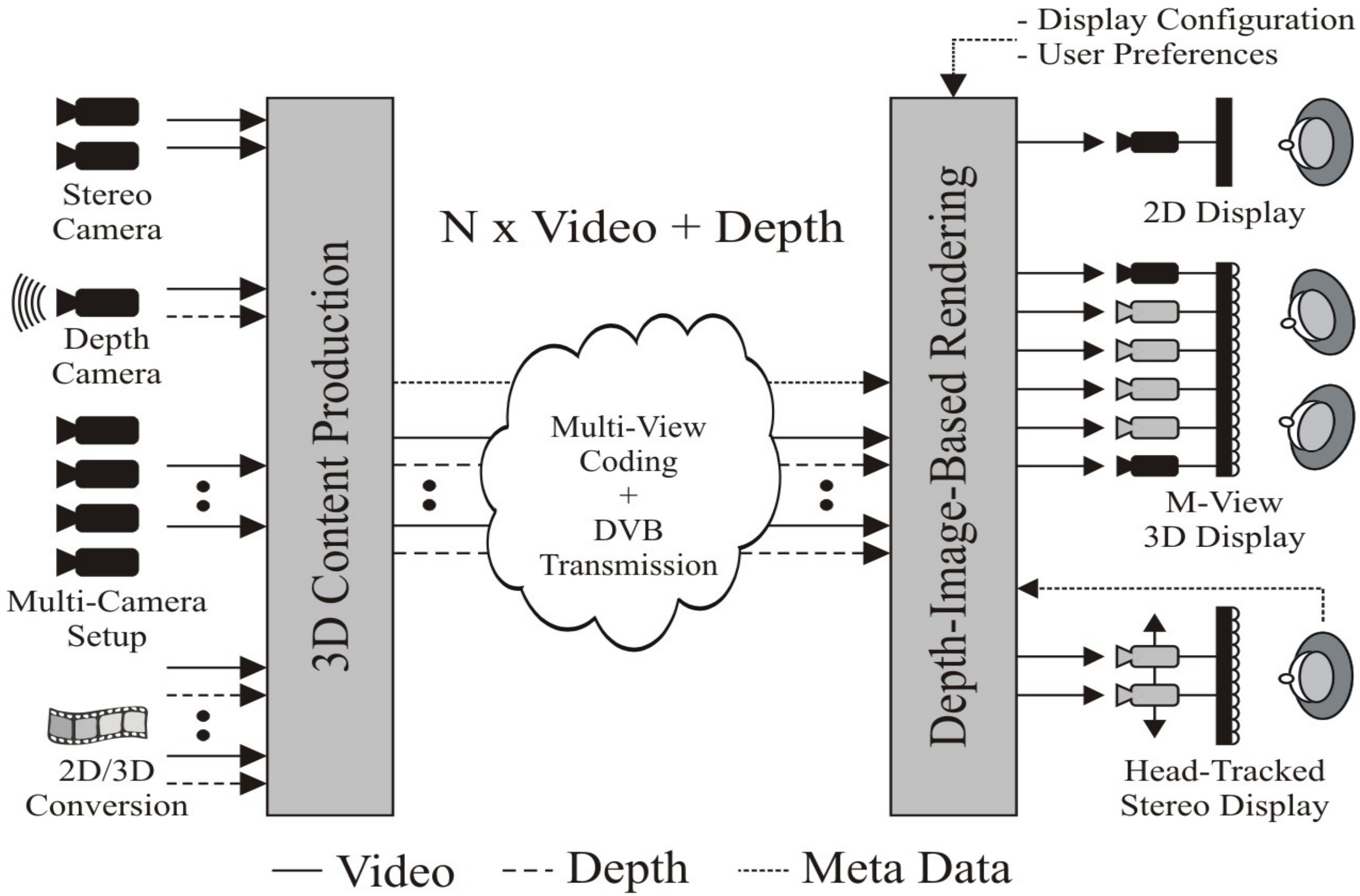


Ray-Space Representation



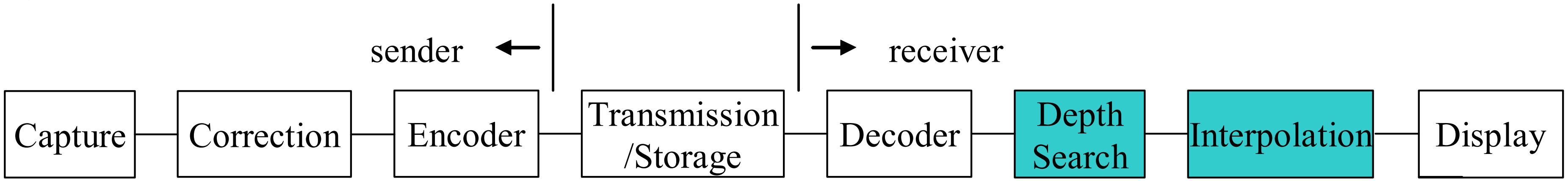
N Video + Depth

- Multi-view Video plus Depth (MVD) data format
  - Free-viewpoint navigation and multi-view video display
  - DIBR(depth image-based rendering)-based multiview generation

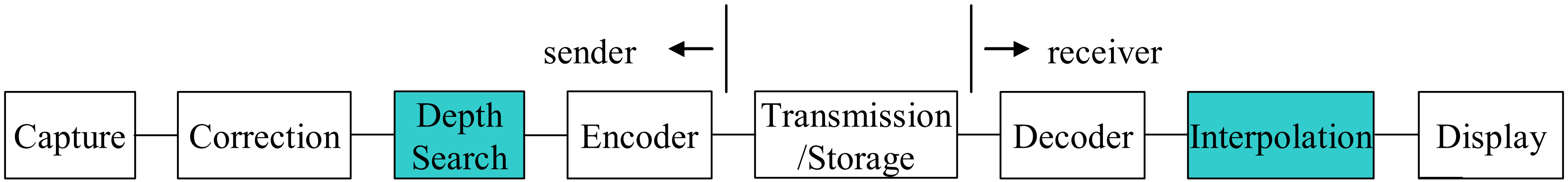


Depth Search and Interpolation

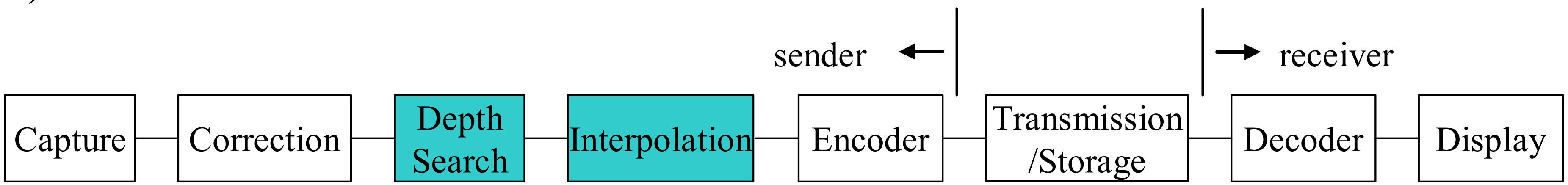
(a) Case A



(b) Case B



(c) Case C

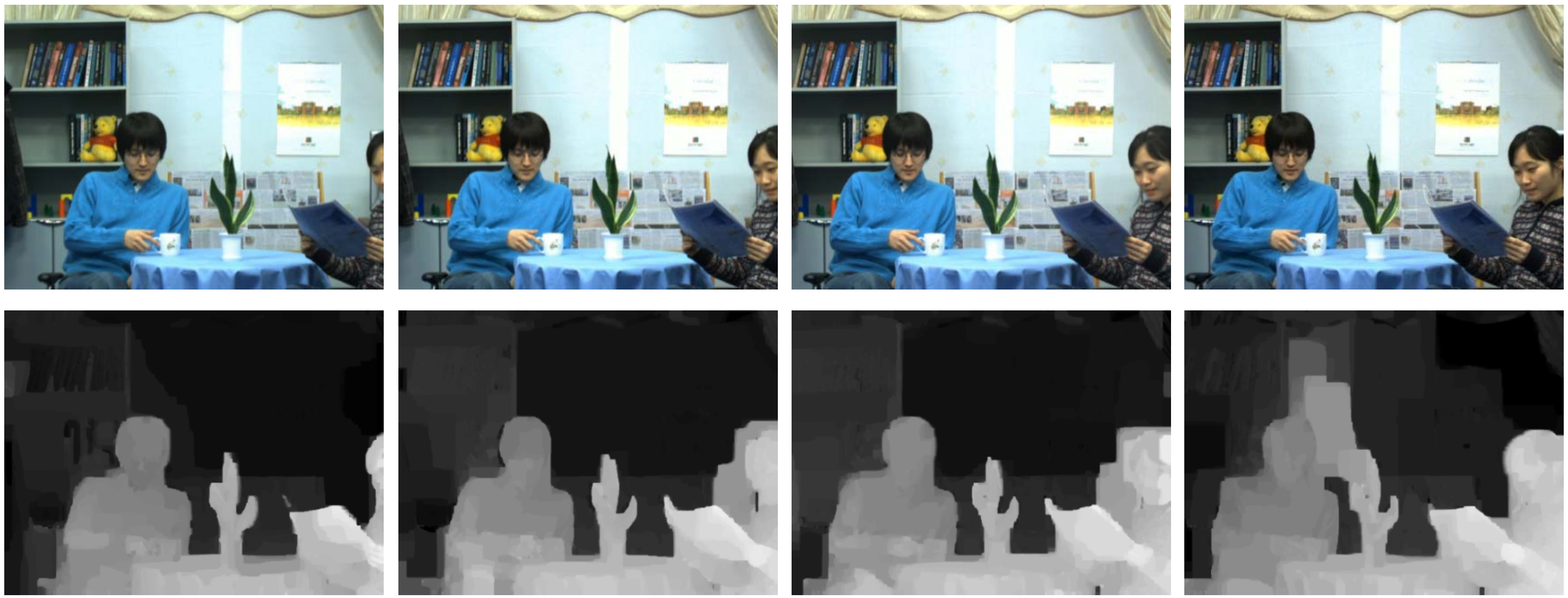


Challenging Issues

- 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

Depth Estimation Reference S/W

- Depth Image
  - Represents distance information from the camera to objects in a scene
  - Quantized values with 8-bit gray scale image
- Depth Estimation Reference Software
  - Depth estimation software based on Graph-cuts
  - Use three views for generating the center depth map



Active Sensor-based Method

- Time-of-flight range sensor
  - Use the infrared range sensor
  - Time-of-flight (TOF) depth sensor
  - Depth range: quantized 8-bit values for 256 levels
- Structured light
  - Projection of a light pattern at the known angle to an object
  - Distortion in the line can be translated to depth

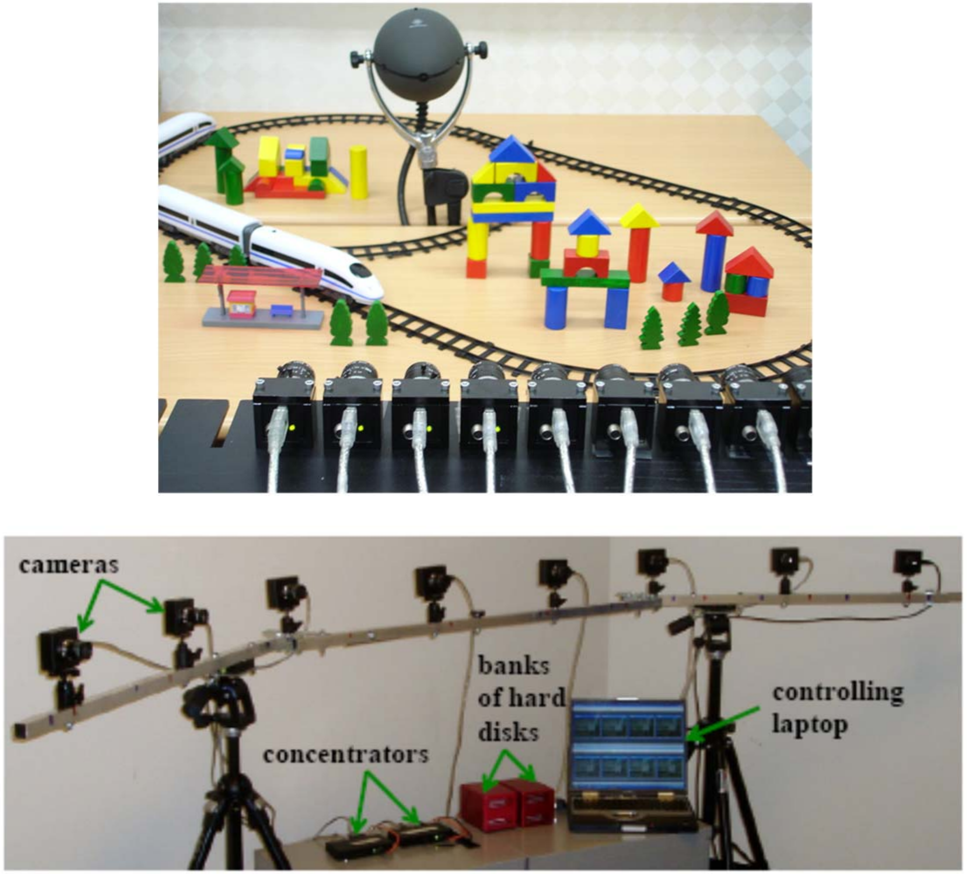
TOF camera

Structured light camera

Passive Sensor-based Method

- Stereoscopic camera
- Multi-view camera

Stereoscopic cameras

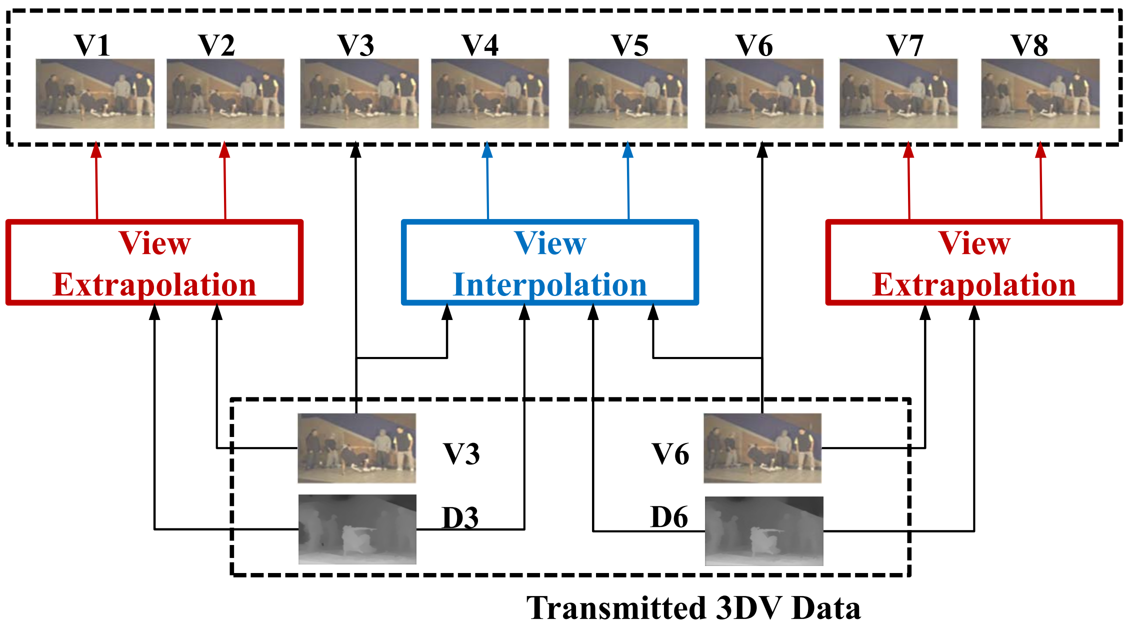


Multi-view camera arrays

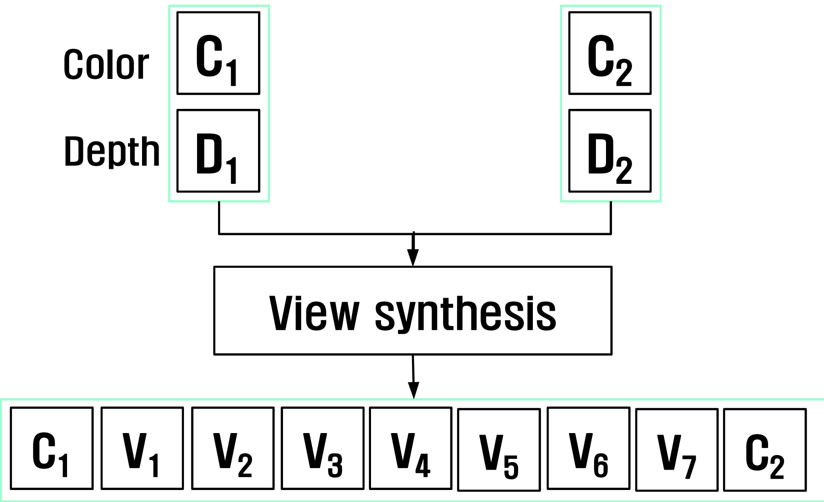
View Synthesis Reference S/W

❖ View Synthesis
- Generate a virtual viewpoint's image using provided view and depth data
- Any viewpoint image can be generated

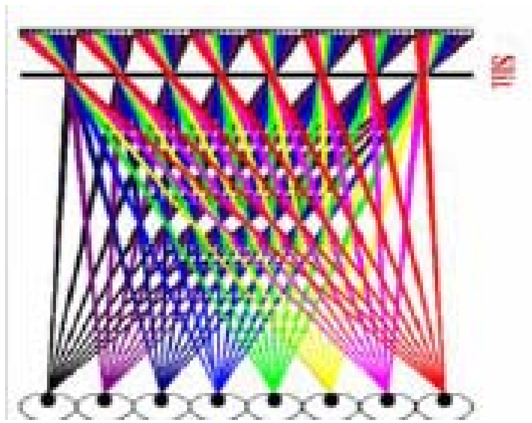
❖ View Synthesis Reference Software
- DIBR (depth image-based rendering) based virtual view generation



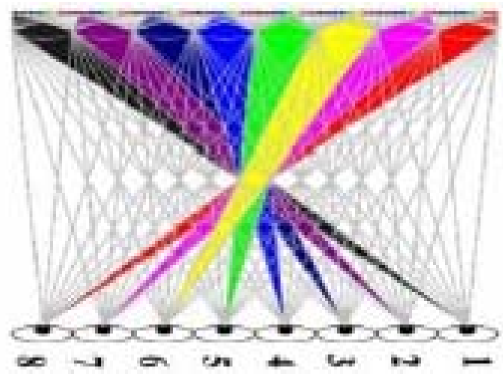
View Synthesis for Multi-view Displays



Auto-stereoscopic Display



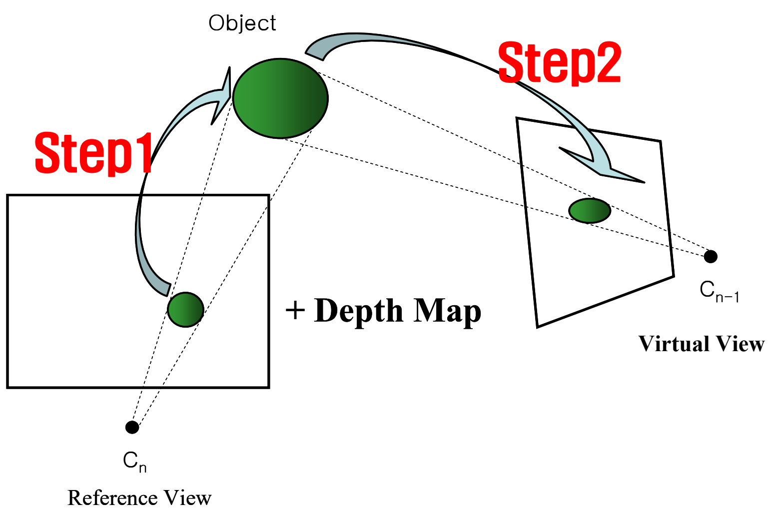
Parallax Barrier



Lenticular Lens

3D Image Warping

- 3D image warping using depth information
- Two operations
  - Project pixels of the reference view into the world coordinates
  - Re-project the image into the virtual view



Step 1: Backward projection from image to world

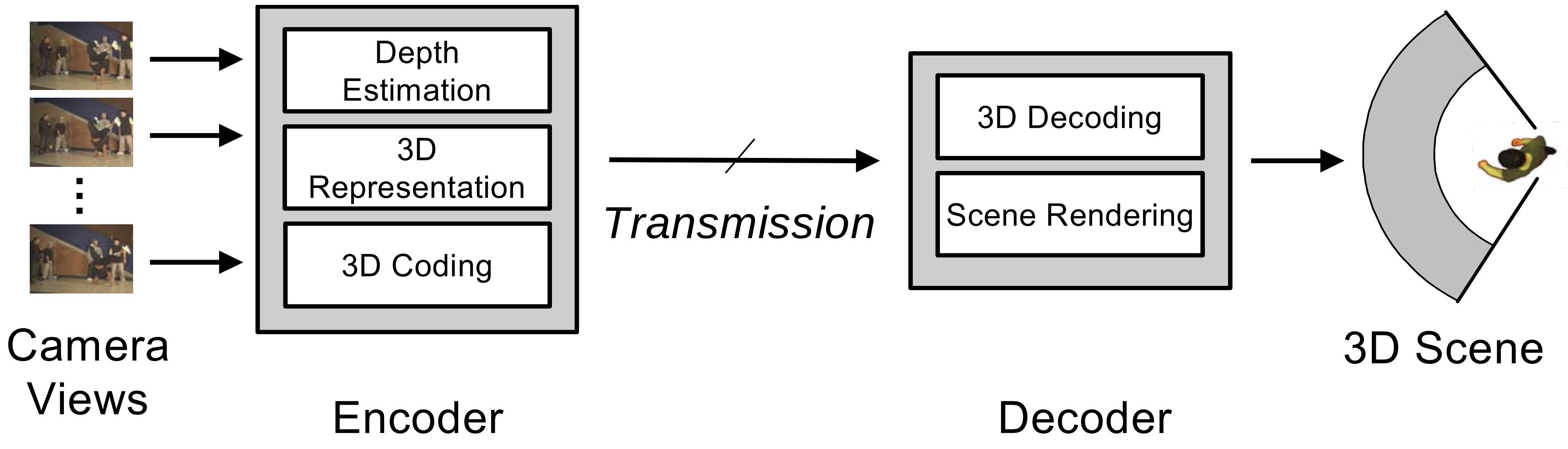
$$\mathbf{M}_r = \mathbf{R}_r^{-1} \cdot \mathbf{A}_r^{-1} \cdot \mathbf{m}_r \cdot \mathrm{d}(\mathbf{m}_r) - \mathbf{R}_r^{-1} \cdot \mathbf{t}_r$$

Step 2: Re-projection from world to target world

$$\tilde{\mathbf{m}}_t = \mathbf{A}_t [\mathbf{R}_t \mid \mathbf{t}_t] \tilde{\mathbf{M}}_r$$

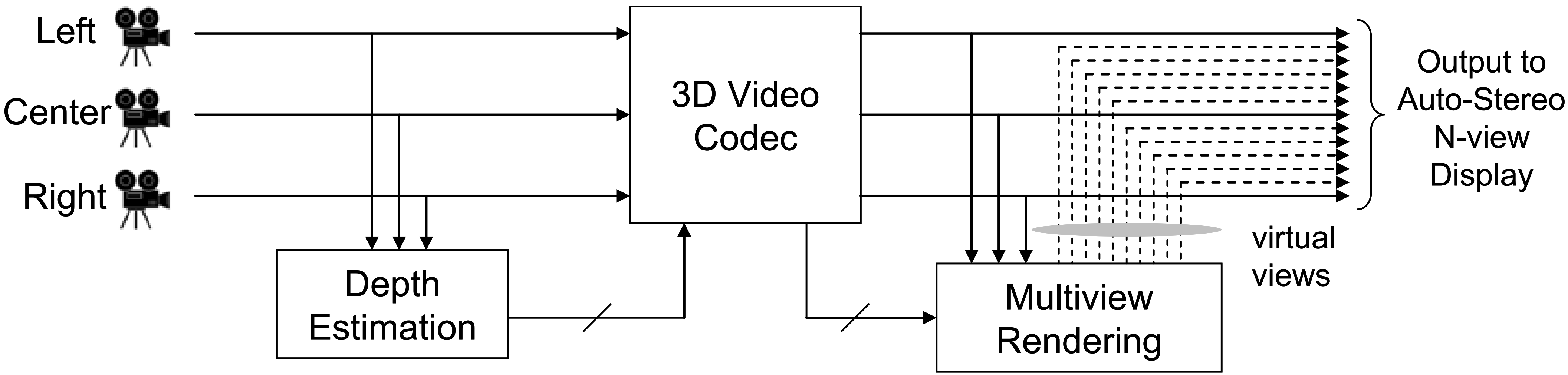
Advanced 3D Video System

❖ Framework for Advanced 3D Video System
- Compatible with 3D video coding under developing by the MPEG group
- Encoder: depth map generation, 3D data conversion, and 3D video coding
- Decoder: 3D data reconstruction, virtual view image generation, 3D video rendering



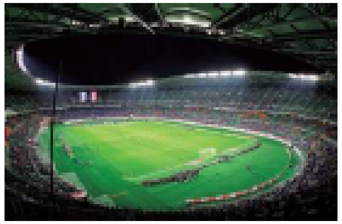
Multi-view Video

- Auto-stereoscopic Displays
  - Wide baseline configuration
  - Three-view configuration
    - Inputs: three texture video + three depth videos
    - Outputs: three reconstructed texture video + virtual view video(s)
  - Reconstructed depth data are used for generating virtual view



Third Phase of FTV

Capture



Data format

Scene Representation

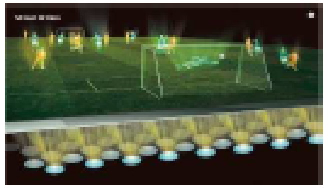
FTV Encoder

Trans-mission

FTV Decoder

View Synthesis

Display



- Wide viewing area
- Flexible camera arrangement

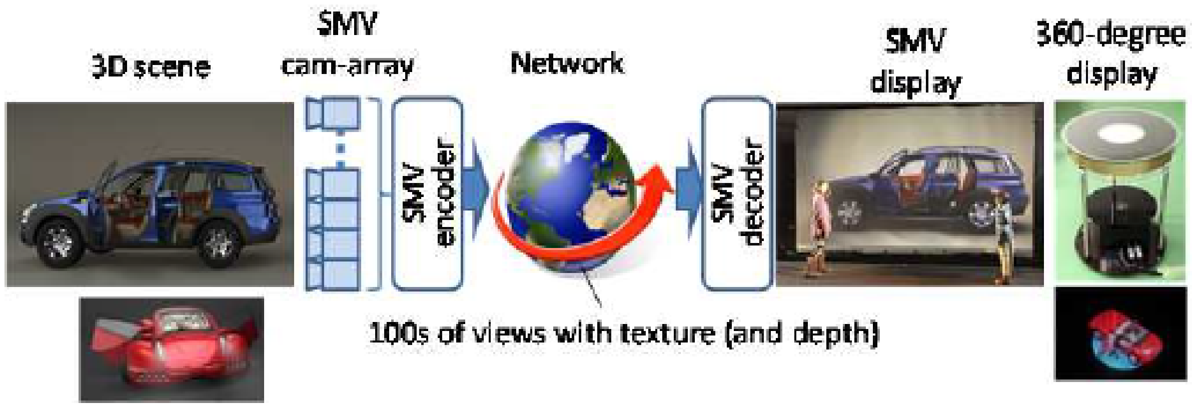
- Supplementary data for fast and robust view synthesis
- Supplementary data for specular reflection, semi-transparency etc.

- 200-500 views
- Free navigation over a wide area
- Robust to misregistration

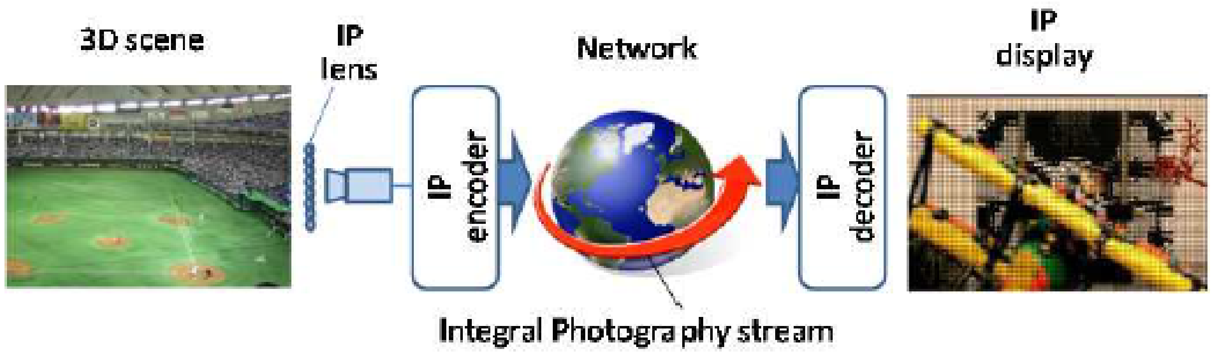
FTV needs higher specifications than 3DV

Target Applications of FTV

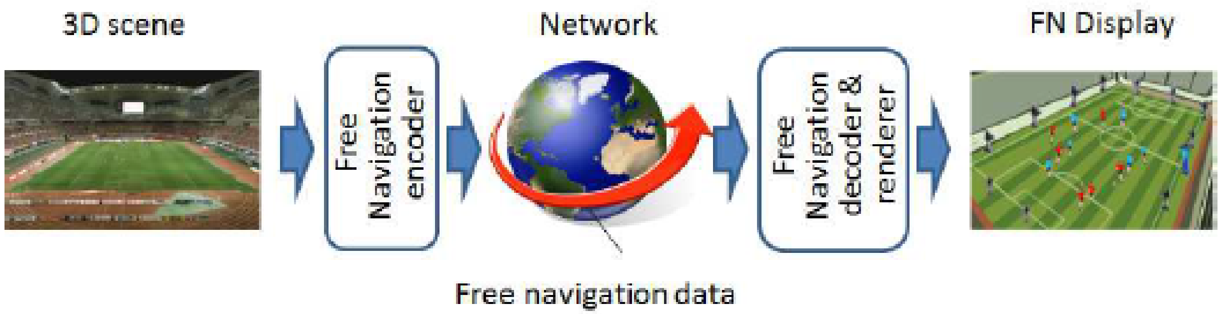
Super multiview with large horizontal parallax



Super multiview with full parallax



Free navigation



MPEG-I

- Launched in the 117th MPEG
- MPEG-I (ISO/IEC 23090)
- MPEG-I Parts
  - Part 1: Immersive Media Architectures
  - Part 2: Omnidirectional Media Application Format (OMAF)
  - Part 3: Immersive Video Coding
  - Part 4: Immersive Audio Coding
  - Part 5: Point Cloud Coding (PCC)

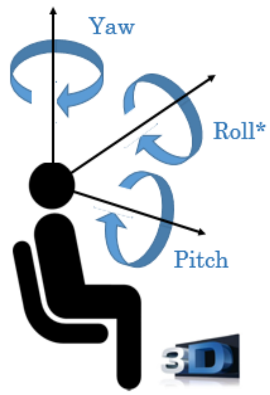
MPEG-I Working Phases

❖ MPEG-I development step

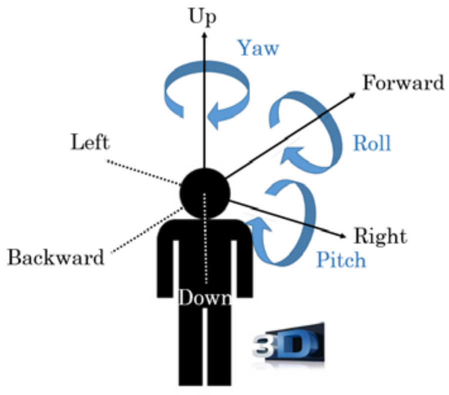
- Phase 1.a : Captured image based multi-view encoding/decoding
- Phase 1.b : Video based multi-view encoding/decoding
- Phase 2 : Video + additional data(depth, point cloud) based multi-view encoding/decoding

❖ MPEG-I classification of available view

- 3DoF+
- Omnidirectional 6DoF
- Windowed 6DoF
- 6DoF



3DoF



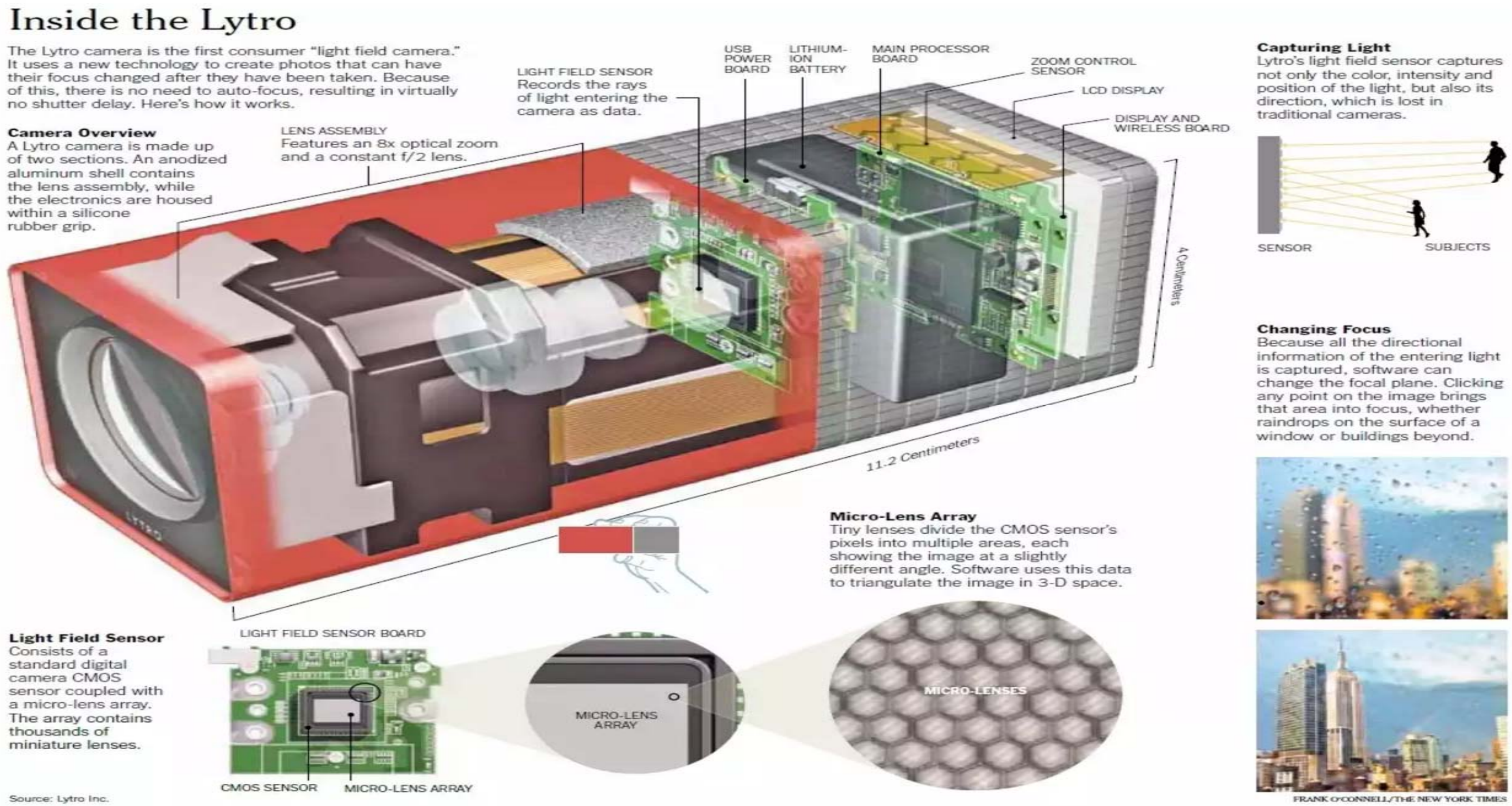
6DoF

❖ MPEG-I ad-hoc group (AhG)

- MPEG-I-Visual/Apps

Light Field Camera

- http://stateofvr.com/?page_id=17270



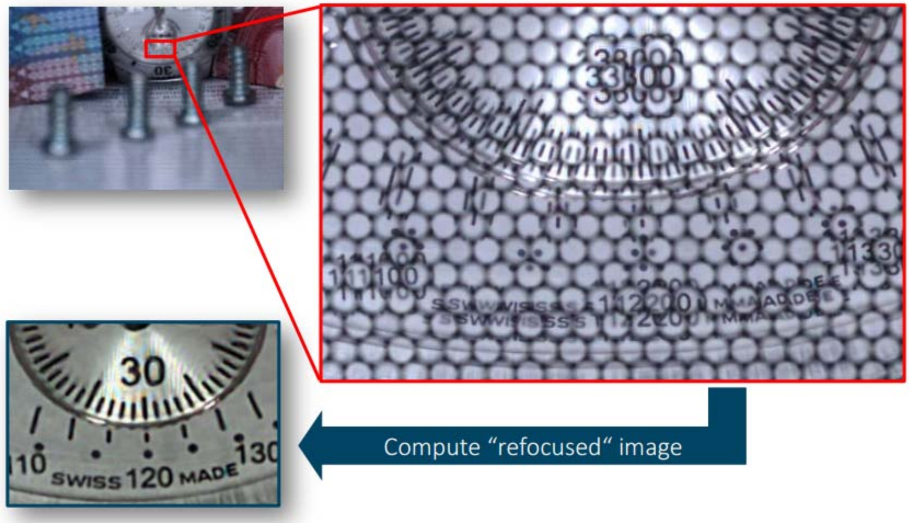
Micro Lens

- Generate sub-aperture images
  - Various optical directional information
- Differently focused image generation

focused in the back using destabilization

Characteristics of Micro Lens

- Lytro Illum (ver.2)
  - Implemented 200,000 micro lens array
- Light field camera and micro lens
  - Provide focused image (embedded camera software)
  - Plentiful color information



Micro lens image and auto-focused result

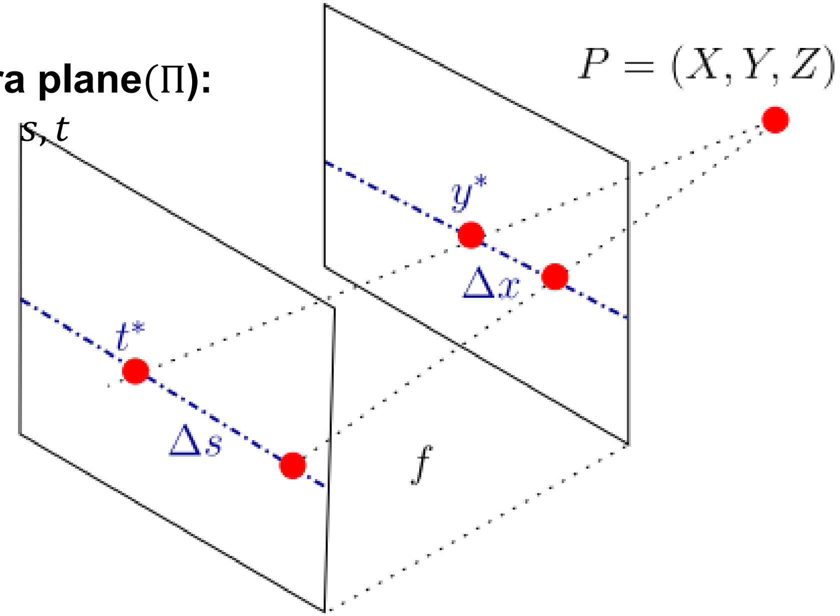
General camera image (left) & light field image (right)

Light Field Camera System

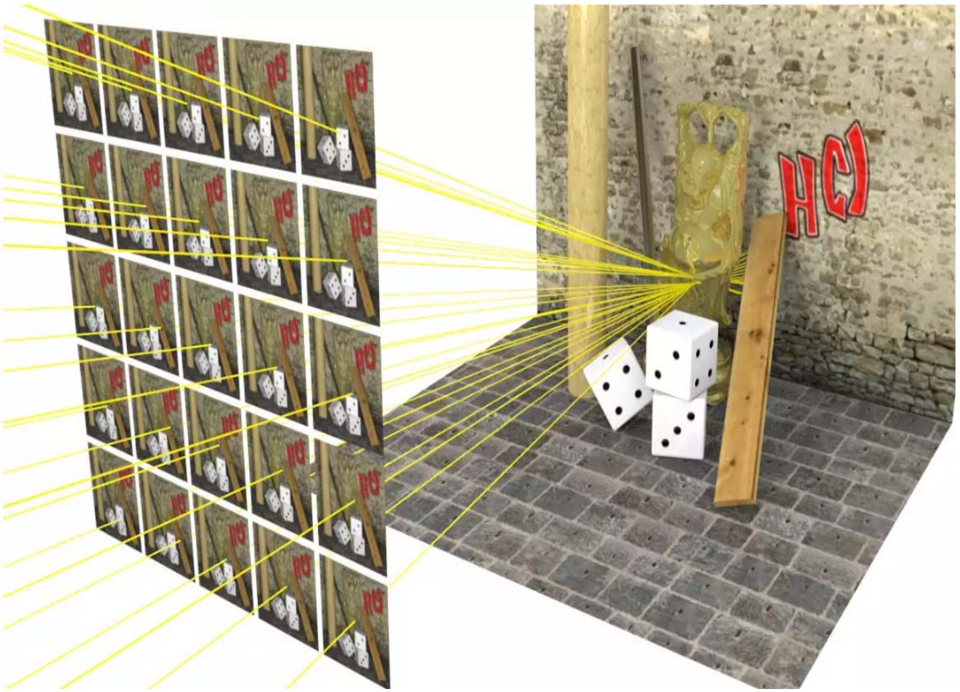
- Image and camera plane
  - Image plane: captured image coordinate plane (x, y)
  - Camera plane: set of image plane (s, t)

Image plane(Ω)**:** x, y

Camera plane(Π)**:** s, t

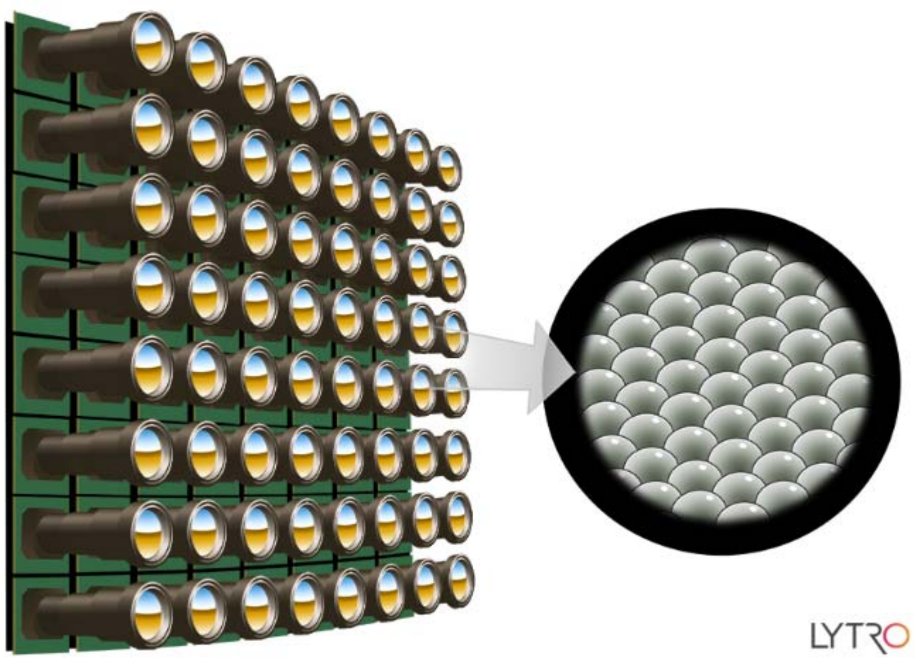


Light field geometry



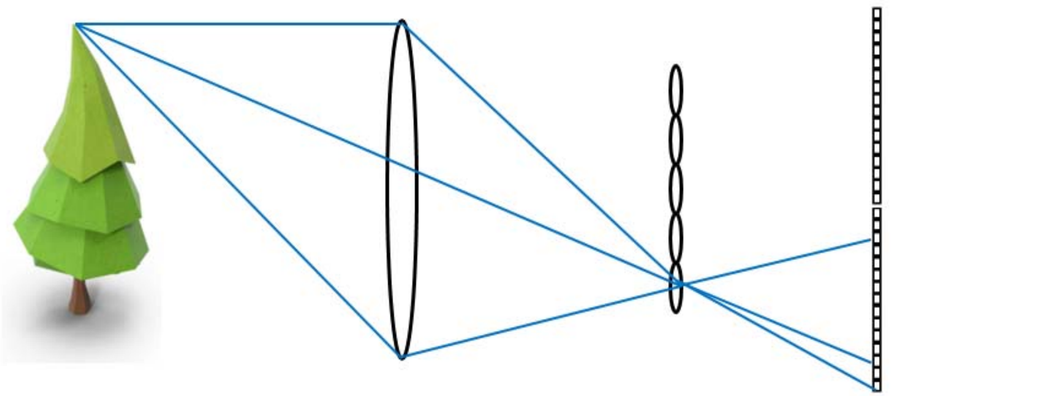
Light Field vs Multi-view Camera

- Structural similarity
  - Multi-view camera: multiple camera array
  - Light field camera: micro lens array
- Camera(lens) distance
  - Multi-view camera: wide
  - Light field camera: narrow

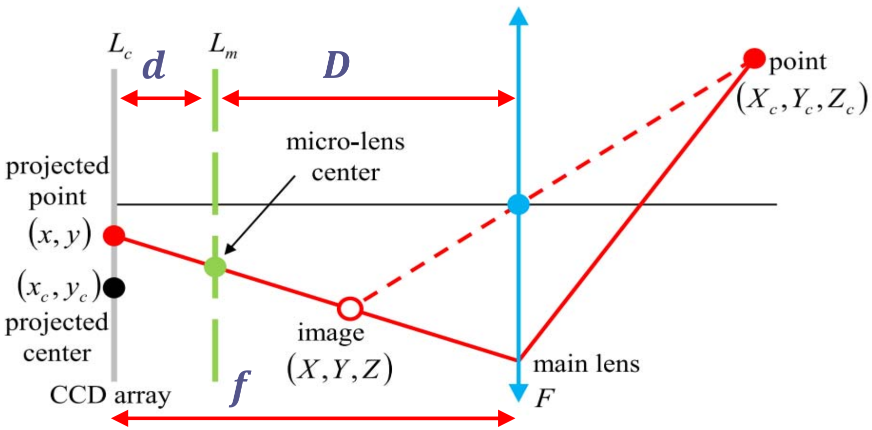


Optical Structure of Light Field Camera

- Main component of light field camera
  - Main lens, micro lens array, image sensor
- Definition of internal light field camera structure
  - D: distance between main lens and micro lens
  - d: distance between micro lens and image sensor
  - f: distance between main lens and image sensor



Optical direction of light field camera



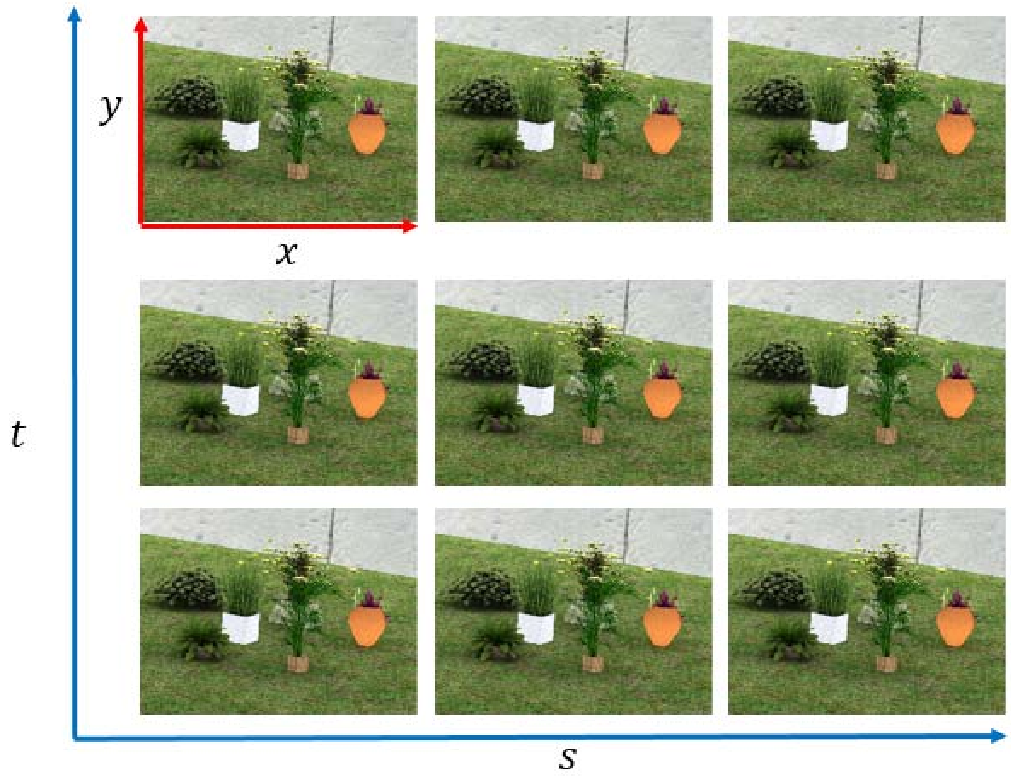
Lens relation in light field camera

4D Light Field

- ❖ Light field coordinate system
  - Composed of image(Ω) and camera plane(Π)

$$L : \Omega \times \Pi \rightarrow \mathbb{R}, \quad (x, y, s, t) \longmapsto L(x, y, s, t)$$

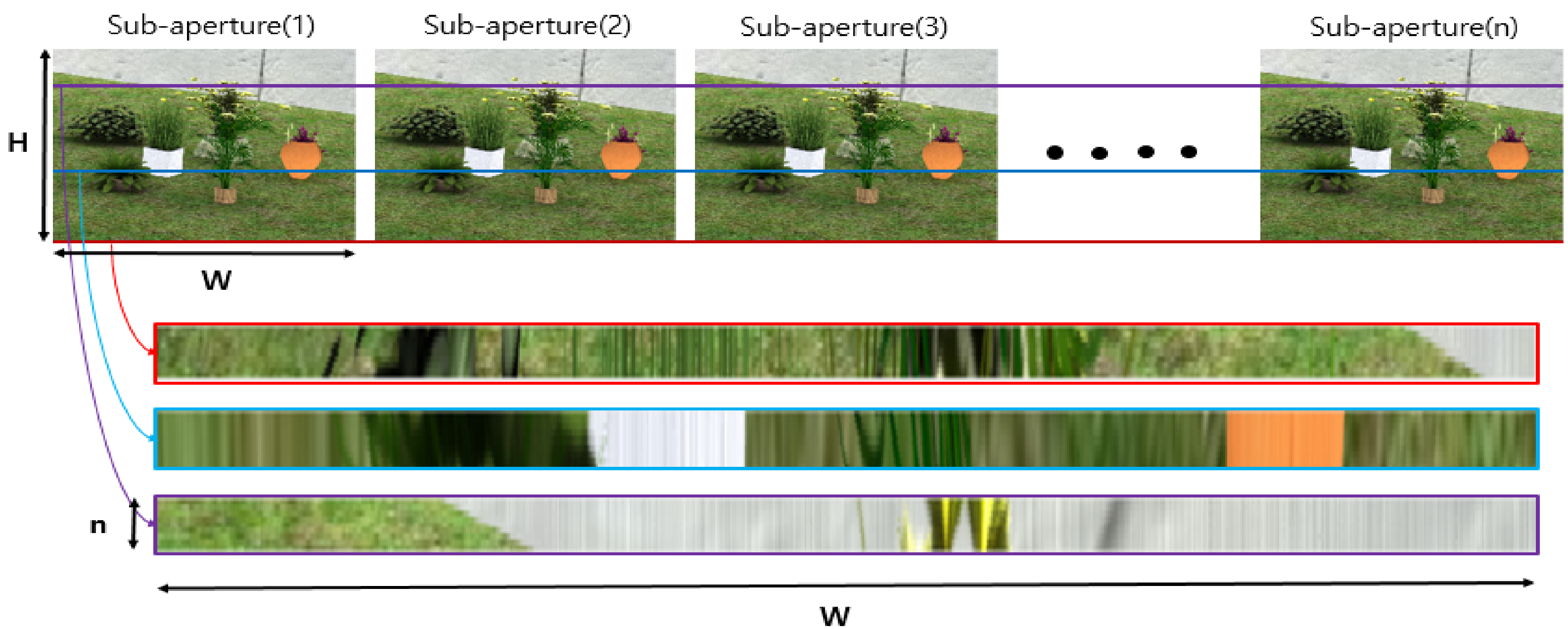
$$(x, y) \in \Omega\,, \ (s, t) \in \Pi$$



4D light field image structure

Light Field EPI Generation

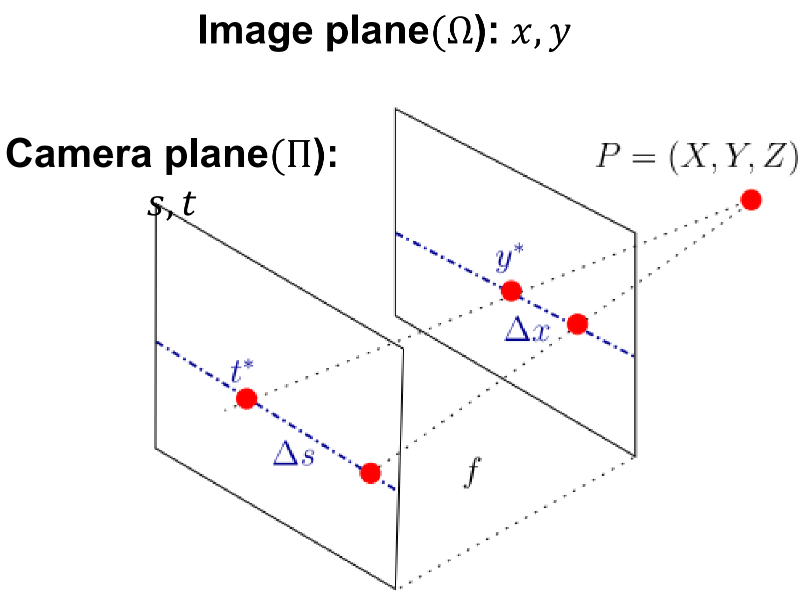
- Epipolar plane image (EPI)
  - Horizontally or vertically fixed line stacked image
- Characteristics of each EPI
  - EPI width = sub-aperture width
  - EPI height = number of sub-aperture
  - Normalize angular tendency for range stabilization

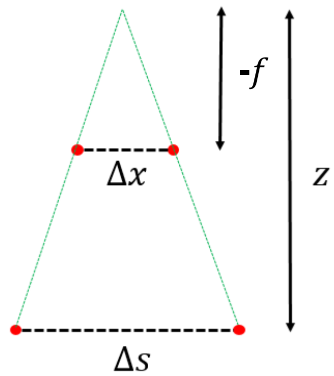


EPI data generation

Real Depth from Light Field EPI

❖ Triangular proportion





❖ Depth estimation

- Light field camera parameter (focal length)
- Coordinate variation in camera and image plane

$$\Delta x : \Delta s = -f : Z \quad \Rightarrow \quad \Delta x = -f\frac{\Delta s}{Z} \quad \Rightarrow \quad Z = -f\frac{\Delta s}{\Delta x}$$

Disparity of Light Field EPI

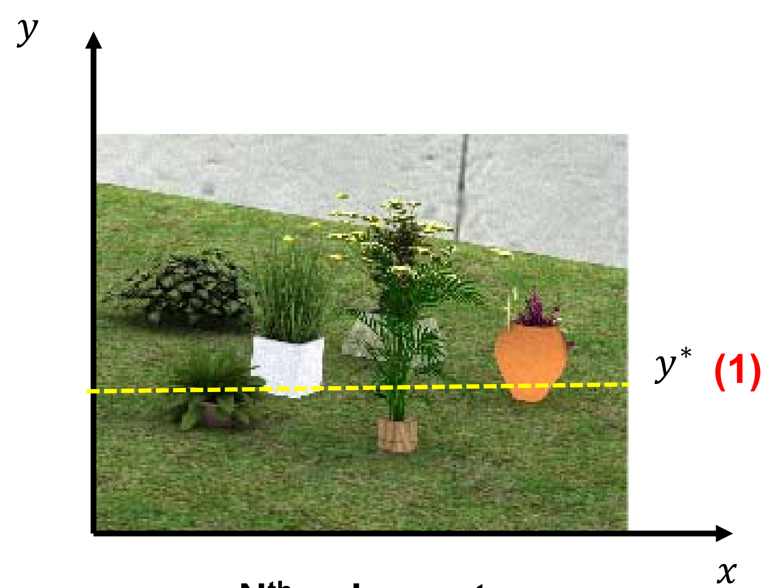
- Abbreviation of real depth
  - Interchangeability between real depth and disparity

$$Z = -f\frac{\Delta s}{\Delta x} \qquad \rightleftarrows \qquad d_{y^*,t^*} = \frac{f}{Z} = \frac{\Delta x}{\Delta s} = tan\theta$$

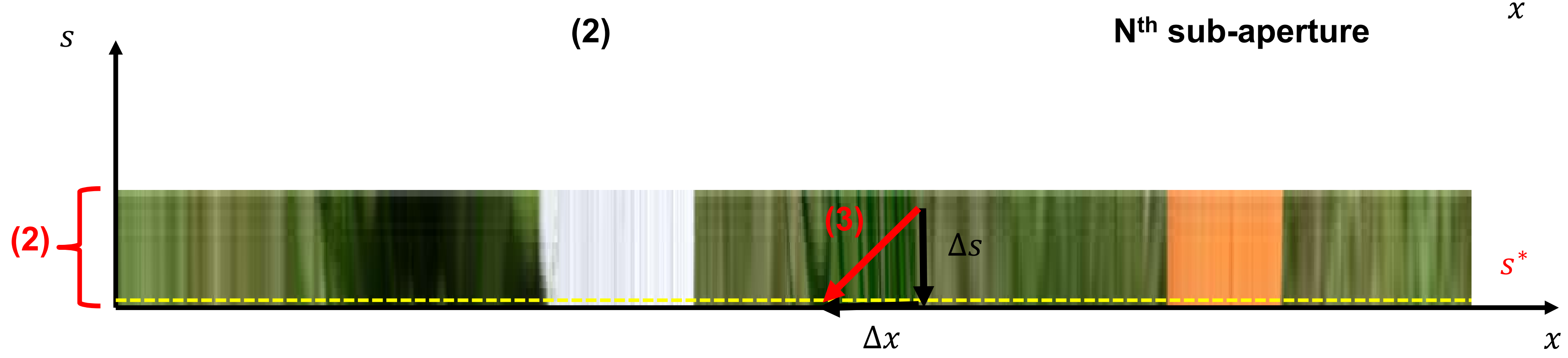
- Disparity from light field EPI
  - Need an angular value of EPI
  - Variation ratio determine angular value

Graphical Representation of EPI

- Procedure for disparity estimation
  - **(1)** Assume a fixed row line y^*
  - **(2)** Stacking row line of each sub-aperture image
  - **(3)** Measuring the angular value(θ) of EPI for each pixel



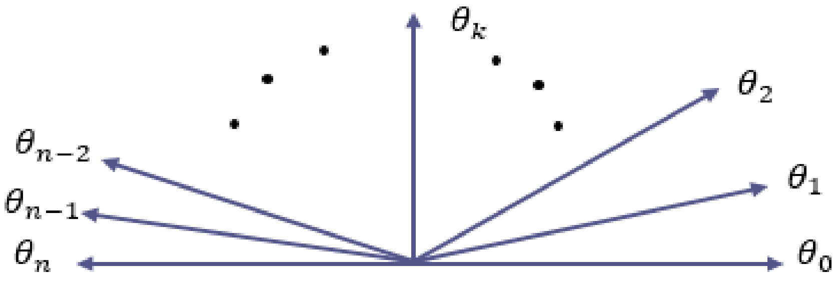
N^th sub-aperture



EPI of horizontally fixed(y^*) coordinate

Cost Computation for Optimal Angle

- Inspect all angular directional candidate
  - Find an optimal angular value for depth estimation[1]
  - Angular search range from $0^{\circ}(\theta_0)$ to $180^{\circ}(\theta_n)$



Angular search range

$$\theta(l, \mathrm{P}) = arg \min_{\theta \in A} C(E_l, P)$$

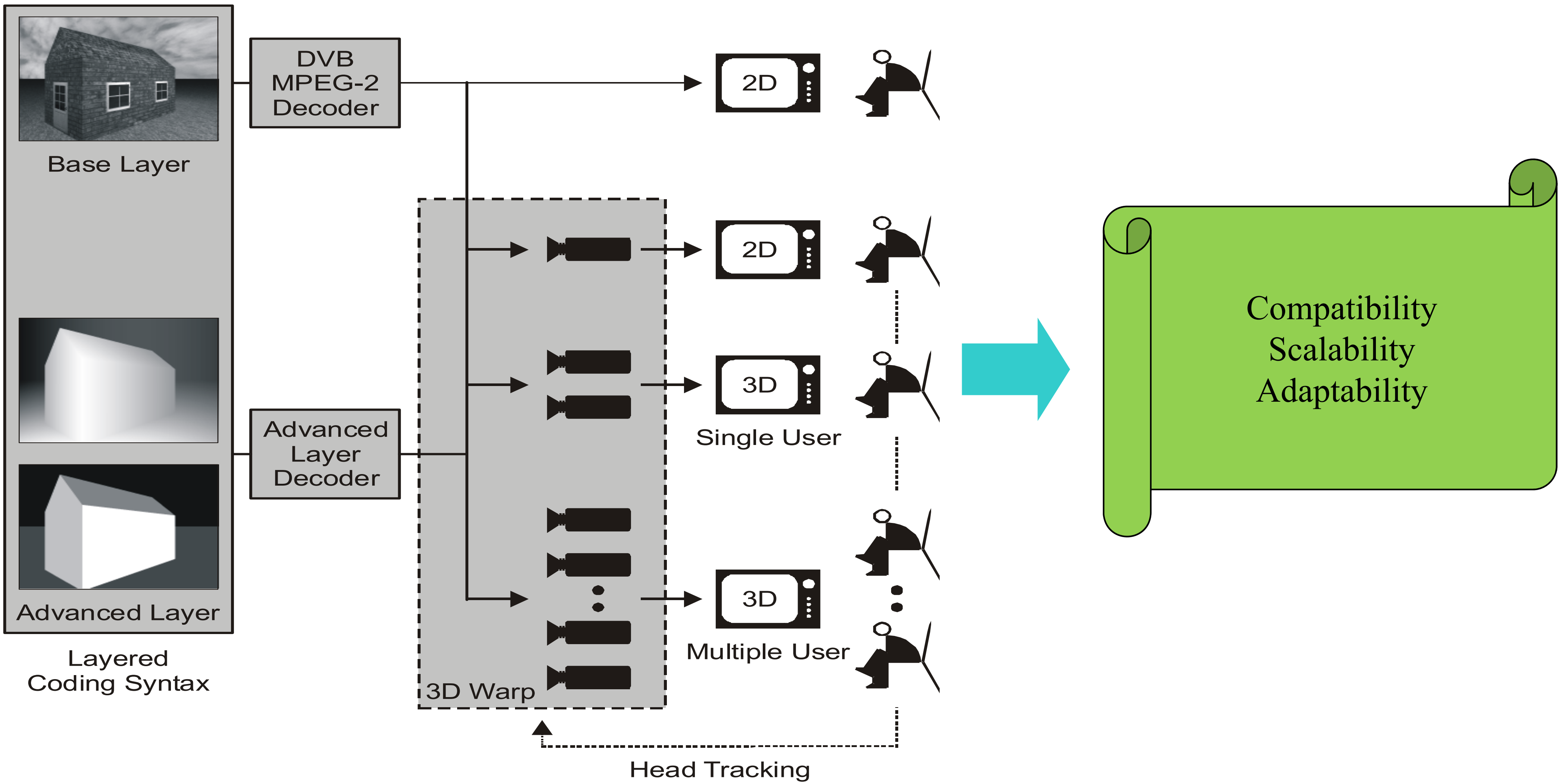
- Cost function for optimal angle selection

$$\mathrm{C}(E_l, P) = \frac{1}{N(\theta)} \sum_{P_n \in N(\theta)} \begin{matrix} (1-\alpha)|E(P_n) - E(P)| \\ +\alpha[|E_x(P_n) - E_x(P)| + |E_y(P_n) - E_y(P)|] \end{matrix}$$

- $N(\theta)$: number of neighbor pixels
- E_x $and E_y$: x and y directional EPI gradient

[1] Y. Zhang, H. Lv, Y. Liu, and H. Wang, "Light-field Depth Estimation via Epipolar Plane Image Analysis and Locally Linear Embedding," IEEE Trans. On CSVT, vol. 27, no. 4, April 2017.

Broadcast 3-D Video

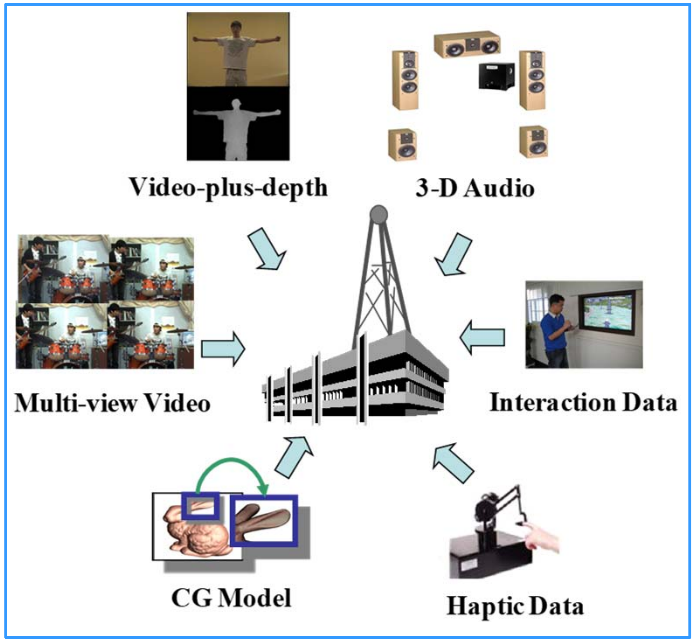


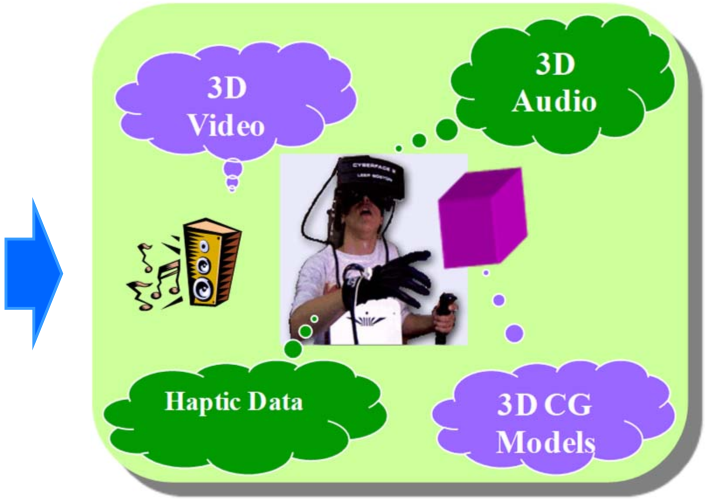
3-D Television System

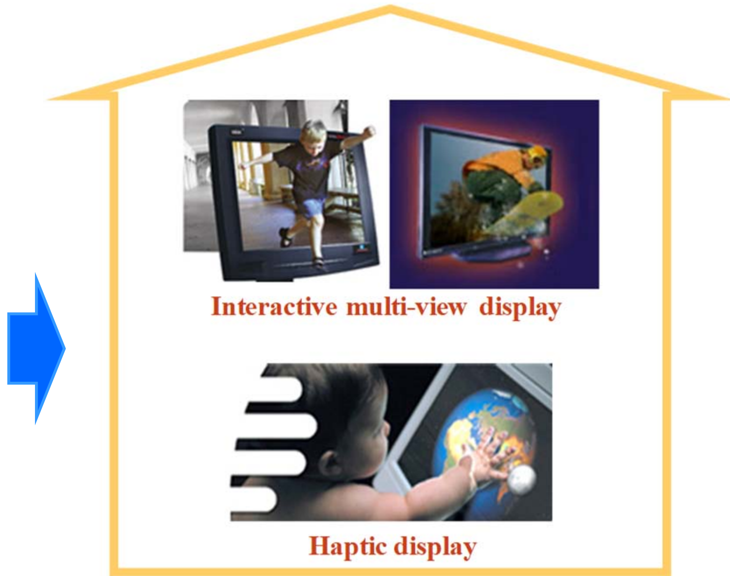
Capture → Correction/ Conversion → Encoder → Transmission/ Storage → Decoder → Interpolation → Display

- 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

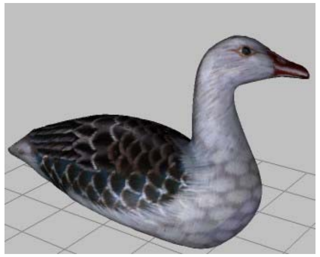
- Conventional 2-D audio-visual information
  - Not enough to give us vivid feeling through our five senses
- Multi-modal immersive media
  - Data overcoming the spatial-temporal limits
  - 3-D video, multi-channel audio, computer graphic data, haptic data
- 3-D video is an essential part of future multimedia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

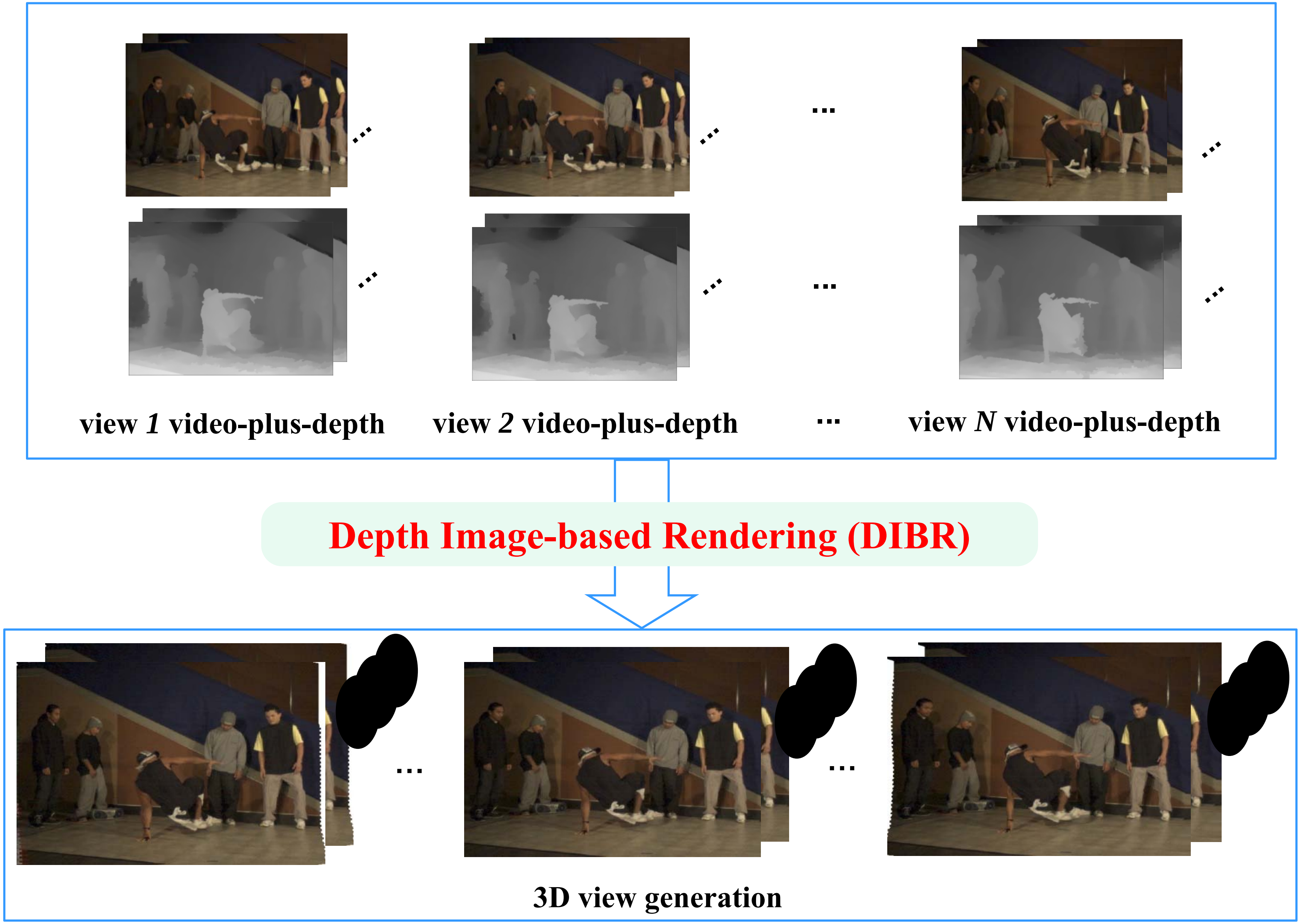
Color image sequence | Depth map sequence

※ 'Breakdancers' sequence from MSR

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

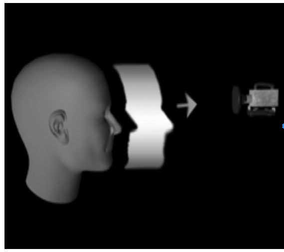
Backward-compatibility and easy adaptability to the current 2-D digital system

Multi-view Video-plus-depth



Time-of-flight Depth Camera

Z-Cam developed by 3DV System, Ltd



Imprinted light wall

Front cut of light wall

Rear cut of light wall

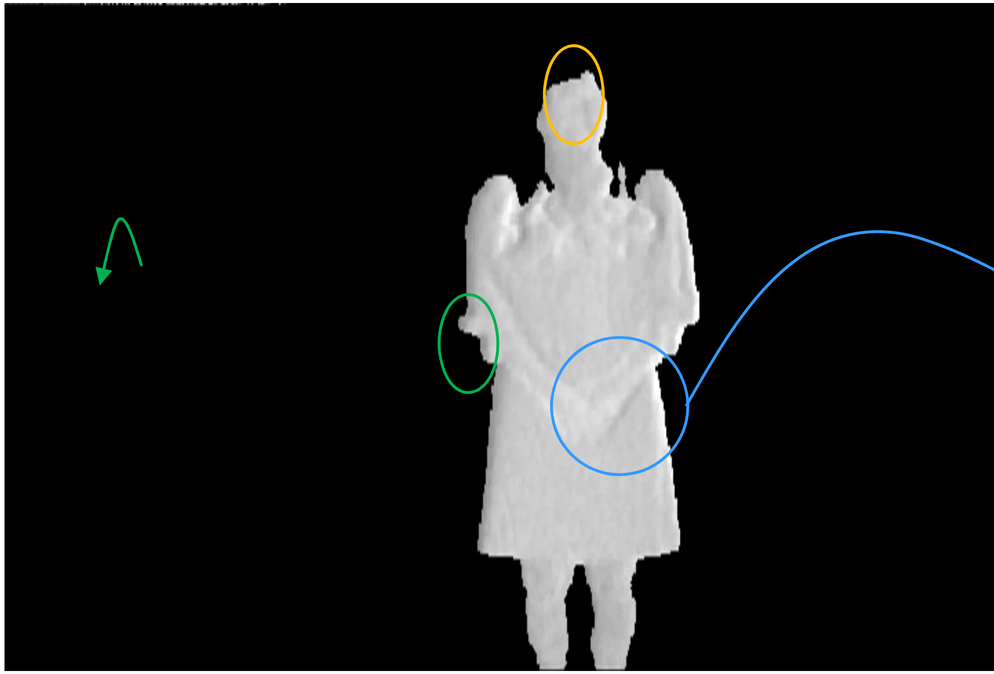
※ G.J. Iddan and G. Yahav, "3D Imaging in the Studio and Elsewhere…,"
Proc. of SPUE Videometrics and Optical Methods for 3D Shape Measurements, 2001.

Real-time capturing of synchronized color and depth information

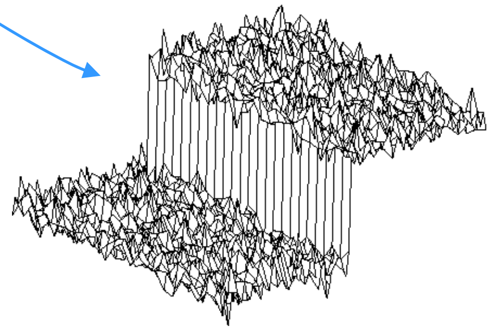
Problems in Depth Map

No depth information

Boundary mismatching

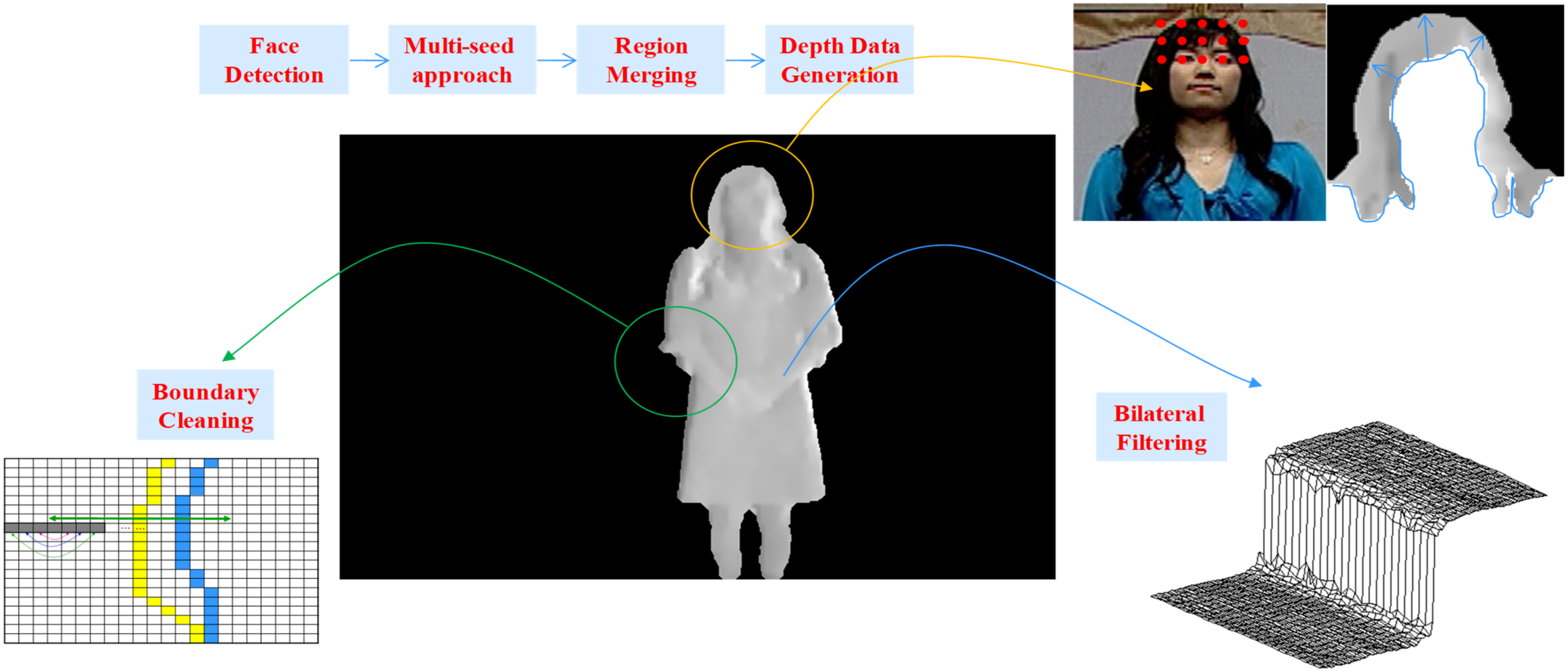


Original depth map captured by a depth camera

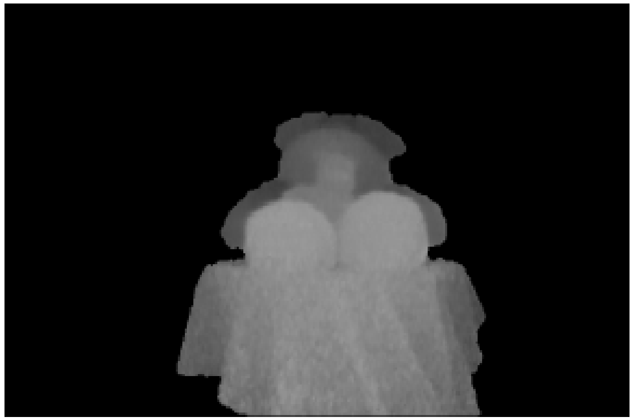


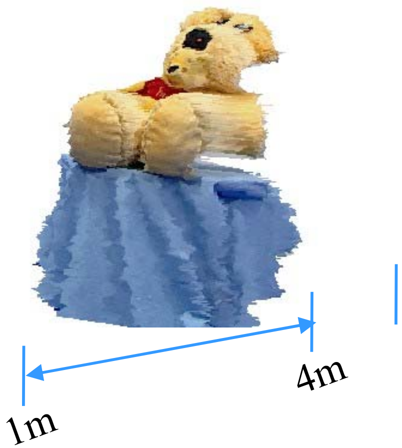
Optical noises
& No calibrated depth

Depth Map Enhancement



Limitation in Depth Measurement





Low-resolution Depth Map

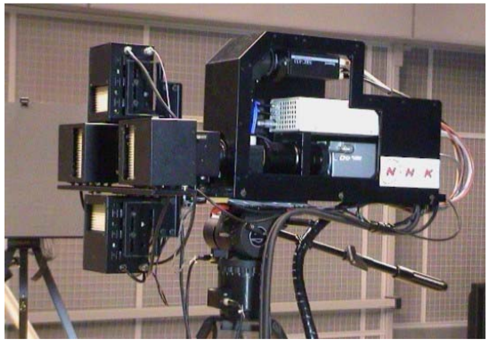
Z-Cam

720

486

720

486



NHK Axi-vision HDTV Camera

1280

720

1280

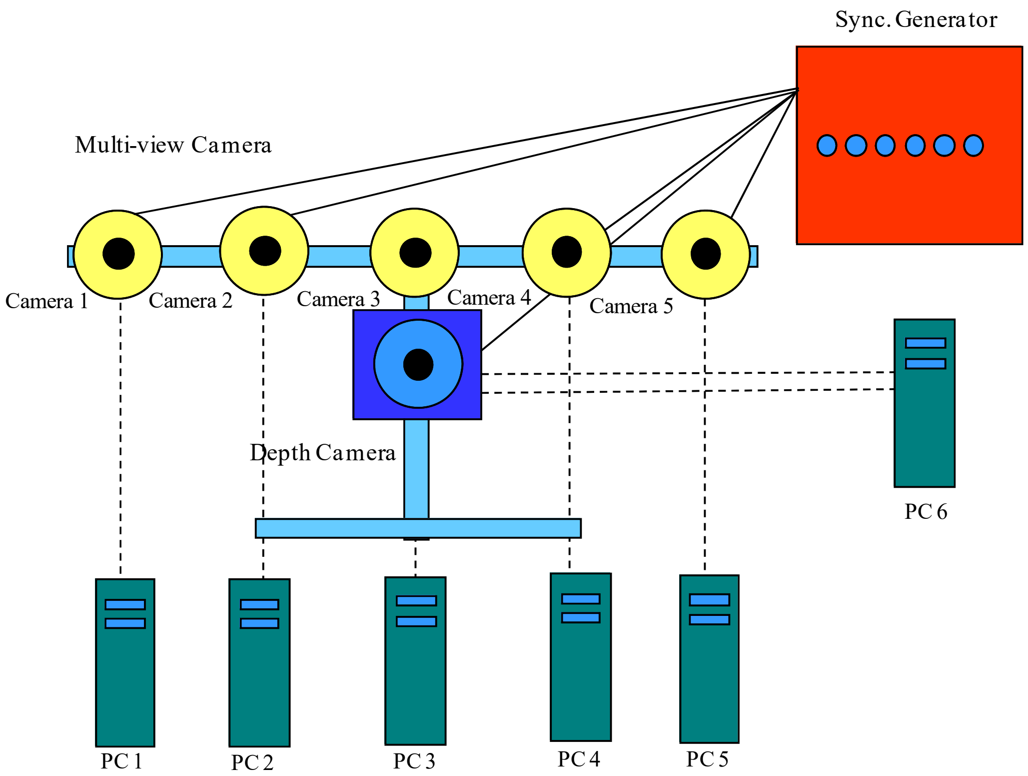
720

※ 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

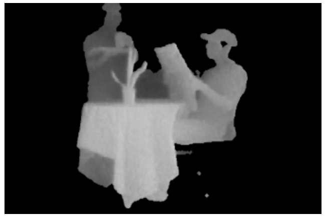


Device	Specification	Details
Multi-view camera (Cannon XL-H1)	Output Format	NTSC or PAL (16:9 ratio, High Definition)
Depth Camera (Z-Cam)	Depth Range	0.5 to 7.0m
	Field of View	40 degrees
	Output Format	NTSC or PAL (4:3 ratio, Standard Definition)
Sync. Generator (LT 443D)	Output Signal	SD/HD Video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

Baseline distance between two cameras: 20cm

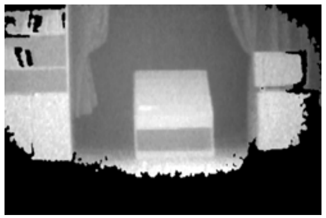
Captured Images by HCS

Color Image
(720x486)



Depth Map
(720x486)

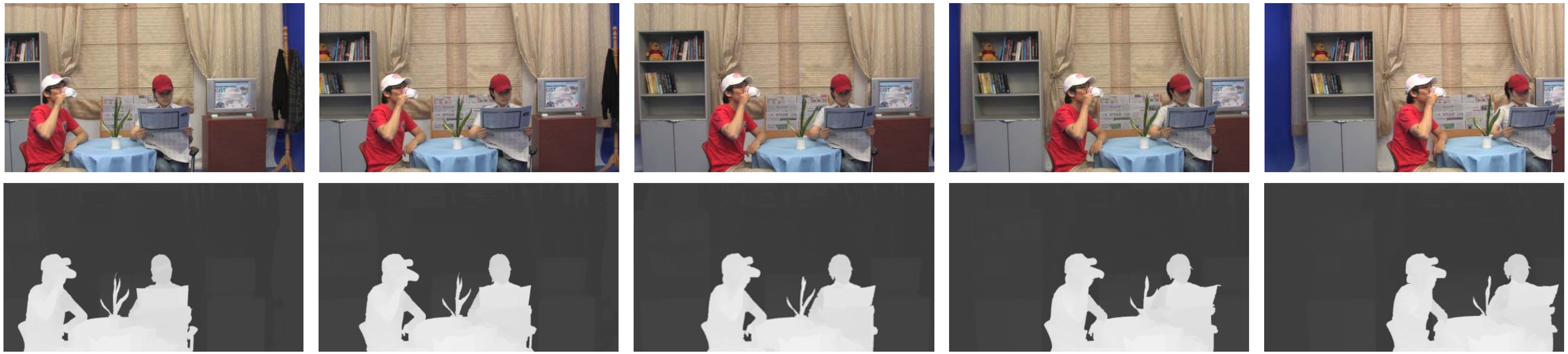
Background Color Image
(720x486)



Background Depth Map
(720x486)

Multi-view Image (1920x1080)

Output Images from HCS



view 1 view 2 view 3 view 4 view 5

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

(u_0,v_0) : Coordinate of the principal point

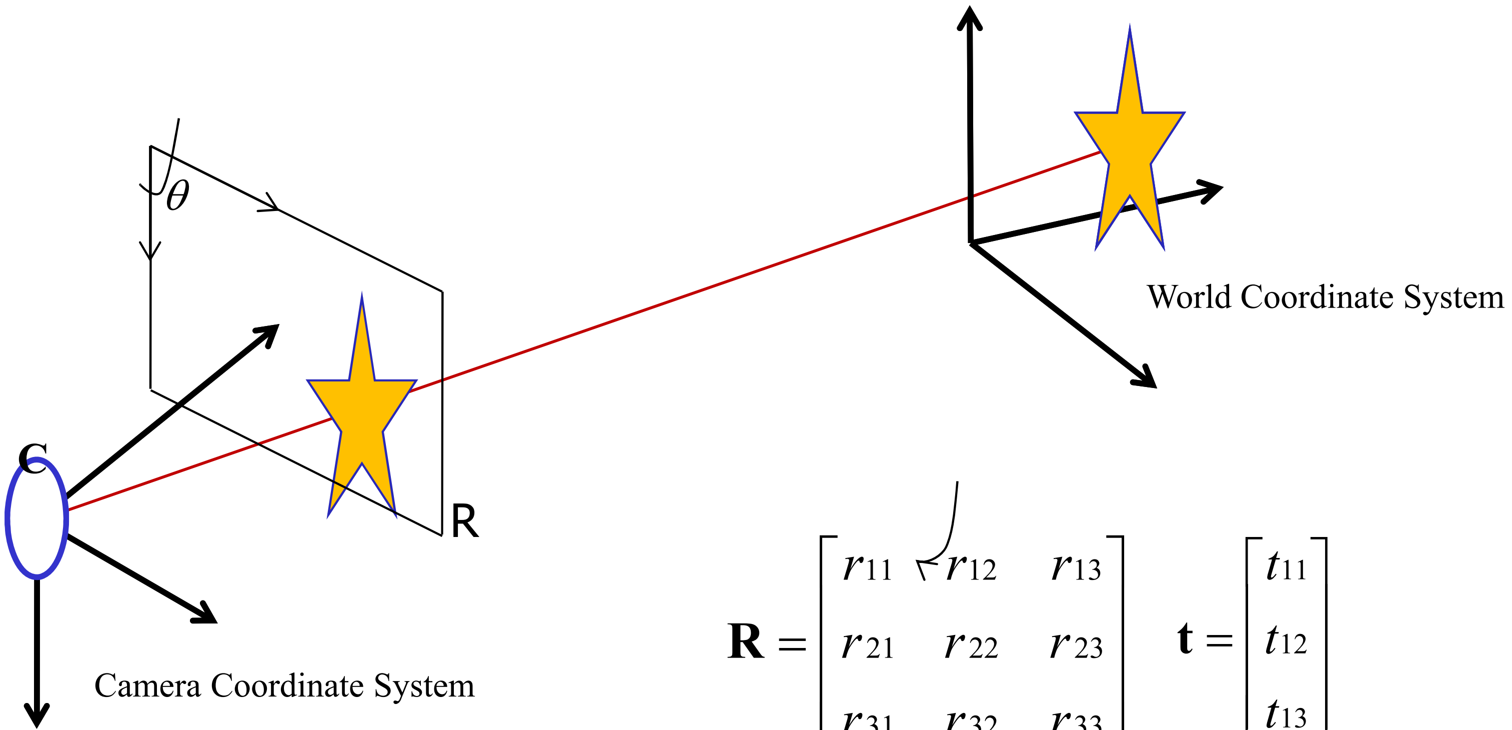
- Principal point: intersection between the Z-axis (principal axis) of the camera coordinate system and the image plane

γ : Skew parameter

- Non-orthogonality between u and v axes

Camera Extrinsic Parameters

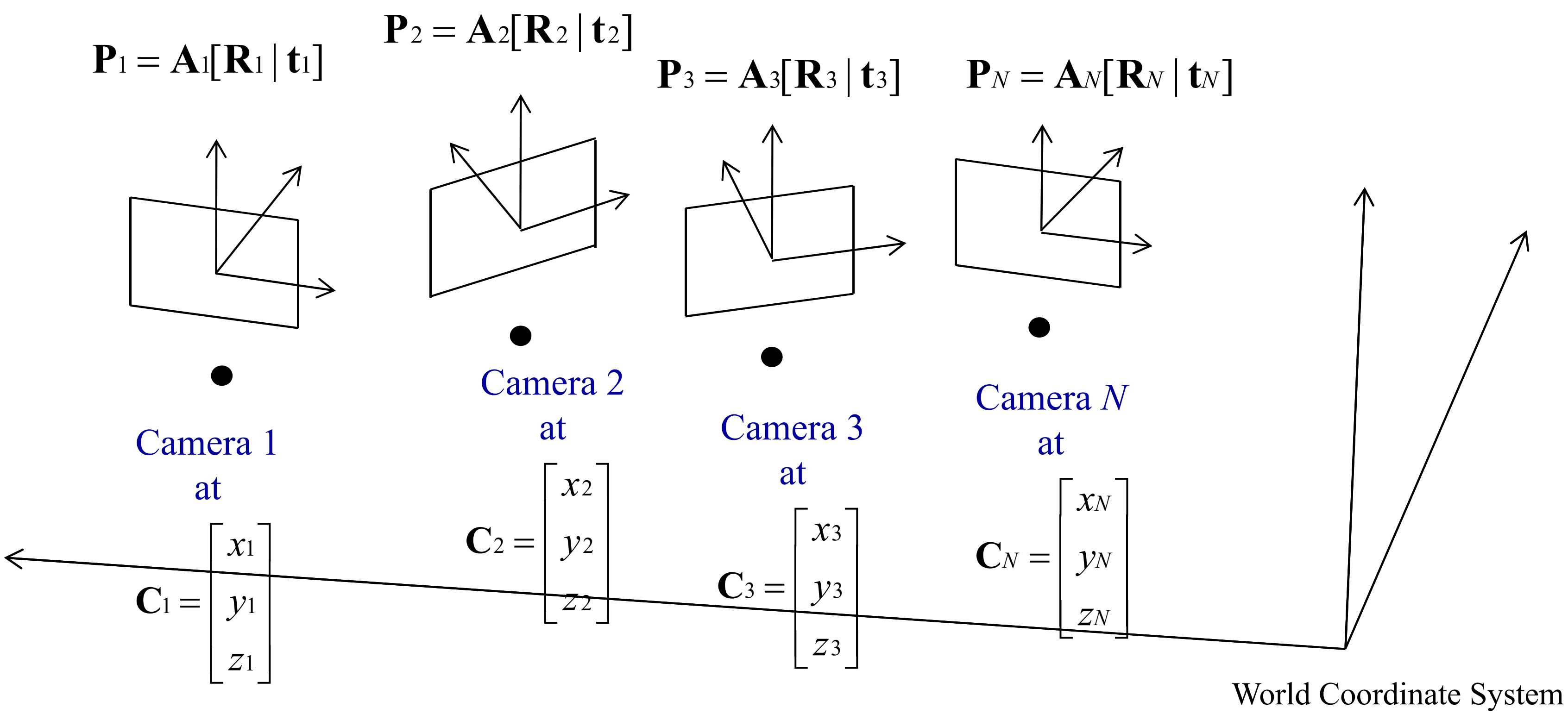
- Represented by
  - 3x3 rotation matrix **R**
  - Translation vector **t**
- Indicate camera orientation and position



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

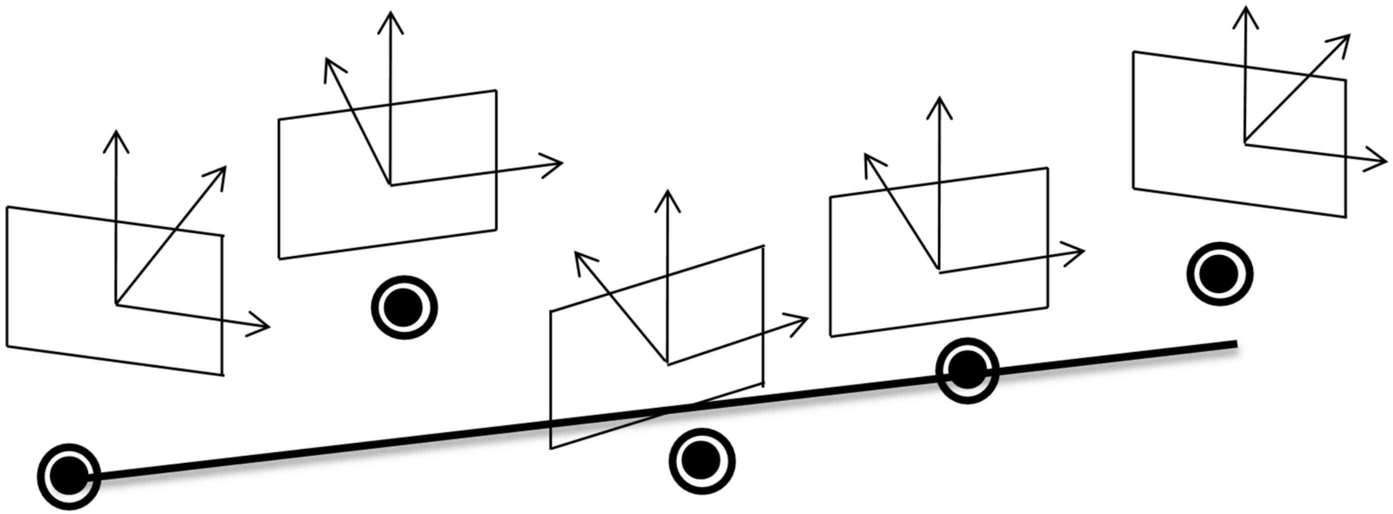
Camera Parameter Extraction

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



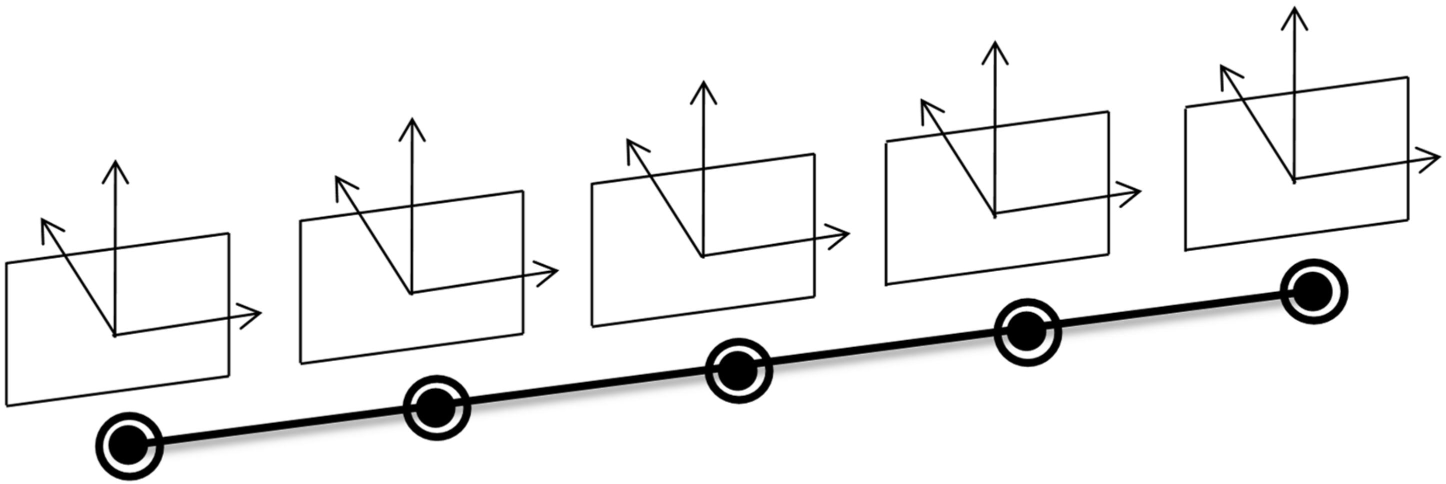
Problems of 1-D Parallel Camera Array

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



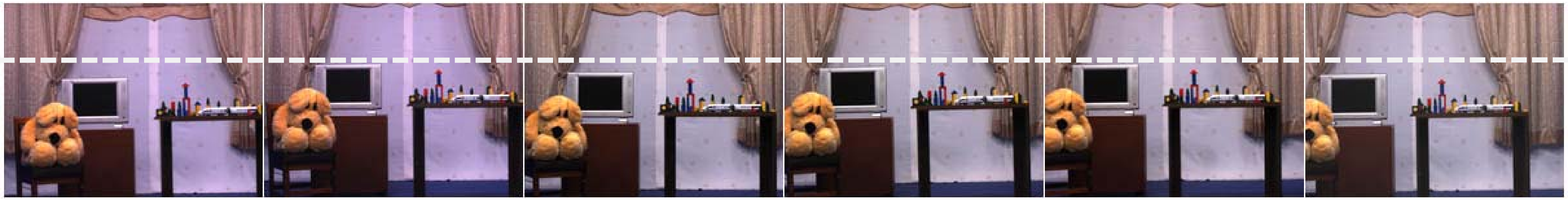
Ideal 1-D Parallel Camera Array

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

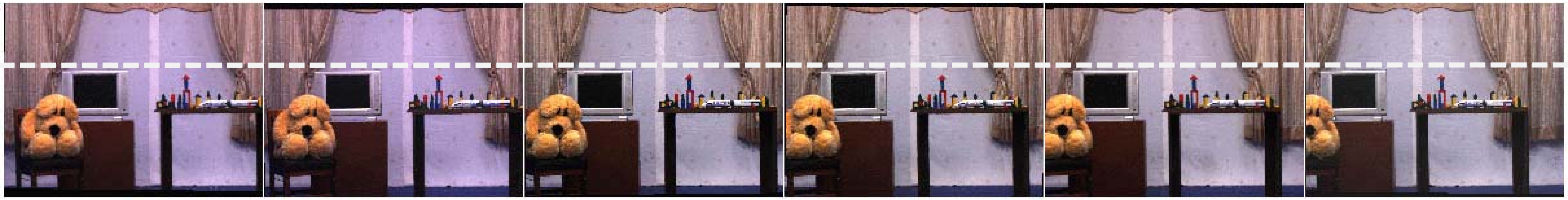


Multi-view Image Rectification

- 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



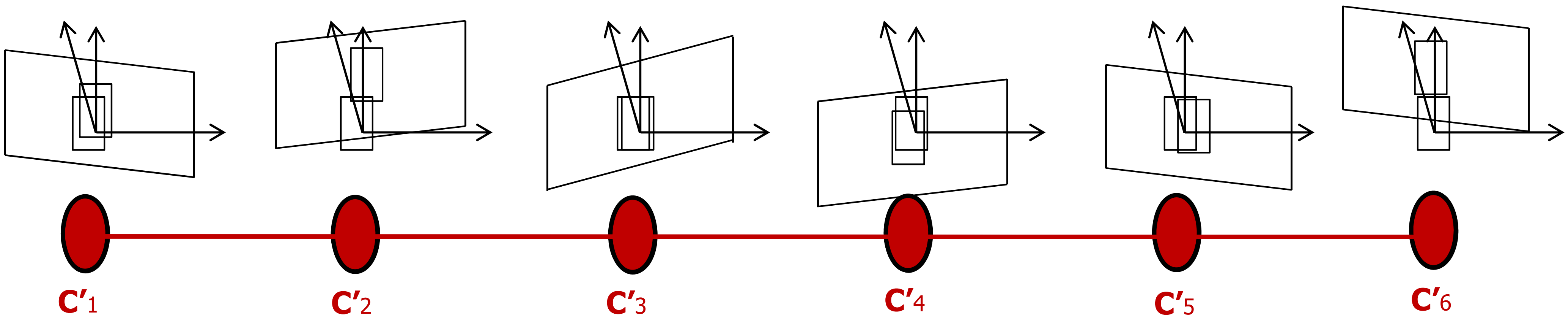
Before



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



Color Inconsistency Problem

Single

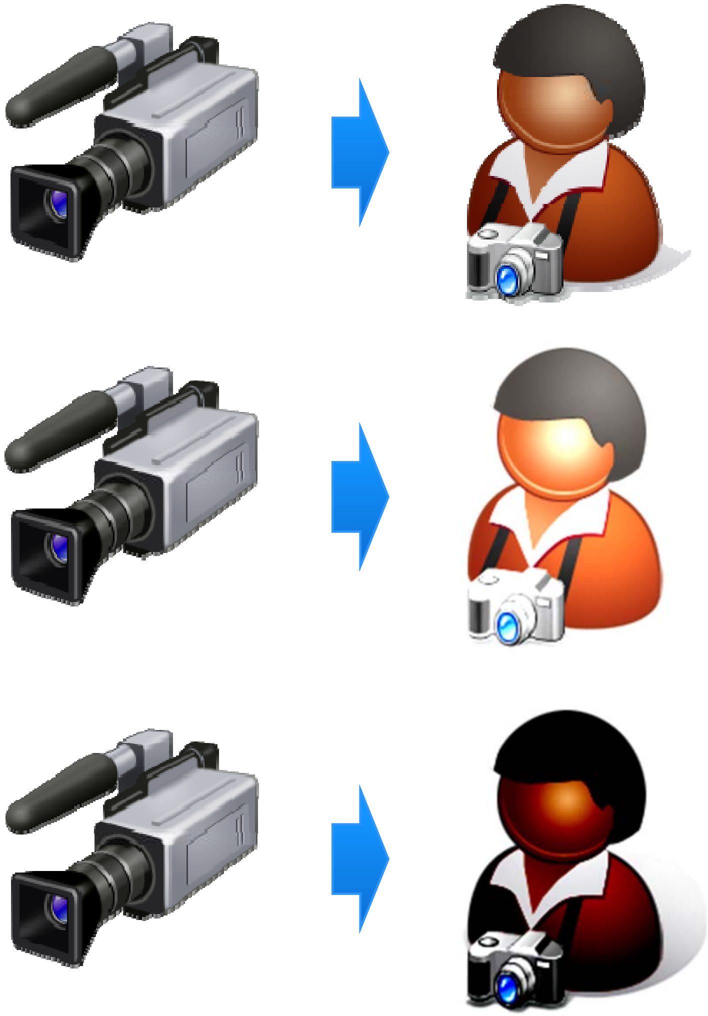
Camera System

Multi-view

Camera System

Color Mismatch Problem

Original Image



Camera Properties : Gain, Offset, Gamma, White Balance

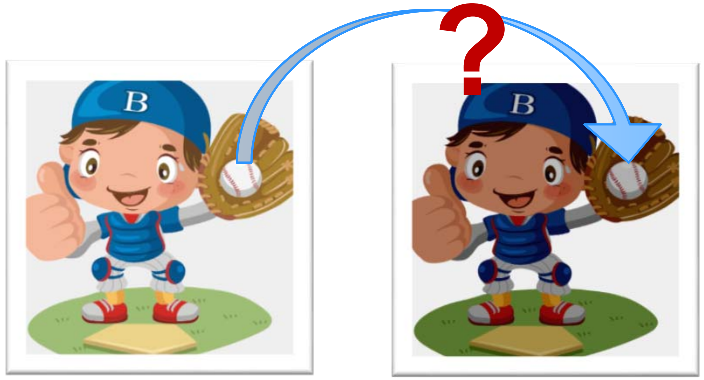
Example of Color Mismatch

Race

Uli

Related Issues

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

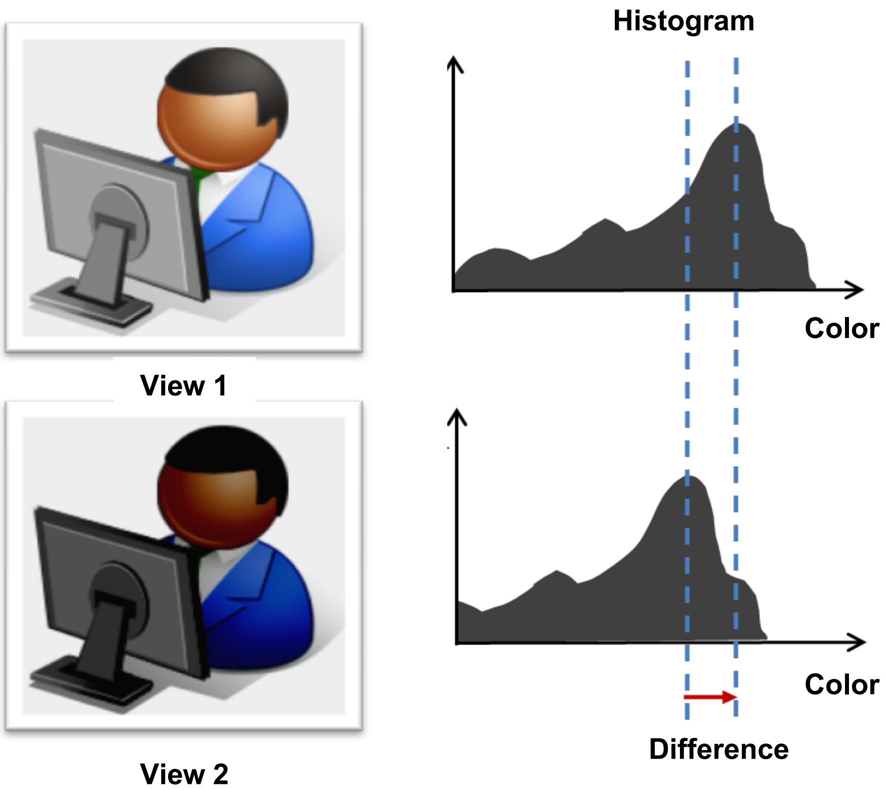


- View Synthesis
  - Refer to neighboring images
  - Cause of stain

Correcting Methods

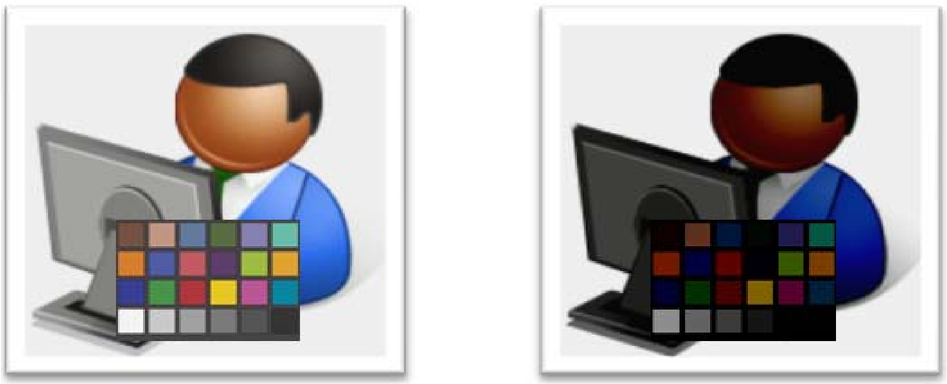
1. Global Property

- Average brightness, histogram



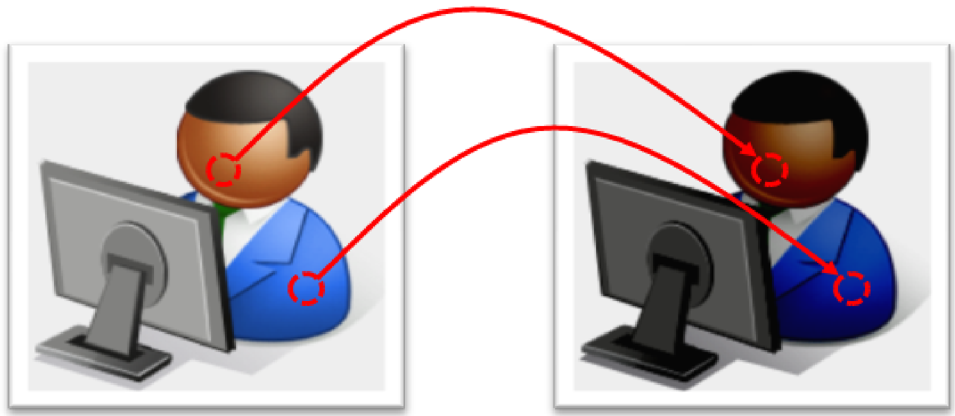
2. Corresponding Property

- Color Chart



View 1 View 2

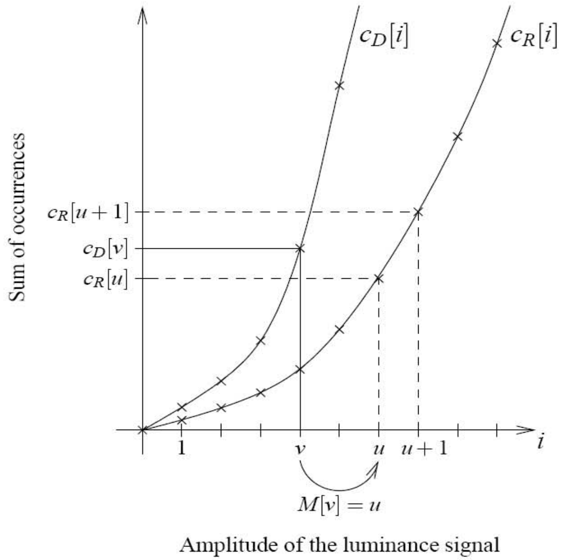
- Correspondence

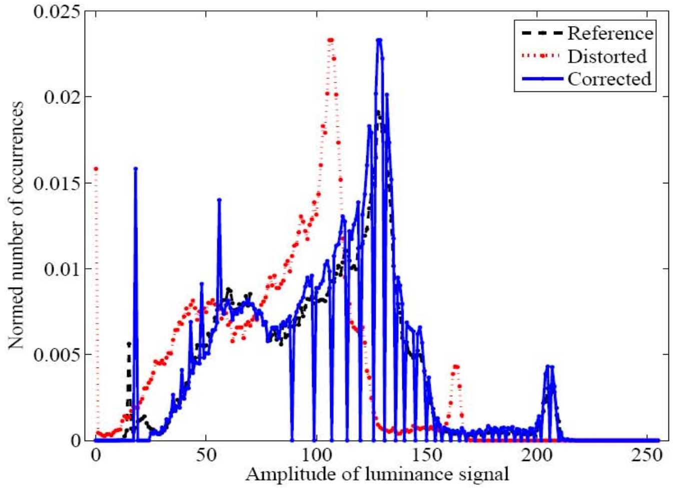


View 1 View 2

Histogram Matching

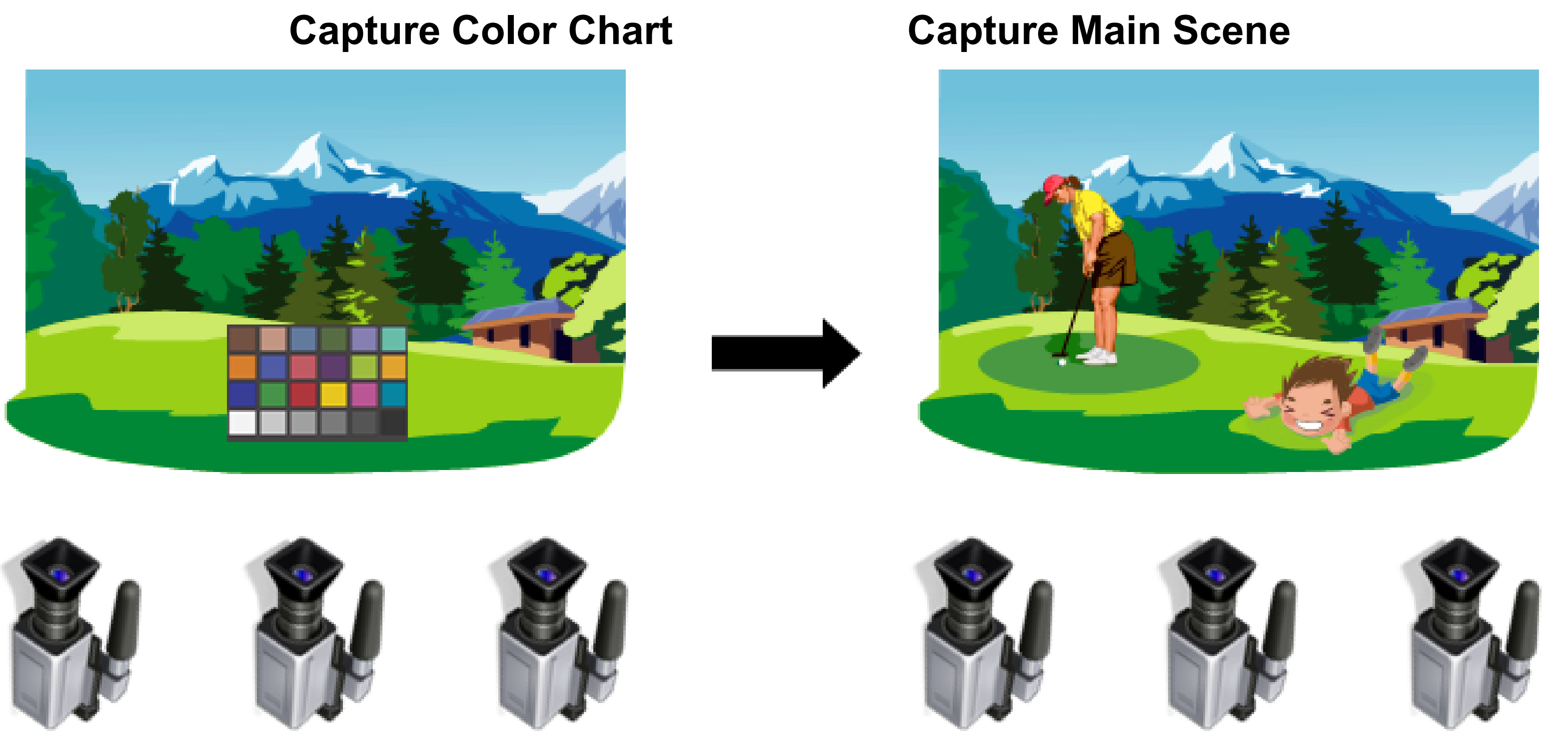
$$M[v] = u \quad \text{with} \quad c_R[u] < c_D[v] \leq c_R[u+1]$$





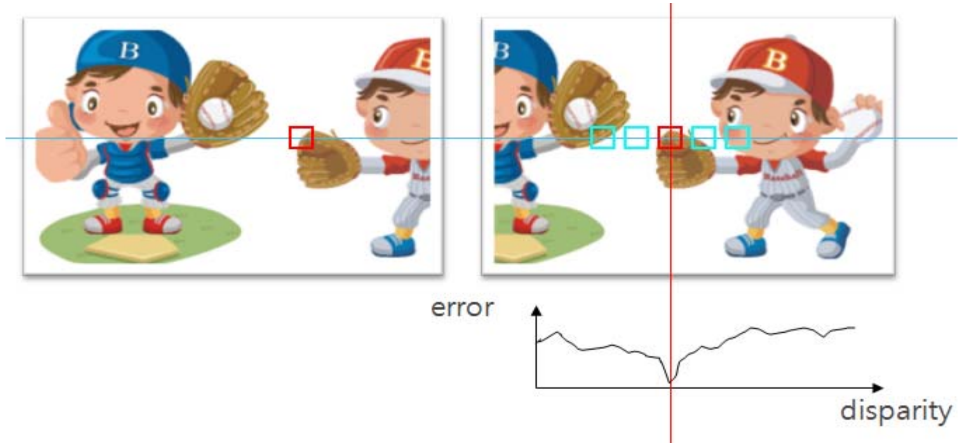
Color Chart

❖ Capturing the Standard Color Chart



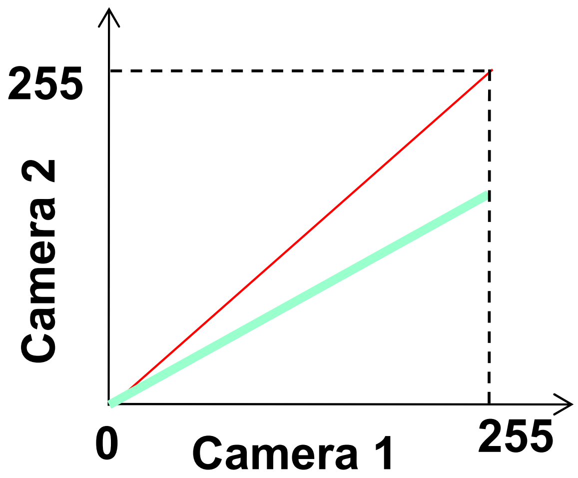
Correspondence

❖ Block-based Method
- Dense correspondence
- Low accuracy



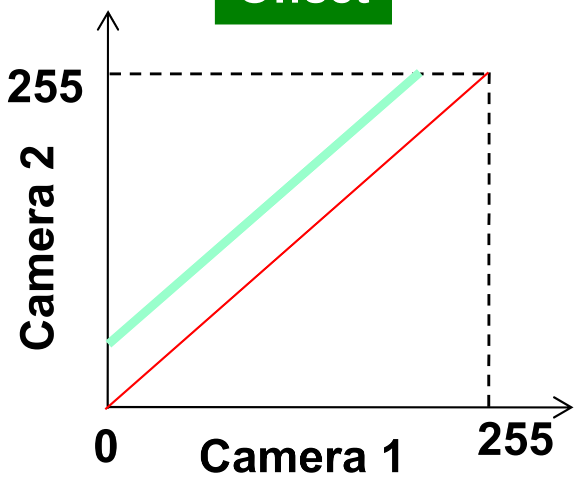
❖ Feature-based Method
- Sparse correspondence
- High accuracy

Camera Characteristic Curve



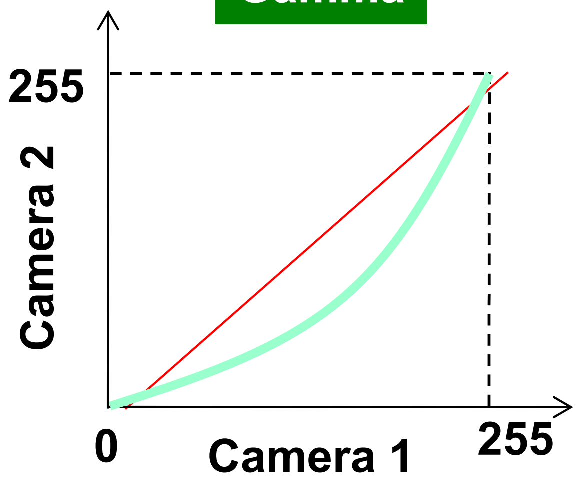
$$P_{ref} = C_{gain} \times P_{tar}$$

Offset

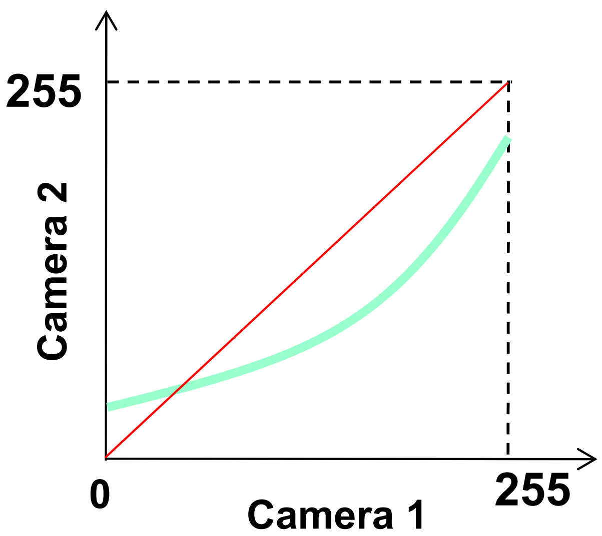


$$P_{ref} = P_{tar} + C_{offset}$$

Gamma



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



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

Corresponding Points

❖ Using SIFT Algorithm

View 3 View 4

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
...	...	...

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

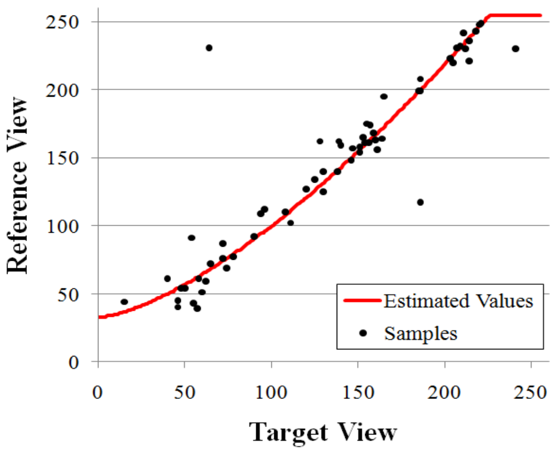
$$\overline{\beta}^{s+1} = \overline{\beta}^{s} + \delta\overline{\beta}$$

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

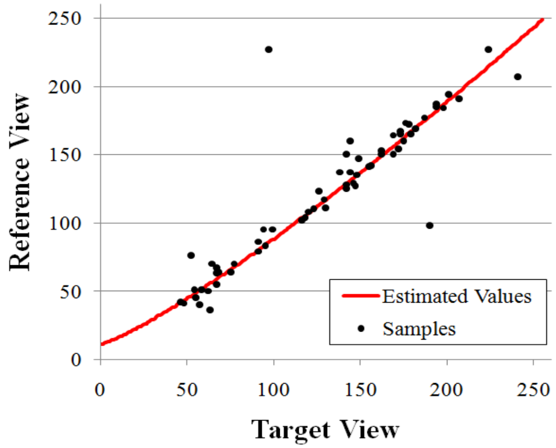
  - Minimize Error

Results of Regression

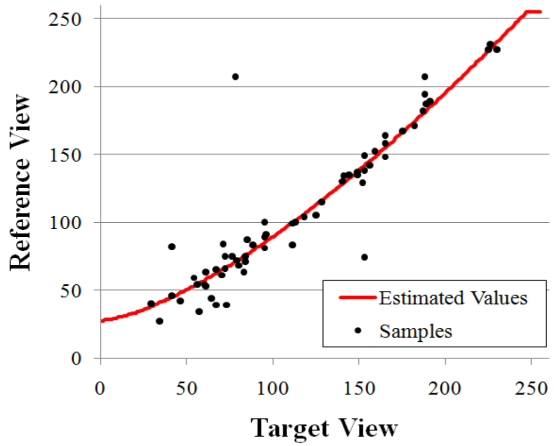
R Channel



G Channel



B Channel



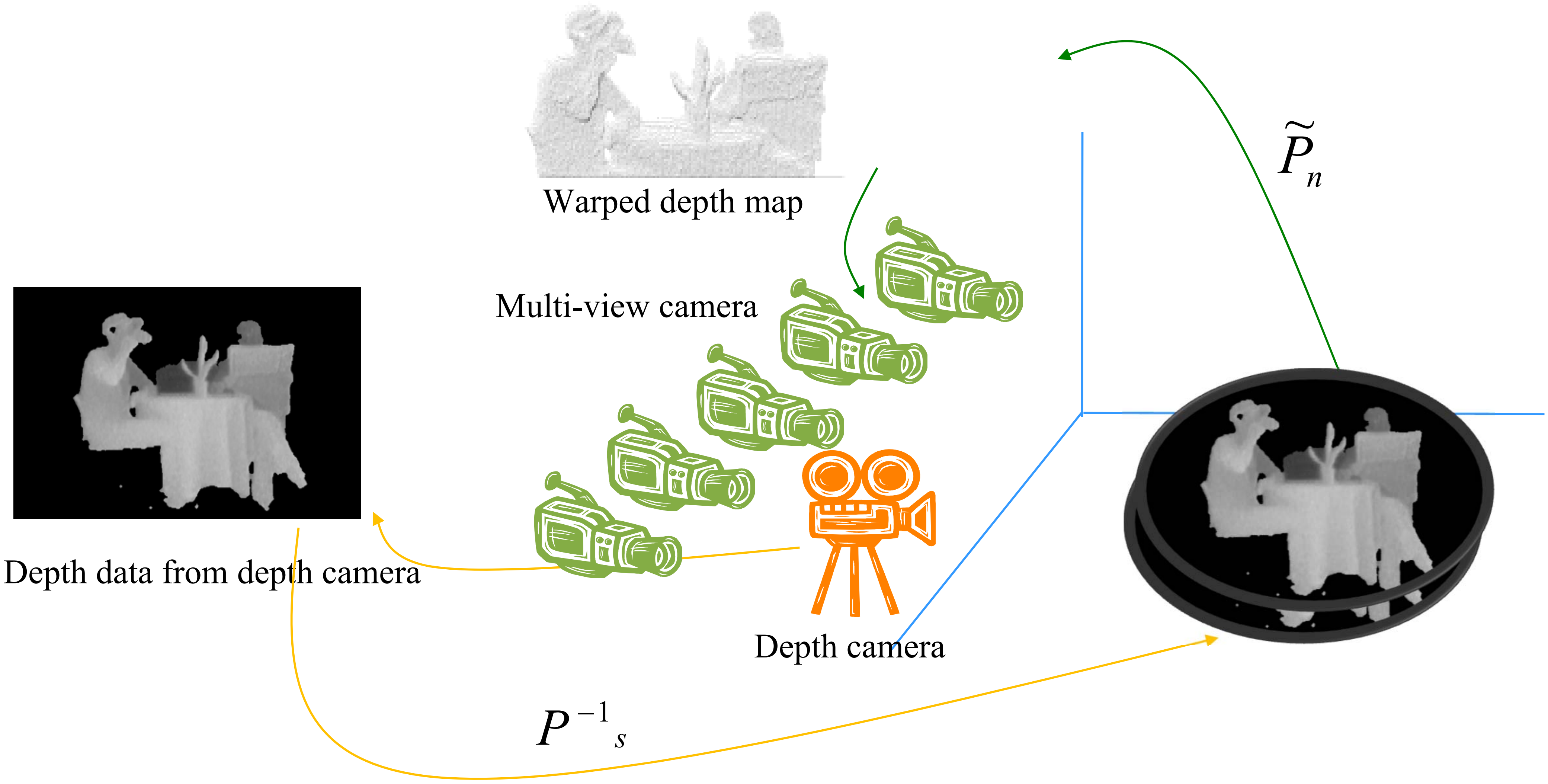
$$P_{ref} = C_{gain}\{P_{tar}/(2^{bitdepth}-1)\}^{C_{gamma}} \times (2^{bitdepth}-1) + C_{offset}$$

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

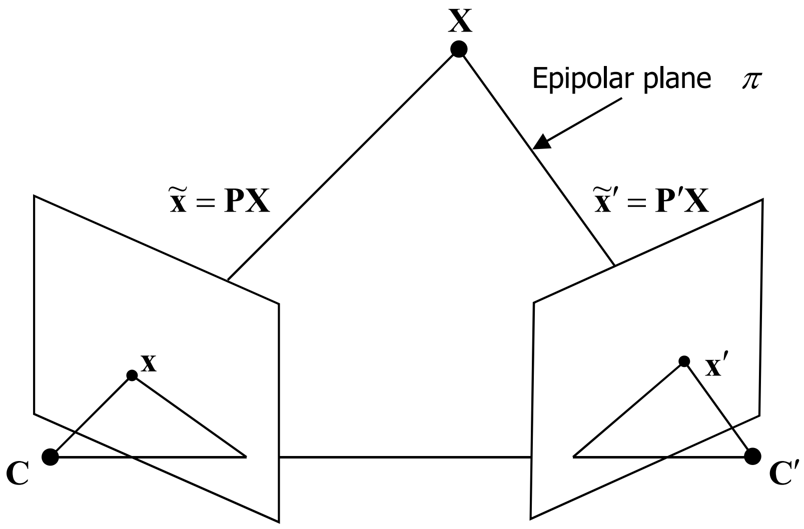
Experimental Result

3-D Warping in HCS

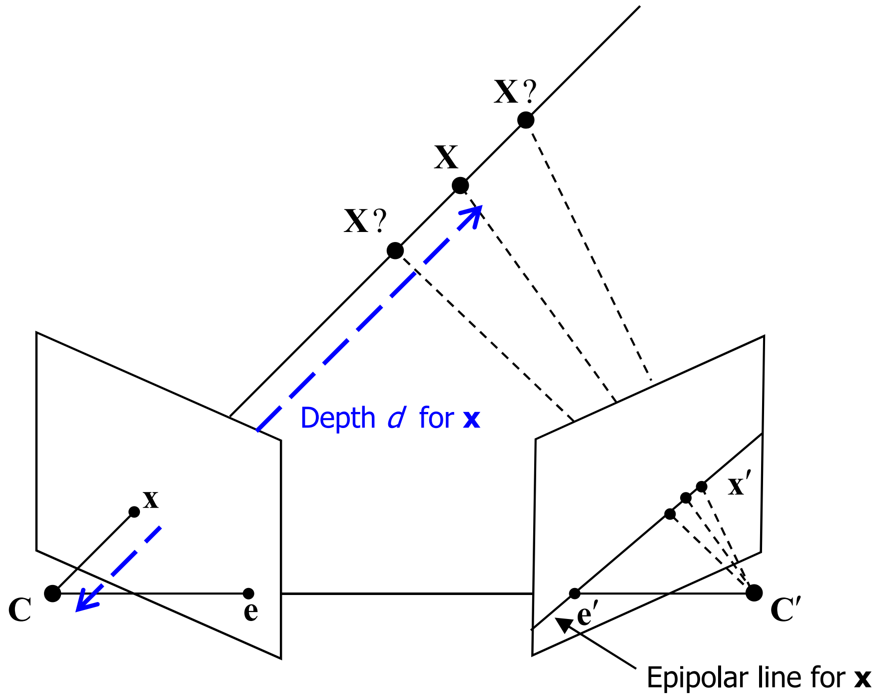


3-D Warping

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



Forward projection



Backward projection

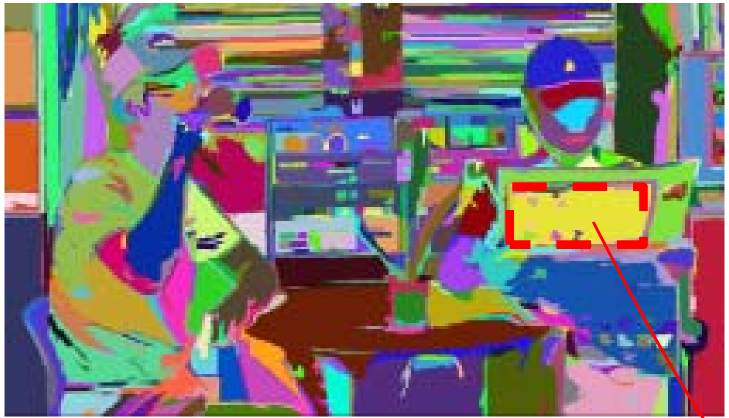
Color Segmentation

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.

Initial Disparity for Each Segment



Color segmentation

Warped disparity map

Initial disparity for each segment is set to the average value of its corresponding warped disparities

Segment-based Stereo Matching

One segment

Assignment of Initial disparity

- Segmentation-based stereo matching using SAD
- Initial SAD from 3-D warping results
- Min Disp SAD in a small search range

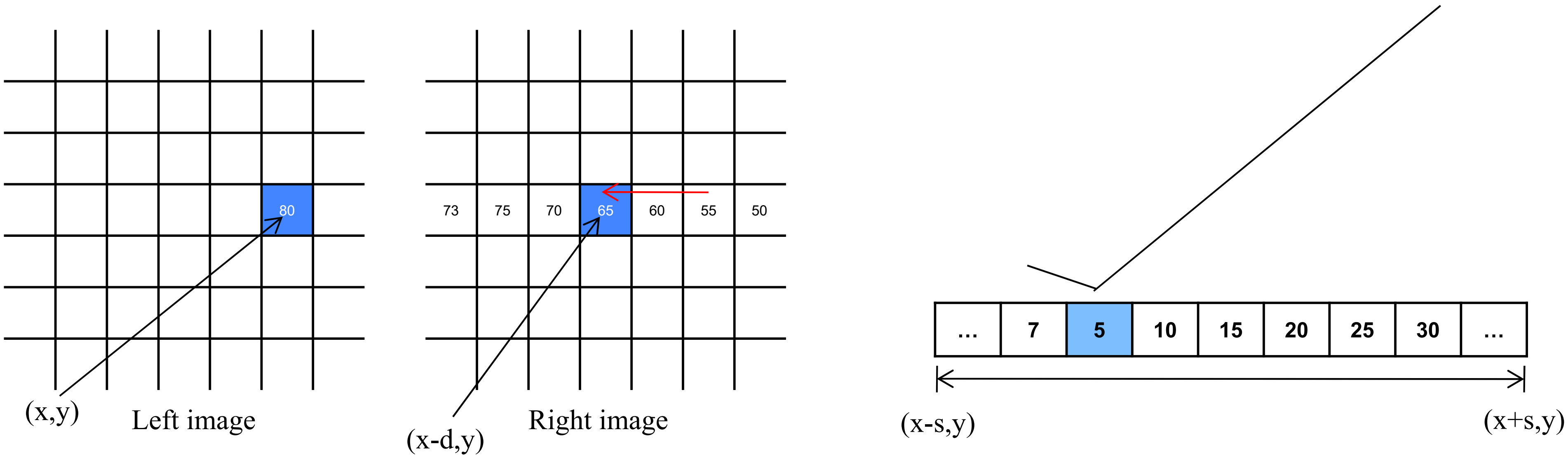
  Search range is set from (*initial disparity*-5) to (*initial disparity* +5)

Refinement of Disparity Map

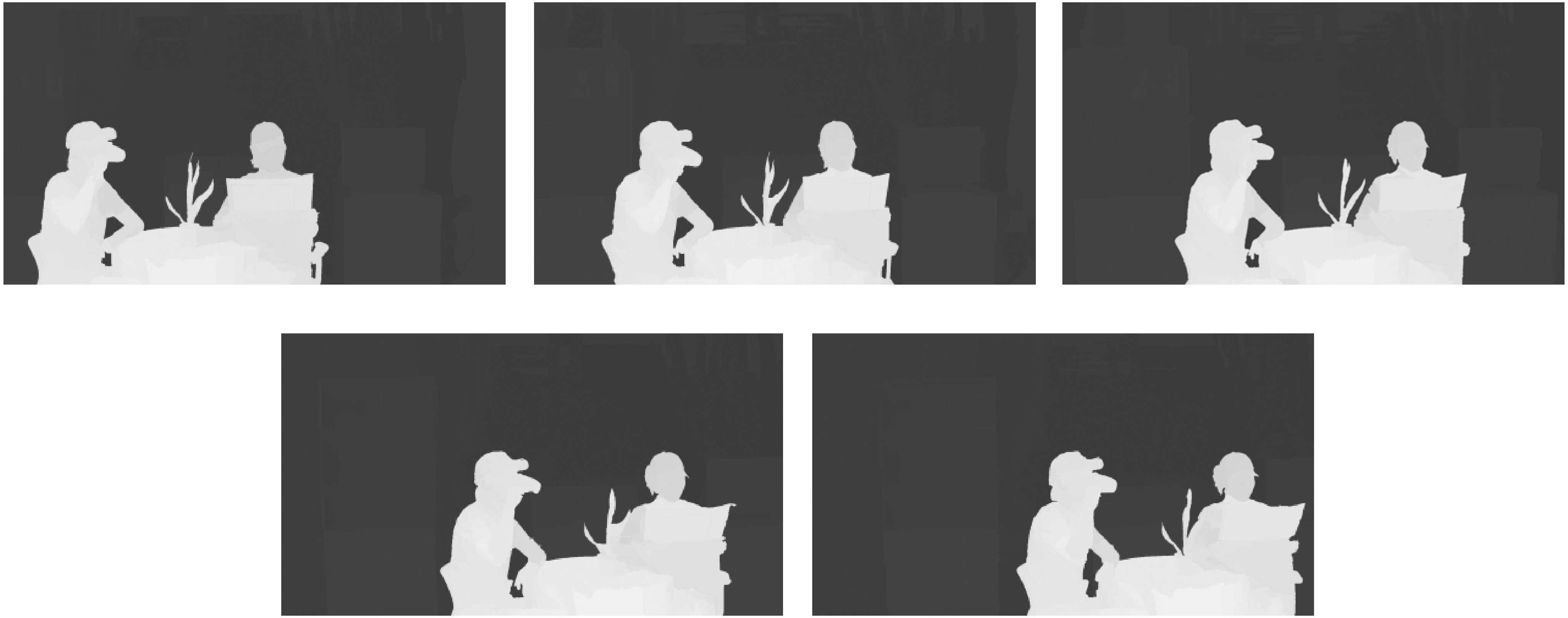
- Modified Belief Propagation (BP)
  - Modified BP
    - Search range restriction using initial depth information
  - Data cost energy function

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

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

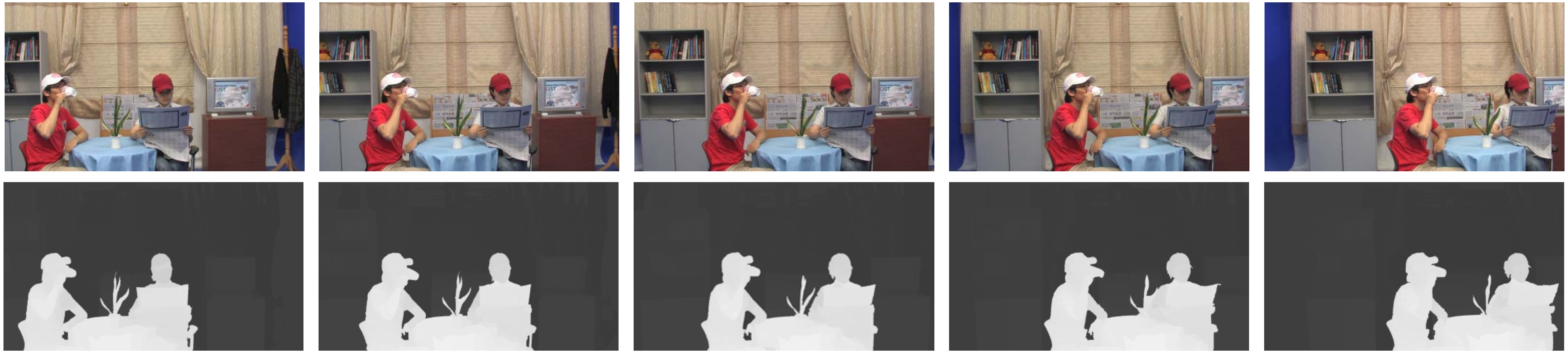


Final Multi-view Depth Map



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

Output Images from HCS

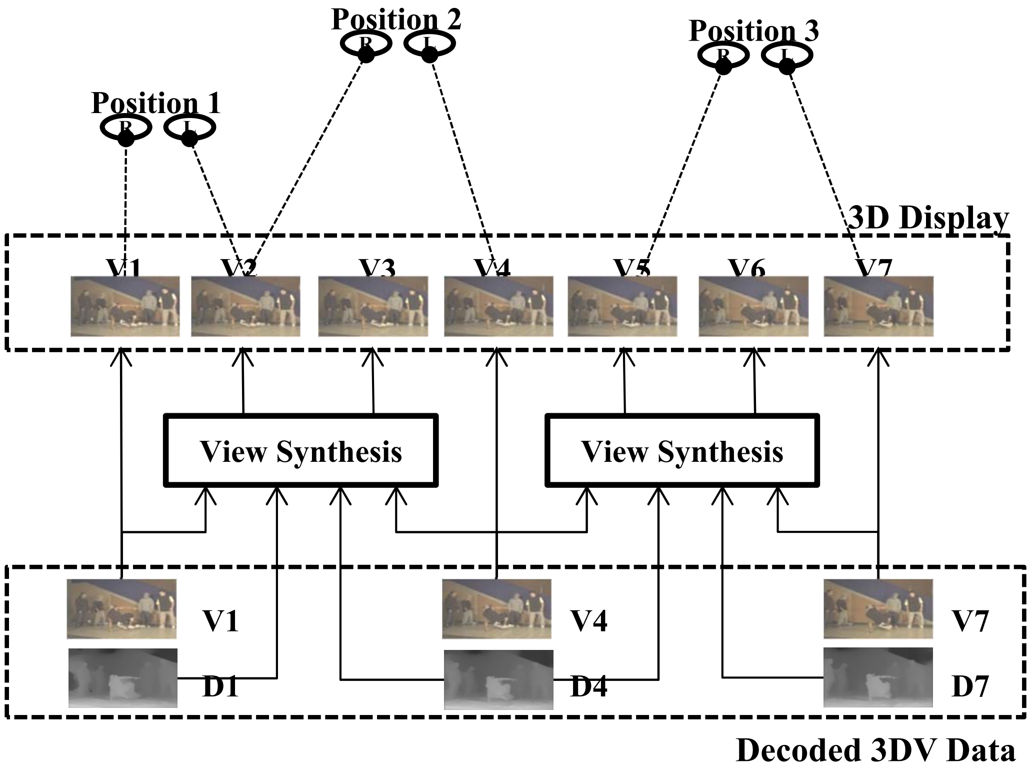


view 1 | view 2 | view 3 | view 4 | view 5

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

3D Rendering using View Synthesis

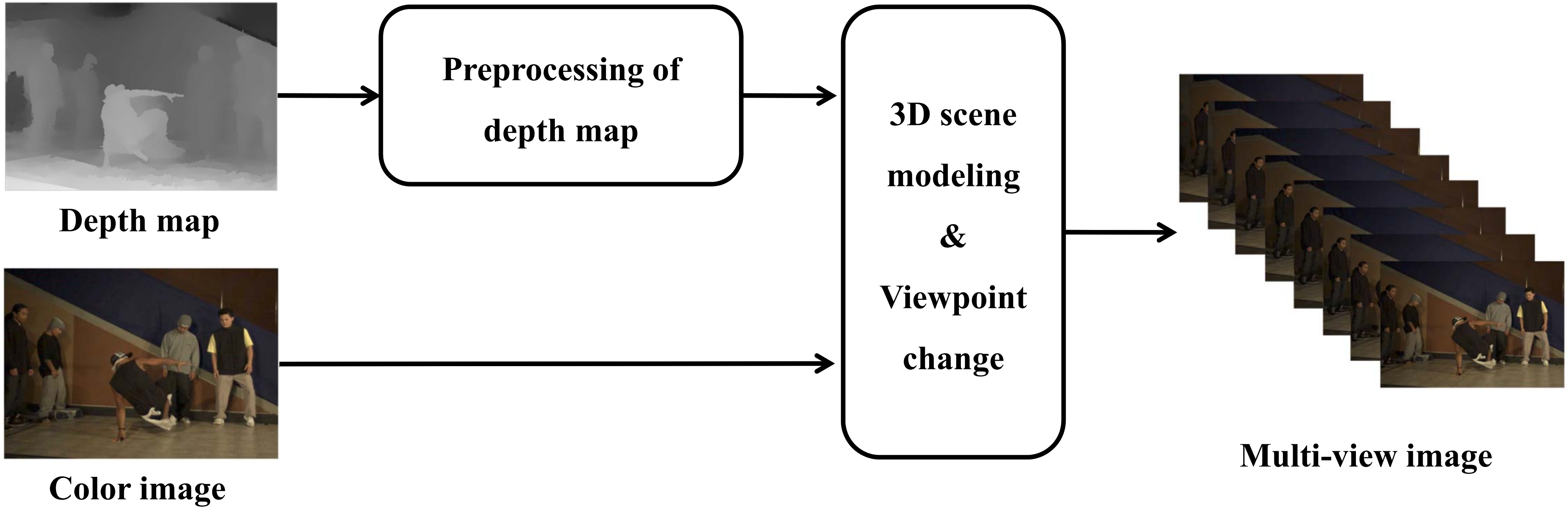
- Render Novel Views using View Synthesis
  - Generate virtual viewpoint images
  - Depth data is necessary for view synthesis



Depth Image-based Rendering

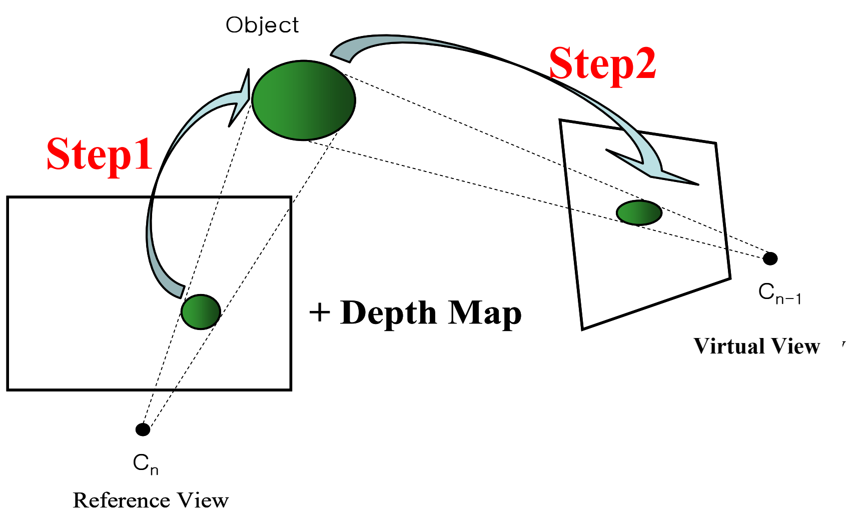
- Construct consecutive 3D video using texture images and depth maps
- Assumption: we have the depth information in advance

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 world coordinates into the virtual view



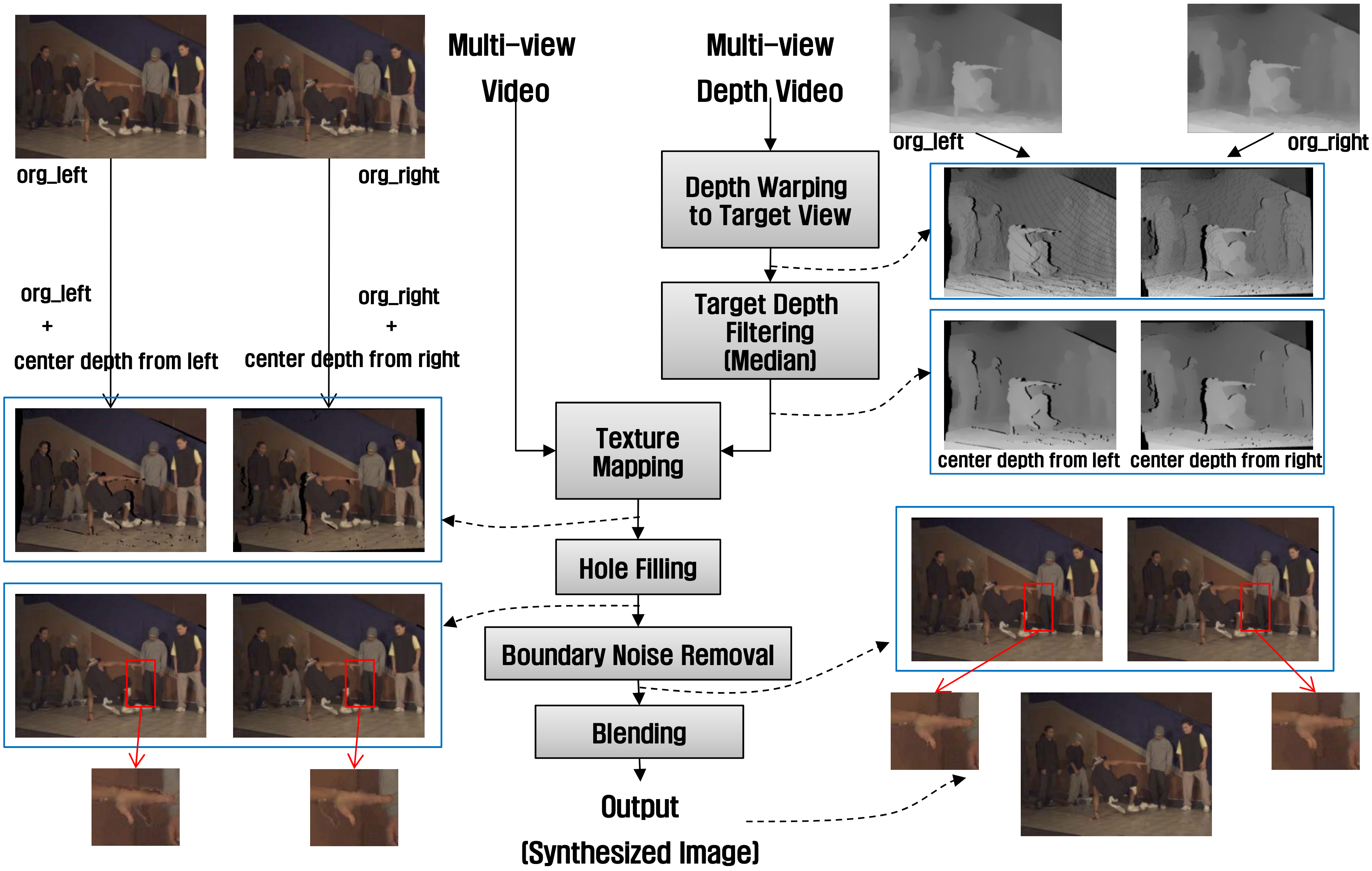
Step 1: Backward projection from image to world

$$\mathbf{M}_r = \mathbf{R}_r^{-1} \cdot \mathbf{A}_r^{-1} \cdot \mathbf{m}_r \cdot \mathbf{d}(\mathbf{m}_r) - \mathbf{R}_r^{-1} \cdot \mathbf{t}_r$$

Step 2: Re-projection from world to target world

$$\tilde{\mathbf{m}}_t = \mathbf{A}_t [\mathbf{R}_t \mid \mathbf{t}_t] \tilde{\mathbf{M}}_r$$

Procedure of View Synthesis



Warped Image

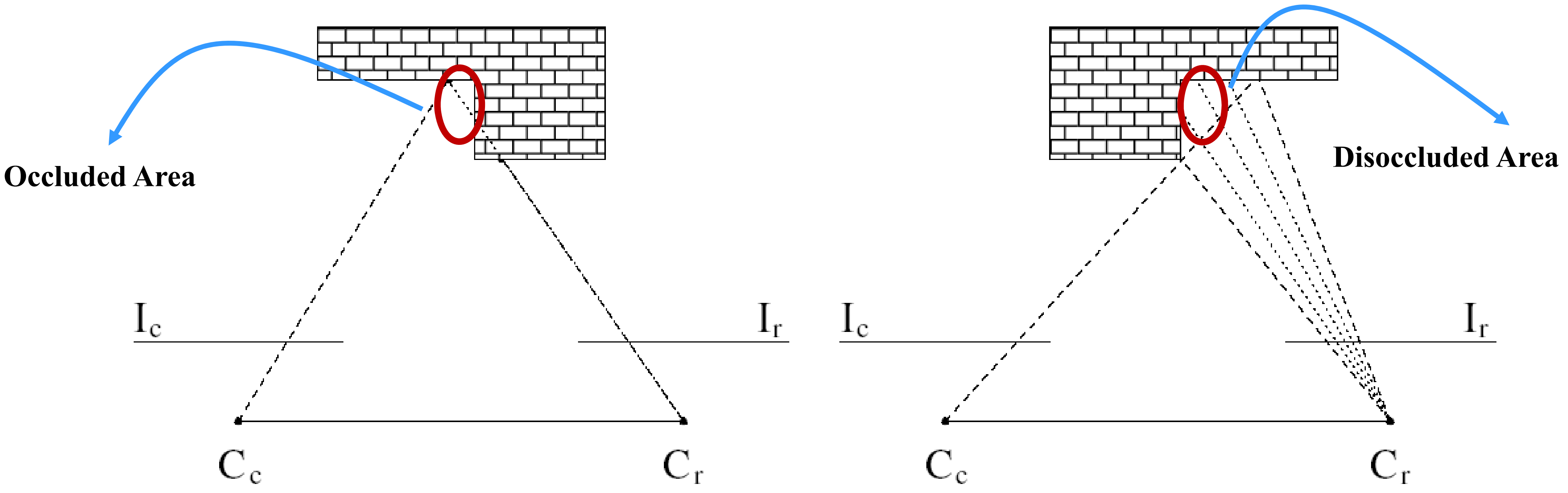
- Depth-based 3D Warping
  - Use warped depth image in virtual view
- Hole Area
  - Newly Exposed Area (white)

< 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

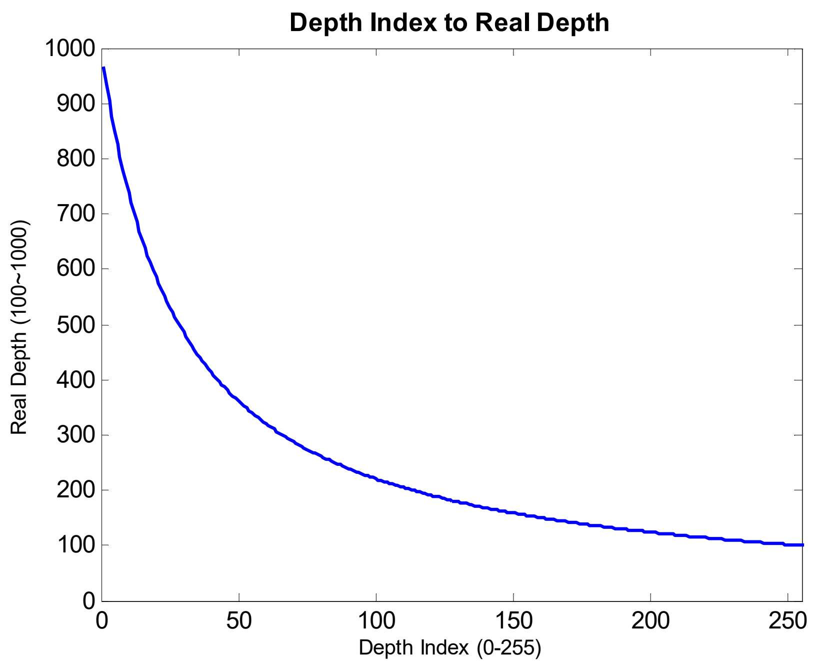
Real depths may not be perfectly represented by the depth map.

- 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

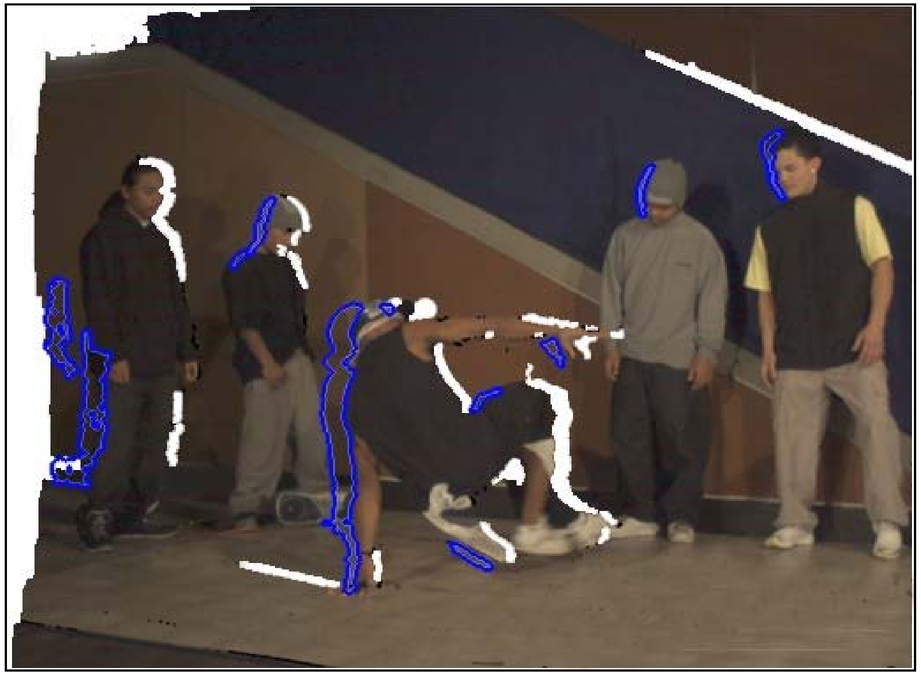
$$Z = \frac{1.0}{\frac{d}{255.0} \cdot \left(\frac{1.0}{Z_{near}} - \frac{1.0}{Z_{far}} \right) + \frac{1}{Z_{far}}}$$



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

$$A_V = \alpha \cdot A_L + (1-\alpha) \cdot A_R$$

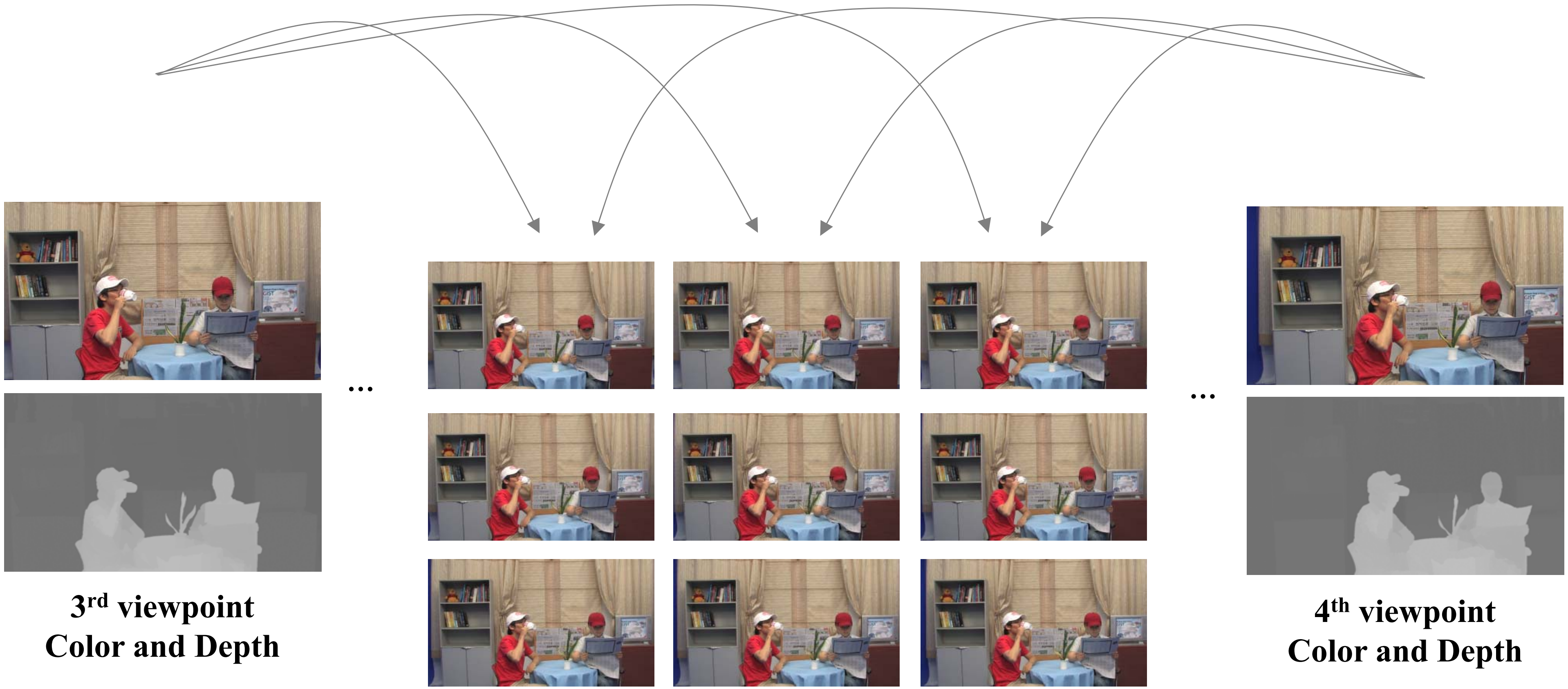
$$R_V = \alpha \cdot R_L + (1-\alpha) \cdot R_R$$

$$t_V = \alpha \cdot t_L + (1-\alpha) \cdot t_R$$

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

3-D Scene Rendering

Generated Intermediate Views



3rd viewpoint
Color and Depth

Intermediate images at virtual viewpoints

4th viewpoint
Color and Depth

Conclusion

- 3D Video for AR/VR Systems
- 3D Image Representation
- MPEG Activities for 3D Video Coding
  - 3D Audio Visual (3DAV)
  - Multi-view Video Coding (MVC)
  - Free-viewpoint TV (FTV)
  - 3D Video Coding (3DVC)
  - MPEG-I Visual
- Challenging Issues of FTV
  - Capture and calibration with flexible camera arrangement
  - Data format and view synthesis
  - Real-time generation of super multi-view images
  - Free navigation over a wide area

Acknowledgments

- MPEG/JCT-3V Contributions
  - http://mpeg.chiariglione.org/standards/exploration/free-viewpoint-television-ftv
- GIST RBRC Members
- GIST VCL Members
  - https://vclab.gist.ac.kr/publications_sci_journals

Thank you for your attention!

Any questions?

Prof. Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST)

Realistic Broadcasting Research Center (RBRC)

+82-62-715-2211

hoyo@gist.ac.kr

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

http://vclab.gist.ac.kr/

Contact Information

Prof. Yo-Sung Ho

- +82-62-715-2211
- hoyo@gist.ac.kr
- http://vclab.gist.ac.kr/

Realistic Broadcasting Research Center (RBRC)

- +82-62-715-2258
- http://rbrc.gist.ac.kr/

Gwangju Institute of Science and Technology (GIST)

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