3D perception, representation and compression

O. Le Meur
olemeur@irisa.fr

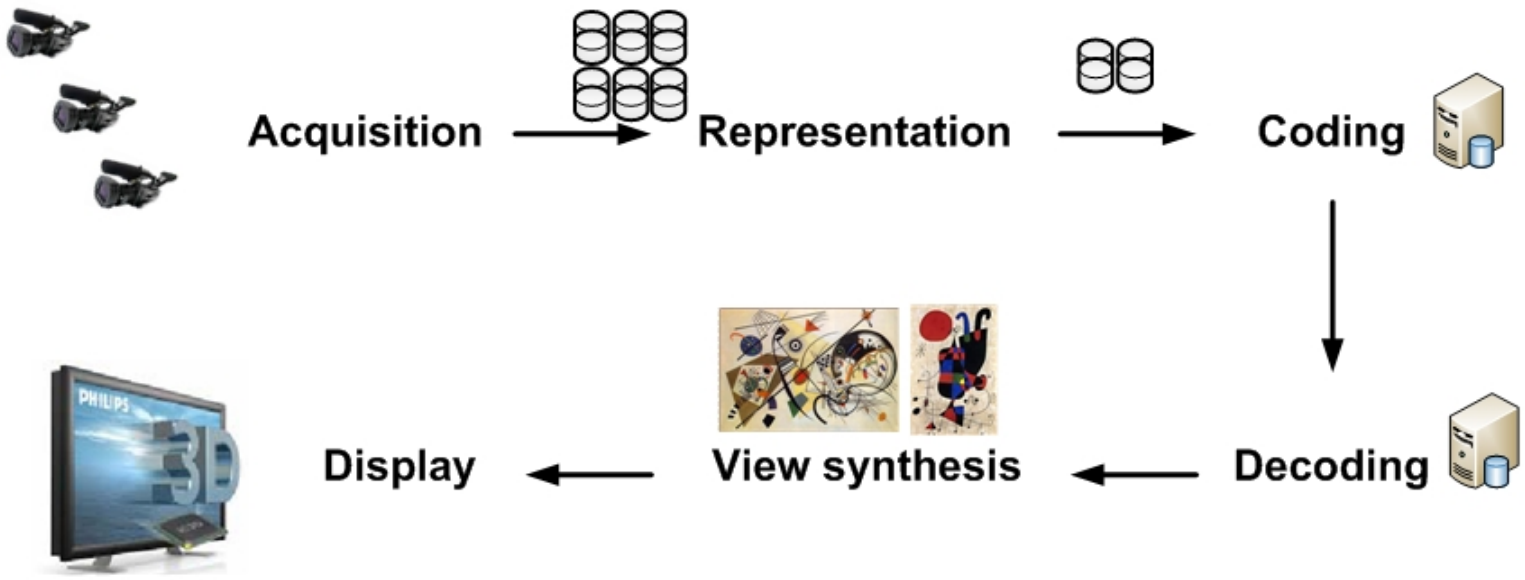
Univ. of Rennes 1
http://www.irisa.fr/temics/staff/lemeur/

November 21, 2011

3D perception, representation and compression

Introduction

Global architecture



Problems:

- Acquisition: synchronization, calibration...
- Compression: compact representation
- Rendering: photo-realistic virtual view generation

3D Perception

The depth...

How do we see depth?

This is a complex process based on different mechanisms:

- Oculomotor;
- Monocular inference of the depth (visual inferences based on certain rules);
- Binocular inference of the depth.

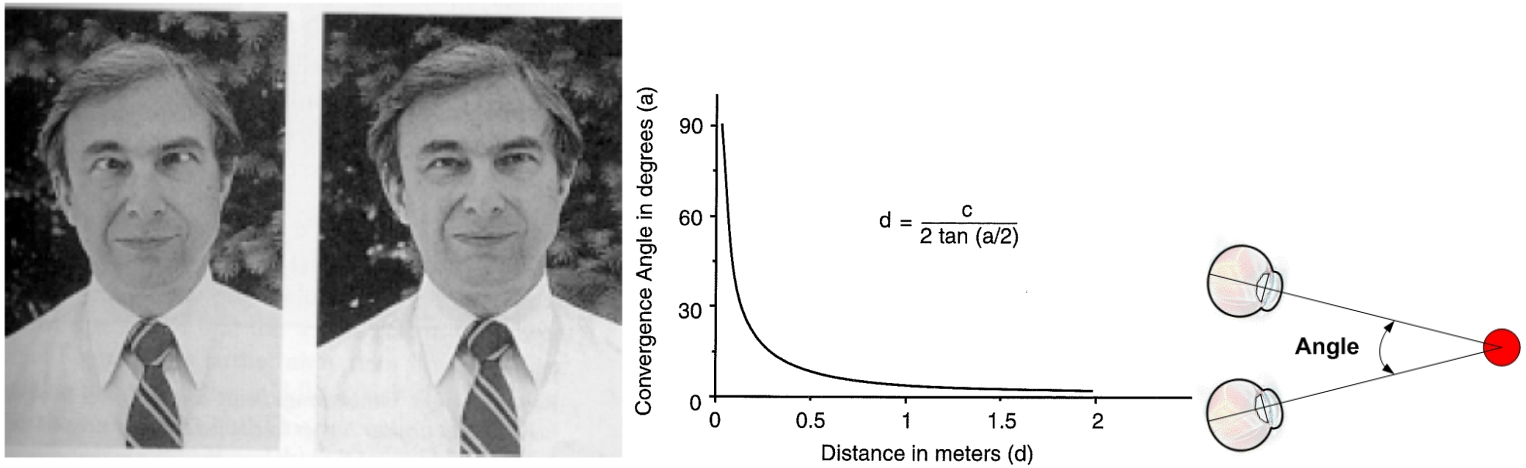
Oculomotor

Definition (Oculomotor cues)

These cues are based on the ability to sense the position of our eyes and the tension in the eye muscles.

Oculomotor cues are due to *convergence* and *accommodation*.

- Convergence: when looking at nearby objects our eyes move inwards.
- Accommodation: when looking at far away objects, our eyes move outward.



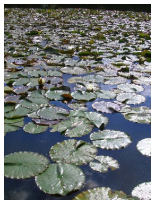
http://www.ics.uci.edu/~majumder/vispercep/chap8notes.pdf

Monocular inference of the depth

Definition (Monocular cues)

Monocular cues are the ones that are obtained from the 2D image of only one eye.

- Linear perspective: it describes the tendency of parallel lines to appear to converge at the horizon (vanishing point)
- Relative size (size constancy): relative and absolute size
- Texture gradient
- Interposition
- Monocular Movement Parallax: When our heads move from side to side, objects at different distances move at a different relative velocity.
- Shadow, lighting, aerial perspective...



Monocular inference of the depth

Ponzo Illusion, inference of the depth from our semantic understanding.....

Depth perception changes perception of line length.

Binocular inference of the depth

Definition (Binocular cues)

Binocular cues depend on the images from both eyes. Our eyes are placed about 6cm apart and hence they get a different view of the objects in the environment that appear in both eyes (this is called STEREOPSIS).

Extracted from `http://web.media.mit.edu/~mhirsch/byo3d/`

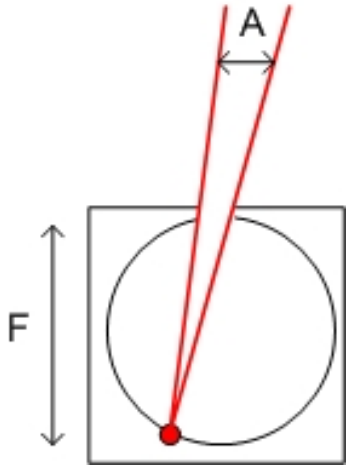
The overlap of the view from two eyes is significant but the viewpoint is different. This difference is called *binocular disparity* and is converted to depth information.

Binocular disparity

Preamble:

- F, focal length $\rightarrow$ diameter of eye $\approx 25mm$;
- A, apperture $\rightarrow$ diameter of pupil $\approx 5mm$.

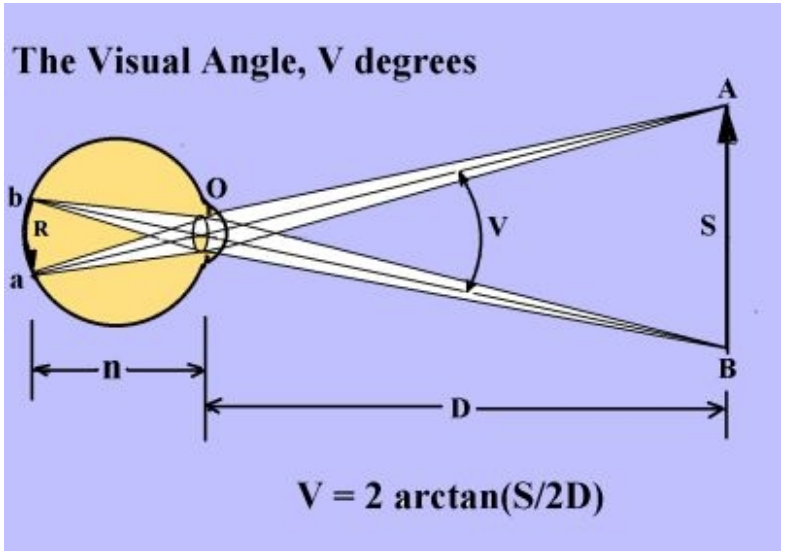
The following ratio is called the *F-number*: $\frac{F}{A}$



In photography, a large F-number means the pixel gather light from a small angle.

Binocular disparity

Preamble: visual angle



For visual angle smaller than about 10 degrees, a simpler formula can be used:

$$\tan V = \frac{S}{D}$$

However, we also have:

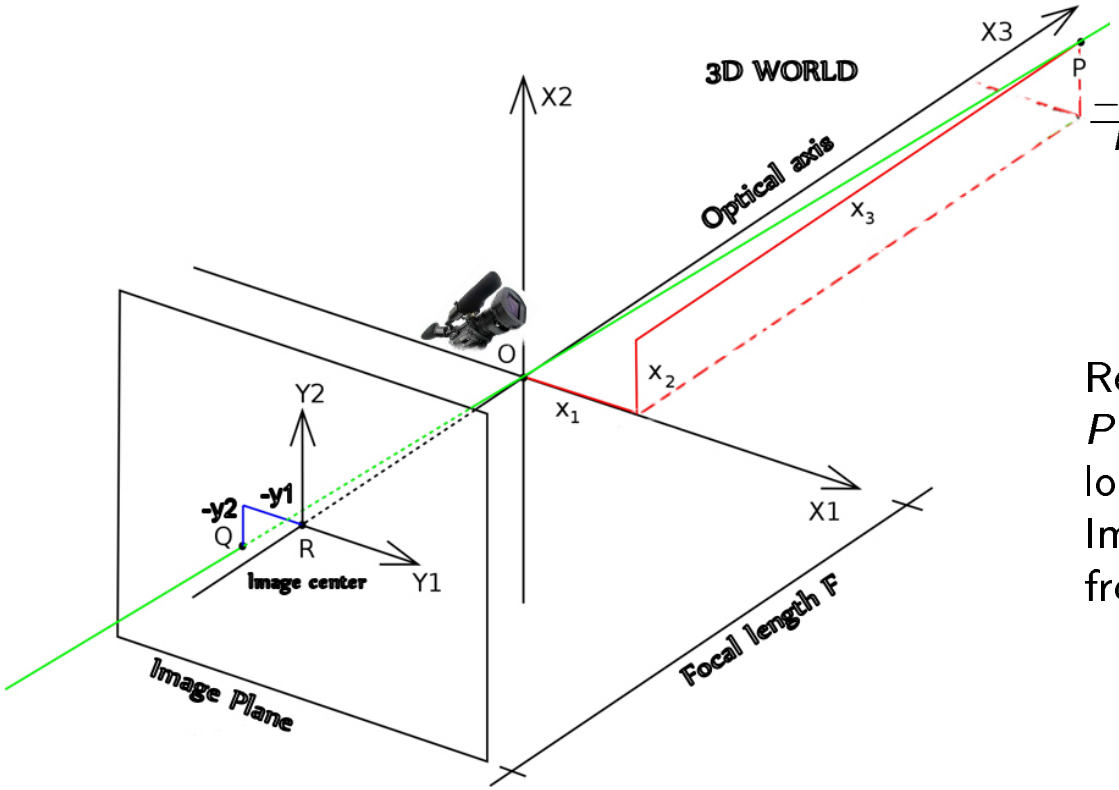
$$\tan V = \frac{R}{n}$$

If one looks at a one-centimeter object at a distance of one meter and a two-centimeter object at a distance of two meters, both subtend the same visual angle of about 0.01 rad or 0.57°. Thus they have the same retinal image size $R \approx 0.17$ mm.

Binocular disparity

Definition (Pinhole camera model)

The pinhole camera model describes the mathematical relationship between the coordinates of a 3D point and its projection onto the image plane of an ideal pinhole camera.



$$\frac{-y_1}{F} = \frac{x_1}{x_3} \text{ and } \frac{-y_2}{F} = \frac{x_2}{x_3}$$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = -\frac{F}{x_3} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Remark: all points of the line PO are projected into the same location in the image plane(!!). Impossible to infer the depth x_3 from a single image.

Binocular disparity

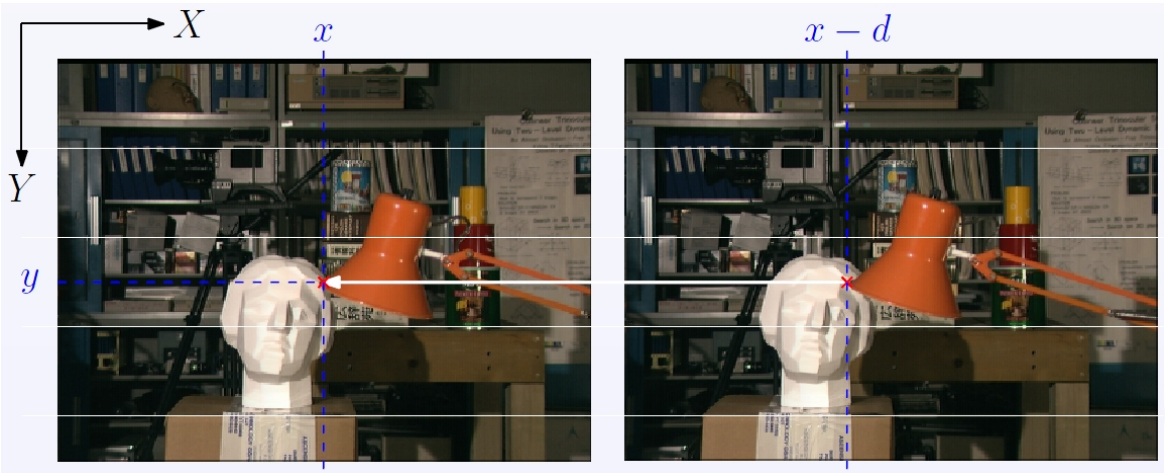
Two eyes (cameras) see the world from slightly different positions. Suppose eyes are separated by T_X (This distance is sometimes called the baseline of the camera pair).

With these assumptions, a 3D point with coordinates (X_0, Y_0, Z_0) in the left camera's coordinate system would have coordinates $(X_0 - T_X, Y_0, Z_0)$ in the right camera's coordinate system.

These 3D points would project to a different x value in the left and right images. The difference in x position is called the binocular disparity.

Definition (Binocular disparity)

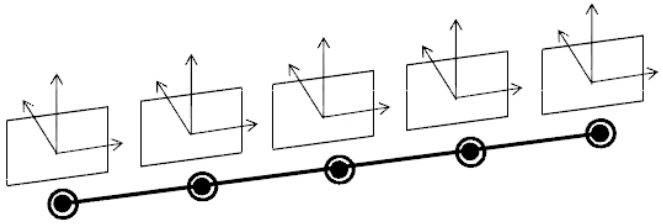
$$d = x_l - x_r = \frac{X_0}{Z_0} F - \frac{X_0 - T_X}{Z_0} = F \frac{T_X}{Z_0} \text{ (mm)}.$$



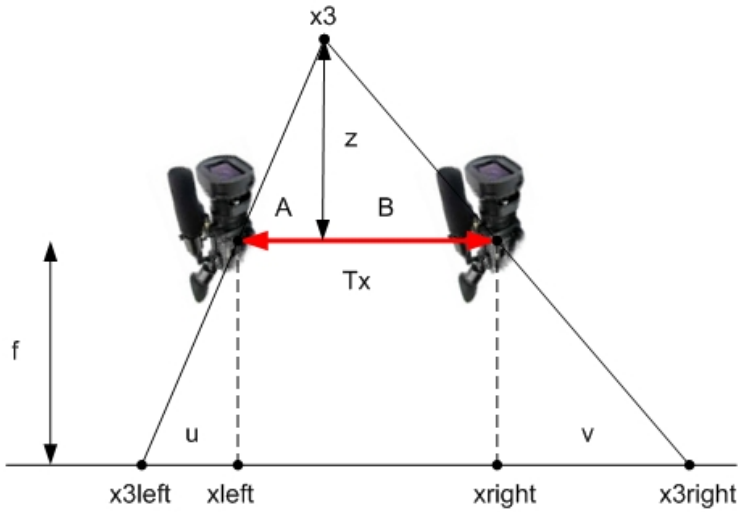
Binocular disparity

Disparity and depth in the parallel camera case:

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



Depth is inversely proportional to disparity in this case:



$$\frac{u}{f} = \frac{A}{z} \text{ and } \frac{v}{f} = \frac{B}{z}$$

$$z = \frac{f \times T_x}{d}$$

with $d = v - u$

Binocular disparity

From http://www.etcenter.org/2010/09/crc-provides-depth-estimation-from-stereoscopic-video/

Source of information about depth

ocular versus optical,
binocular versus monocular,
static versus dynamic,
relative versus absolute,
qualitative versus quantitative.

INFORMATION SOURCE	Ocular/ Optical	Binocular/ Monocular	Static/ Dynamic	Relative/ Absolute	Qualitative/ Quantitative
Accommodation	ocular	monocular	static	absolute	quantitative
Convergence	ocular	binocular	static	absolute	quantitative
Binocular Disparity	optical	binocular	static	relative	quantitative
Motion Parallax	optical	monocular	dynamic	relative	quantitative
Texture Accretion/Deletion	optical	monocular	dyanmic	relative	qualitative
Convergence of Parallels	optical	monocular	static	relative	quantitative
Position relative to Horizon	optical	monocular	static	relative	quantitative
Relative Size	optical	monocular	static	relative	quantitative
Familiar Size	optical	monocular	static	absolute	quantitative
Texture Gradients	optical	monocular	static	relative	quantitative
Edge Interpretation	optical	monocular	static	relative	qualitative
Shading and Shadows	optical	monocular	static	relative	qualitative
Aerial Perspective	optical	monocular	static	relative	qualitative

Extracted from Vision Science, Palmer.

3D Camera

3D Camera

Stereoscopic cameras

Fujifilm Finepix Real 3D W3 (10 Mpixels READ 3DW3; 3D video and pictures (HD); Price < 400 euros).

3D Camera

Multi-view cameras

8-view 1-D Parallel MVV capture setup



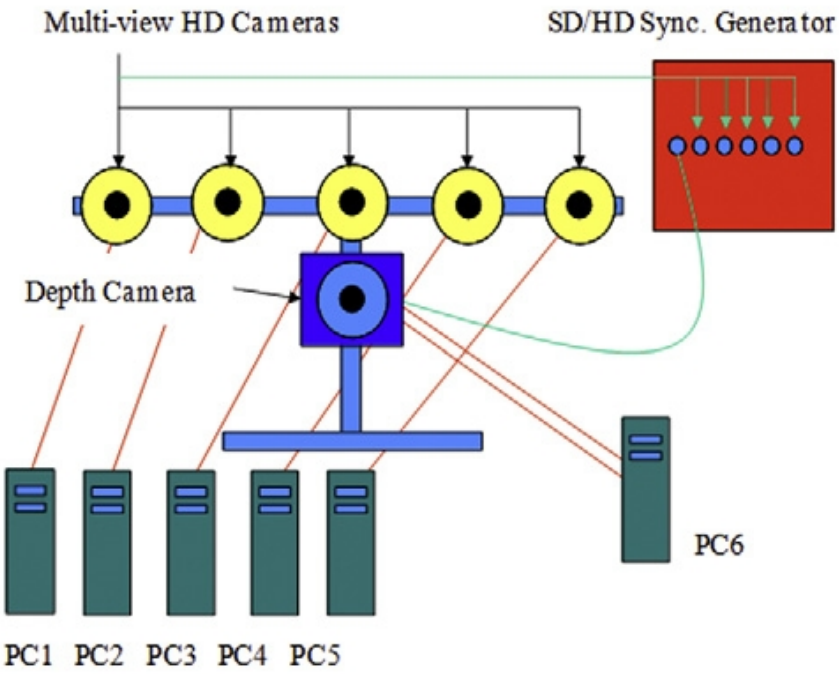
Multi-view cameras systems are classified by the shape and dimension

- Parallel vs convergent camera arrays
- 1-D vs 2-D camera arrays

3D Camera

Hybrid camera system

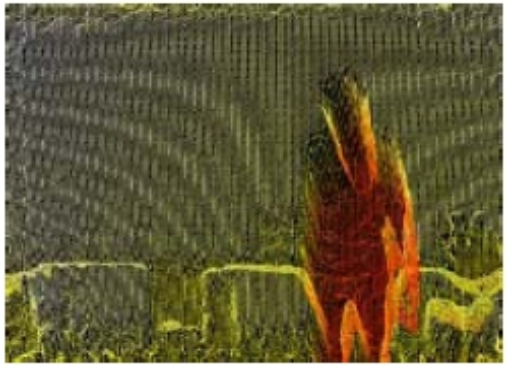
An hybrid camera system combines depth camera and multiple video cameras to generate multi-view video sequences and their corresponding depth maps.



Extracted from E.K. Lee, Y.S. Ho, Generation of high-quality depth maps using hybrid camera system for 3-D video, J. Vis. Commun. Image R. 22 (2011) 73-84.

3D Camera

Hybrid camera system (stereo+depth)



3D Display

3D Display

Taxonomy of 3D displays:

- Glasses-bound stereoscopic:
  - → Head-mounted
  - → Multiplexed (stereo pair with same display surface)
- Unencumbered automultiscopic
  - → Parallax-based (2D displays with light-directing elements)
  - → Volumetric
  - → Holographic

Multiplexed display

Stereo images can be displayed

- **sequentially** at a doubled frame rate (field-sequential schemes).
  - → to avoid image flickering and ghosting effects, the image source must have a high refresh rate and each image must decay completely before the next field is displayed.
- **concurrently** at a regular frame rate (field-concurrent schemes).

The major shuttering techniques include color filters, polarization filters, and LCD shutter glasses.

Multiplexed display

Field-sequential scheme

Definition (Anaglyph)

A stereoscopic picture where the left and right eye images are superimposed, but in different colours. A colour filter over each eye only transmits the image component suitable for that eye and the brain interprets the result in three dimensions.

For a red-blue anaglyph:

- the eye covered by the red filter sees the red parts of the image as white, and the blue parts as black;
- the eye covered by the blue filter perceives the opposite effect. True white or true black areas are perceived the same by each eye.

Multiplexed display

Main disadvantages of anaglyphs:

- Ghosting: If some colour from the left image gets into the right eye (and vice versa) a faintly coloured ghost will be seen. Increasing the parallax (stereo depth) separates the two images and makes ghosting worse.
- Retinal rivalry. If the brightness (luminance) of the two images is not the same in each eye, the effect is unpleasant.
- Wrong colours. There is no hope for getting colours exactly right, because each eye is getting only part of the RGB colour range.

Advantage:

- Anybody with normal vision can see 3D in an anaglyph;
- Red - Cyan stereoscopic glasses are common and cheap, often coming free with magazines showing 3D;
- A singe digital projector can show anaglyphs on a screen for a large audience, who see three dimensions through the same, cheap, coloured glasses used for computer or print viewing.

Anaglyphs are classified by the colour filters used.

scheme	left eye	L	R	right eye	color rendering
red-green	pure red			pure green	monochrome
red-blue	pure red			pure blue	monochrome
red-cyan	pure red			pure cyan (green+blue)	color (poor reds, good greens)

From Wikipedia

Multiplexed display

Field-sequential scheme

Definition (Polarized 3D glasses)

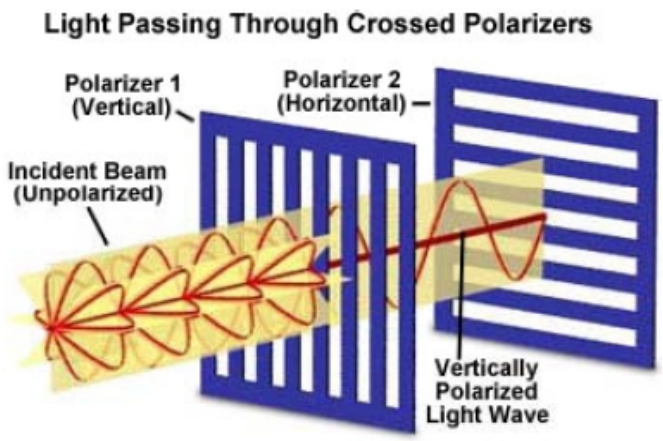
Polarized 3D glasses create the illusion of three-dimensional images by restricting the light that reaches each eye, an example of stereoscopy which exploits the polarization of light.

- Two images are projected superimposed onto the same screen through different polarizing filters.
- The viewer wears linearly polarized eyeglasses which also contain a pair of orthogonal polarizing filters oriented the same as the projector. As each filter only passes light which is similarly polarized and blocks the orthogonally polarized light, each eye only sees one of the projected images, and the 3D effect is achieved.

Multiplexed display

Field-sequential scheme

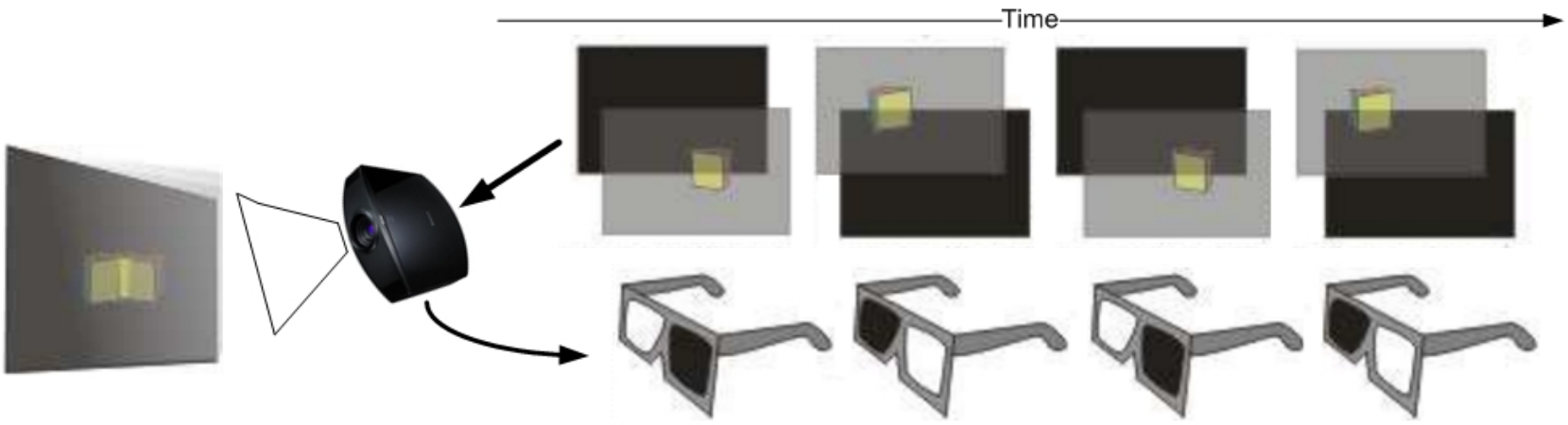
Linearly polarized glasses (orthogonal polarizing filters)



Linearly polarized glasses require the viewer to keep his head level, as tilting of the viewing filters will cause the images of the left and right channels to bleed over to the opposite channel.

Multiplexed display

Field-concurrent scheme

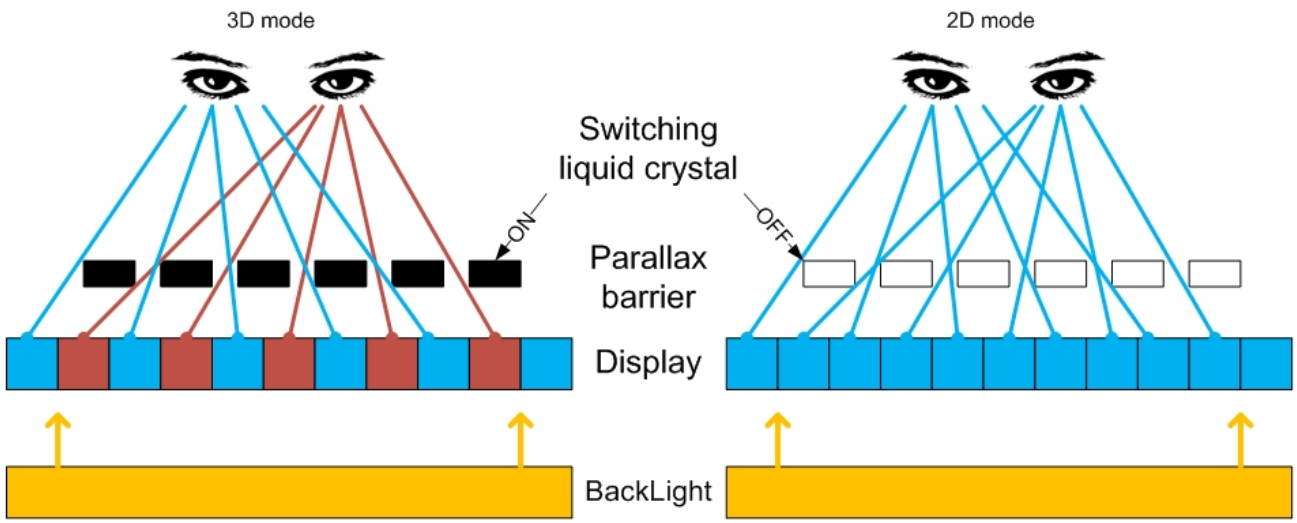


- Projector 1: displays images for left eye alternating with a black image;
- Projector 2: displays images for right eye alternating with a black image;
- Glasses are opaque for one eye at a time, controlled by an infrared shutter signal synchronized with the projectors;
- The images switch so fast that without glasses the images seem to be blended.

Unencumbered automultiscopic

Definition (Autostereoscopic displays)

Autostereoscopic displays have image-separation techniques integrated into the display units. Parallax displays present stereo pairs simultaneously and deliver multiple views directly to the correct eyes.

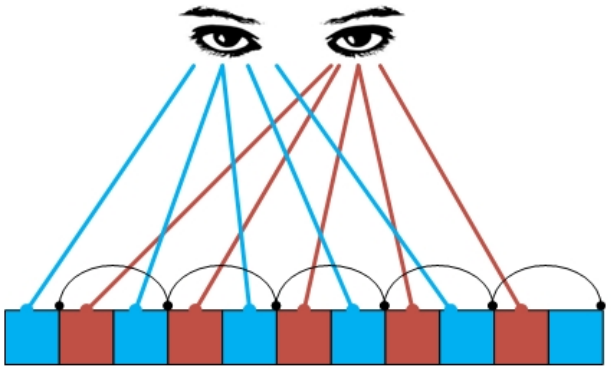


A parallax barrier is created separating the light into two images. This causes different images to be seen by the right and left eyes.

The parallax barrier is transparent. This causes the same image to be seen by the right and left eyes.

Others

- Integral imaging: it consists of a large number of closely packed, distinct micro-images, that are viewed by an observer through an array of spherical convex lenses, one lens for every micro-image.



The cylindrical lenses direct the diffuse light from a pixel so that it can only be seen in a limited angle.

- Multi-view Parallax Barrier / lenticular sheet.

Data representation

Two targets

Two scenarios exist. Notice that they do not exclude each other:

- 3D Video (3DV):
  - → Generation of a 3D depth impression from separate views for each eye
  - → Applications: DVD, Broadcast, medecine, military...
- Free Viewpoint Video:
  - → Generation of intermediate views in between real existing camera views
  - → Applications: production, special effects...

There is a need for **new tools** to deal with the new issues and **new standards** (MPEG) to ensure the developpment of consumer markets.

Build a representation suitable for 3D video: Compact, easy to compress, adapted for view synthesis

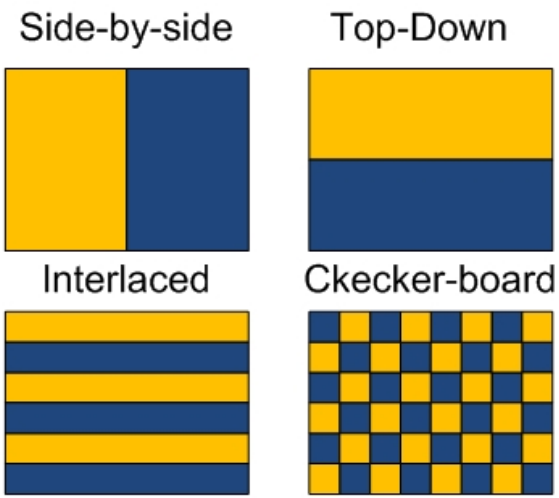
Conventional stereo video

Definition (Conventional stereo video)

A pair of 2D video is acquired, one for the left eye, and the other for the right eye.

Different strategies can be used to represent the two views:

- Full resolution
- Side-by-side
- Top-Down
- Interlaced
- Checker-board



Conventional stereo video

Left view

Right view

Interlaced

Top-Down

Side-by-side

Notice that the original views are downsampled in the horizontal dimension or vertical dimension to fit within the size of the original picture (loss of spatial resolution).

Adapted from I. Daribo's PhD thesis.

Conventional stereo video

Another strategy to represent these data is based on the Binocular suppression theory.

Definition (Binocular rivalry)

Binocular rivalry is a phenomenon of visual perception in which perception alternates between different images presented to each eye. When one image is presented to one eye and a very different image is presented to the other, instead of the two images being seen superimposed, one image is seen for a few moments, then the other, then the first, and so on, randomly for as long as one cares to look.

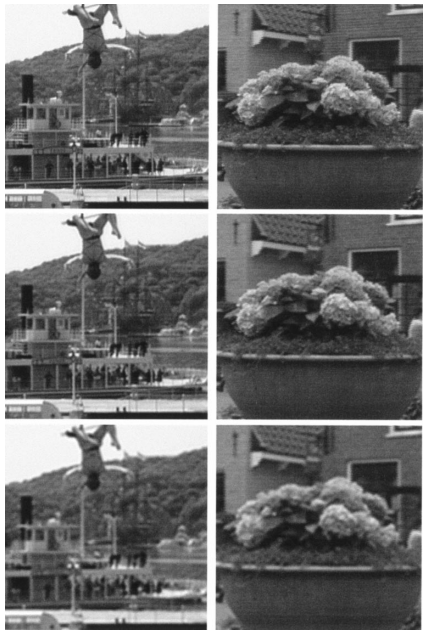
Suppression theory: The main idea is that, despite having two eyes, we see only one of everything (known as singleness of vision) because we see with one eye at a time.

Conventional stereo video

From L. Stelmach, Wa James Tam, D. Meegan, and A. Vincent. Stereo image quality: effects of mixed spatio-temporal resolution. IEEE Transactions on Circuits and Systems for Video Technology, 10(2):188-193, 2000.

They explored the response of the human visual system to mixed-resolution stereo video-sequences, in which one eye view was spatially or temporally low-pass filtered.

- the overall sensation of depth was unaffected by low-pass filtering, while ratings of quality and of sharpness were strongly weighted towards the eye with the greater spatial resolution;
- temporal filtering produced unacceptable results.



They conclude that low-pass spatial filtering of one channel of a stereo pair may be an effective technique to reduce storage and transmission bandwidth of stereoscopic sequences, while retaining all of the depth and most of the quality and sharpness of the unfiltered original image.

Conventional stereo video

Advantages:

- Simplicity
- Backward compatibility

Drawbacks:

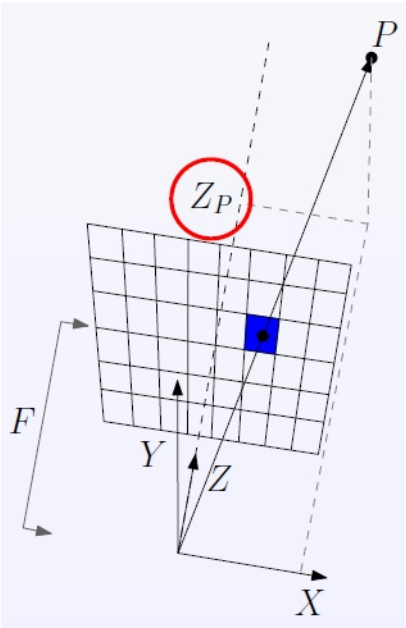
- it can be optimized for only one receiver configuration (size, number of views of the display...)
- 3D impression can not be modified at the receiver side...
- Loss of resolution

Depth map

Definition (Depth map)

Depth maps are 3D information added to an image.

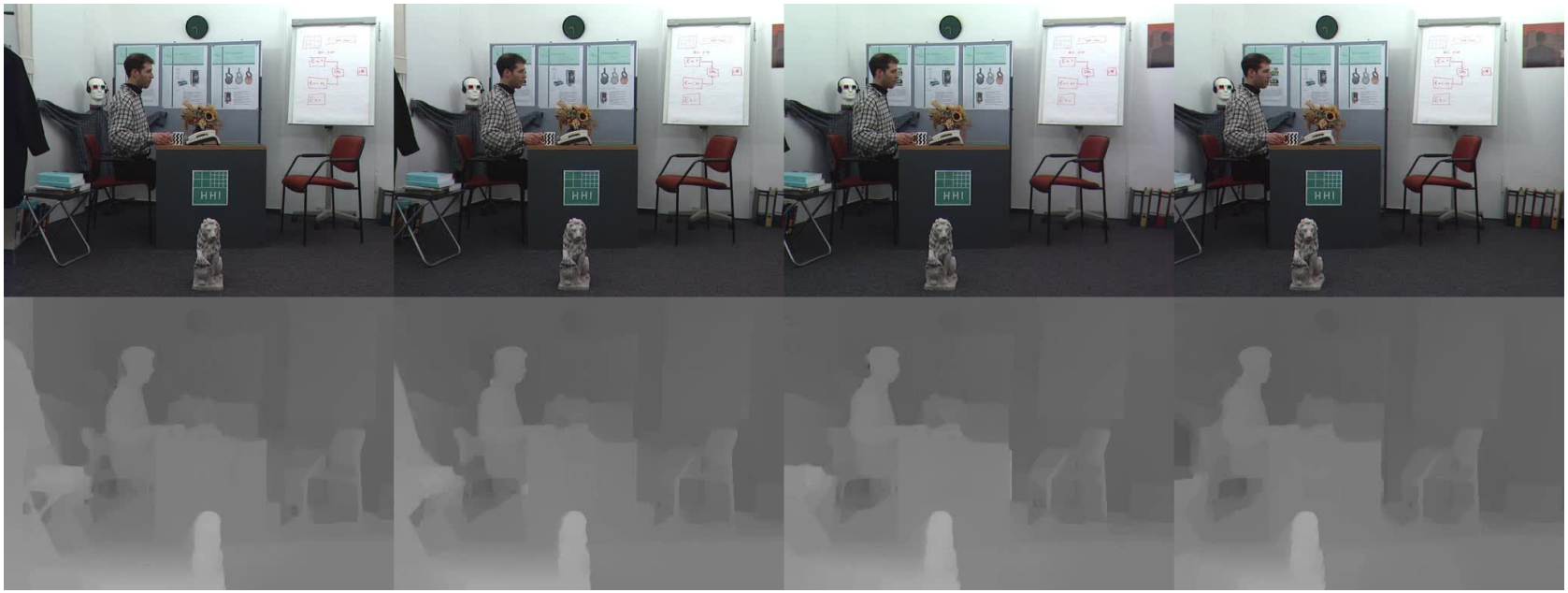
- this map is generally encoded as grayscale images (generally quantized with 8 bits, the closest point has a value 255 and the most distant point has a value 0).
- each pixel in a depth map describes the depth of the corresponding point in the image, relative to the camera
- restricted to a range in between two extremes Z_{NEAR} and Z_{FAR}.



In the classical $(X, Y, Z) \in \mathcal{R}^3$ camera space, the stored value corresponds to the Z coordinate of the point.
Given an image pixel $p = (x, y)$ and a corresponding depth value Z_p, the complete 3D coordinates are retrieved using the camear focal length F:

$$\begin{aligned} X_p &= xZ_p/F \\ Y_p &= yZ_p/F \end{aligned}$$

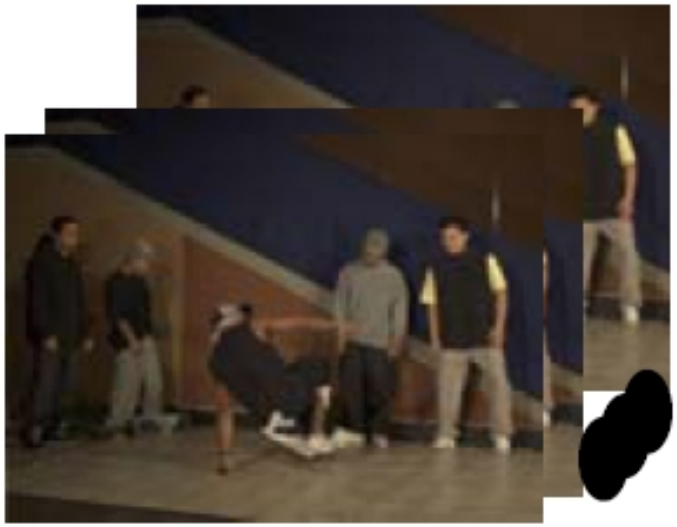
Depth map



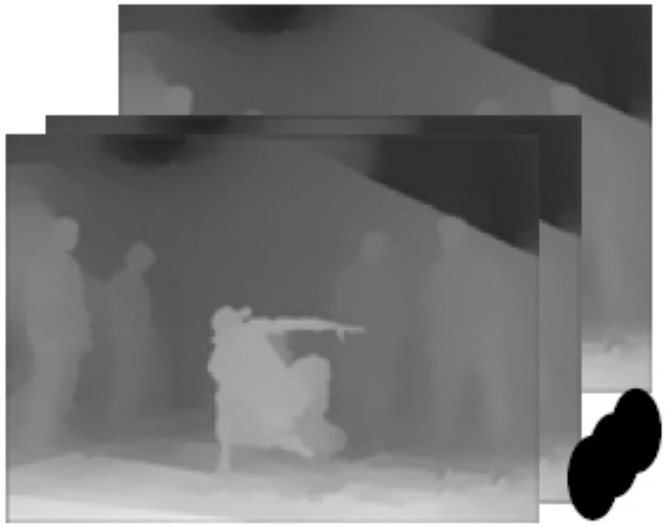
2D video + depth

Definition (2D video + depth)

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



Color image sequence



Depth map sequence

2D video + depth

After decoding, the second color video corresponding to the second view is reconstructed from the transmitted video-plus-depth data by means of DIBR techniques.

DIBR = Depth Image-based Rendering

Challenging Issues:

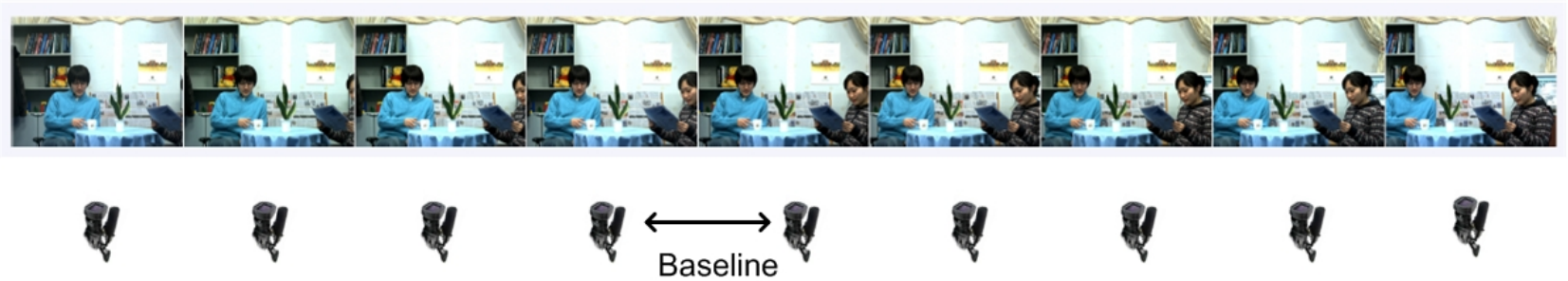
- Depth estimation (estimation from a captured stereo video content VOIR ref 101).
- Depth map coding scheme
- Bit allocation for depth coding
- View synthesis (errors in particular for disoccluded points).

Multiview video

Definition (MVV, MultiView Video)

This is a simple extension of the conventional stereo video. More than 2 views are captured.

- The amount of data to be processed and transmitted increases significantly compared to the conventional stereo data or classical 2D video.
- Multiview autostereoscopic displays



Multiview video + depth

The goal is to render a continuum of output views or a very large number of different output views at the decoder side without increasing the number of input views.

Definition (MVD, MultiView Video+depth)

The MVD format consists of a set of 2D videos with their associated depth video.

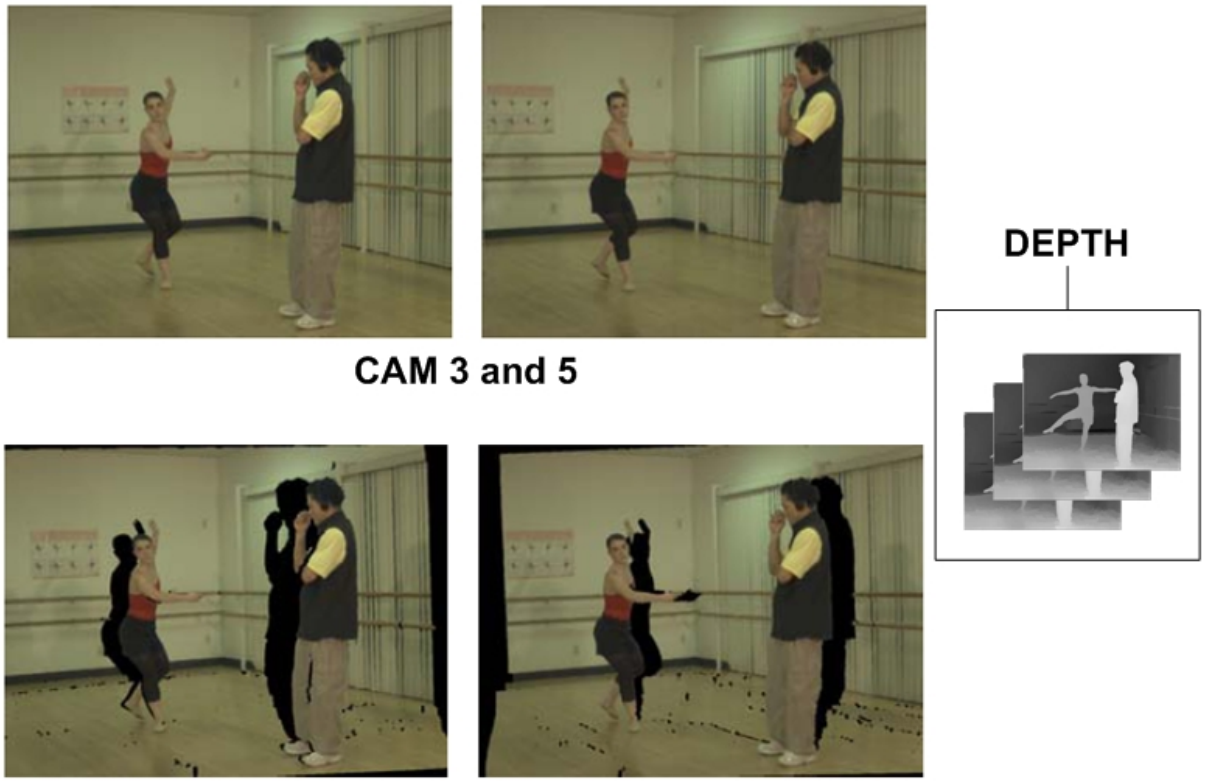
N 2D videos and N depth videos have to be transmitted

Interpolation of virtual intermediate views along all cameras

Multi-view video plus depth (MVD) format for advanced 3D video systems. Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-W100.

Multiview video + depth

Interpolation of virtual intermediate views along all cameras



Generation of virtual views for cam 4
(3 to 4, and 5 to 4)

Layered depth video

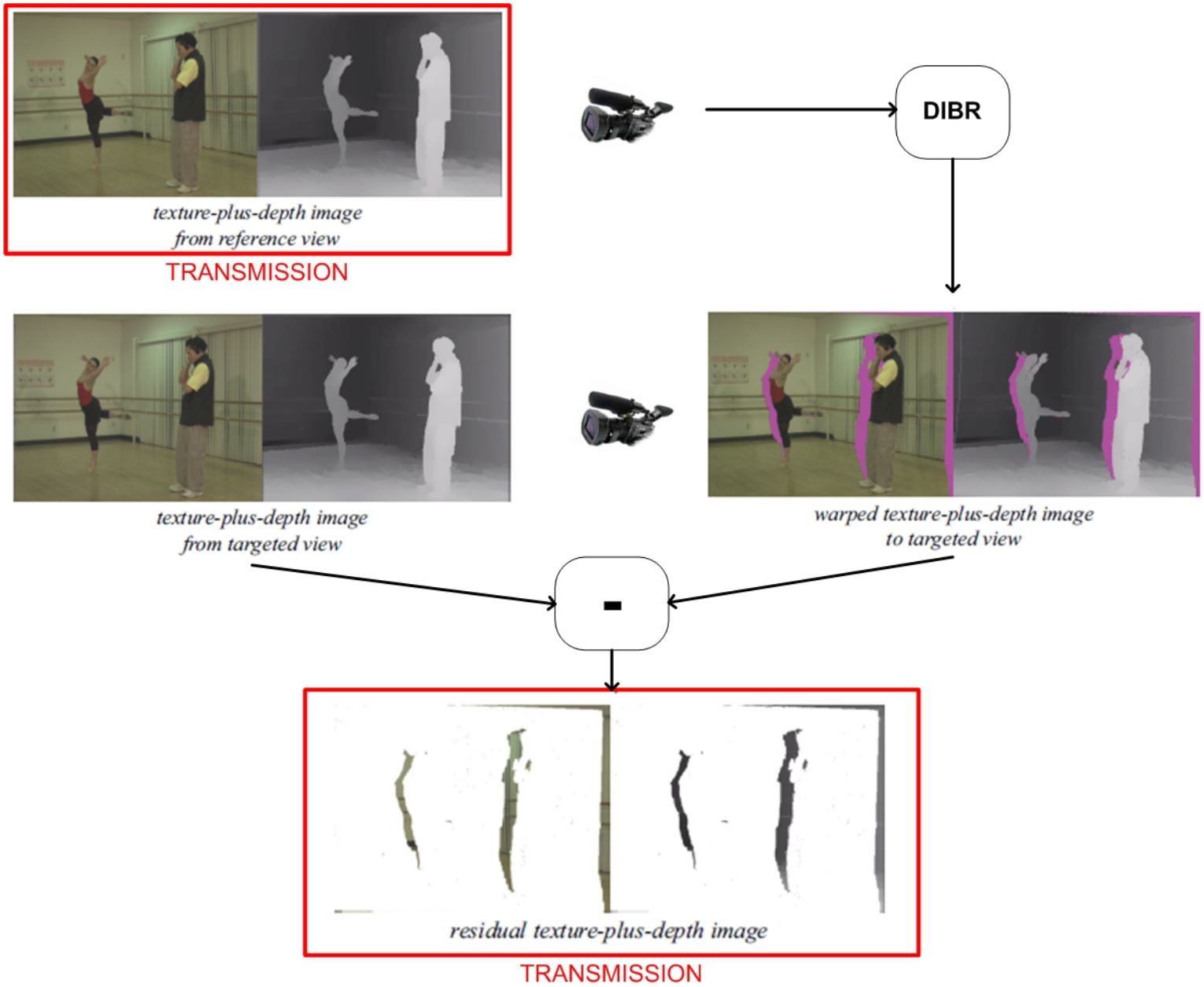
Definition (LDI Layer Depth Image)

A Layered Depth Image is a compact representation of a collection of view from different viewpoints. The first layer consits of the reference view and the subsequent layers contain visible and occluded texture.

1st layer 2nd layer 3rd layer 4th layer ...

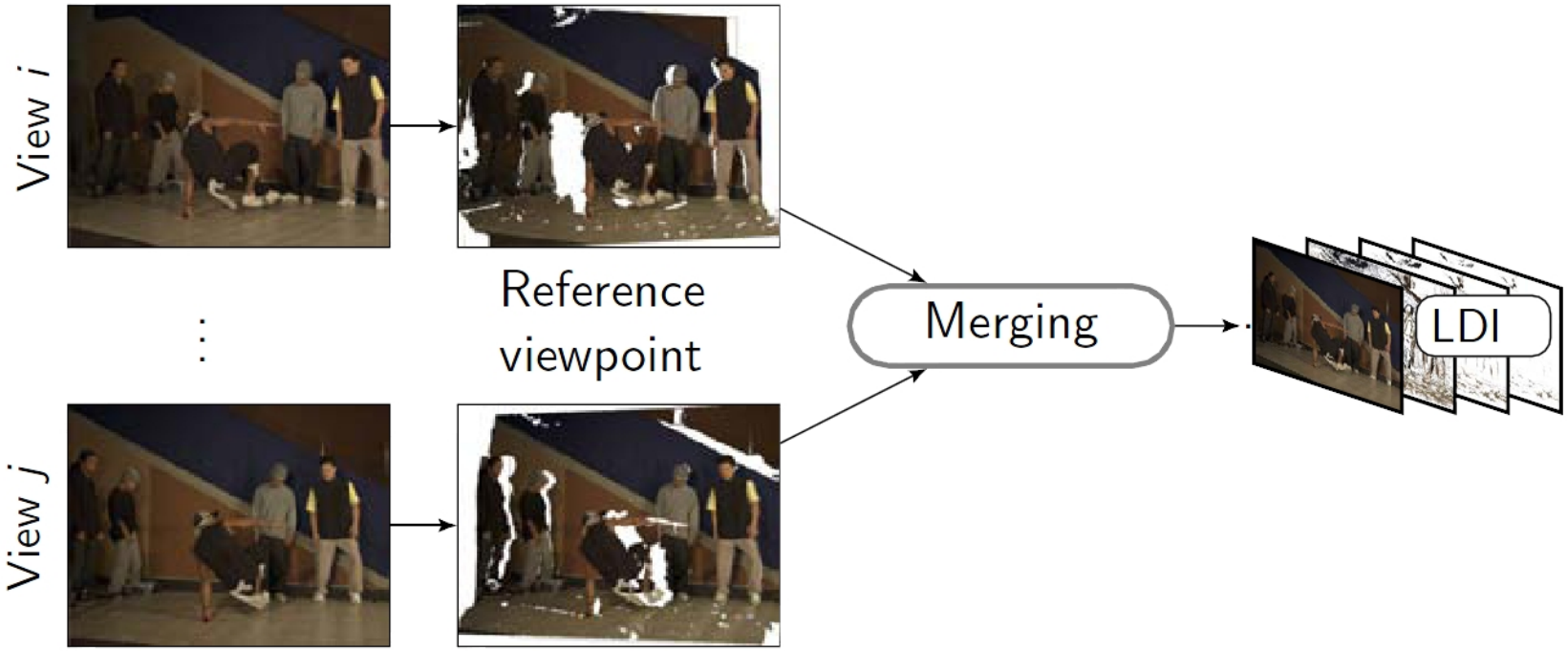
Layered depth video

LDI construction for two views:



Adapted from Daribo's thesis.

LDI construction for N views:
Every input views are warped onto a reference viewpoint, and then merged together.



where the merging consists in eliminating pixels described twice.

Adapted from Jantet's thesis.

Layered depth video

Advantages of LDI:

- Disocclusion can be filled by truth textures
- Camera freedom, virtual camera can move inside a large area
- Compactness

Drawbacks:

- Many layers, partially empty with scattered pixels distribution.
- Ghosting, Blurring...

Layered depth video

Definition (LDV Layer Depth Video)

LDV is the temporally extension of the layered depth image. LDV can be generated from MVD by warping the main layer image onto other contributing input images.

LDV format is composed of:

- A reference view (texture)
- Its associated depth map
- A background texture (residual)
- The background depth map

3D View Synthesis

Goal

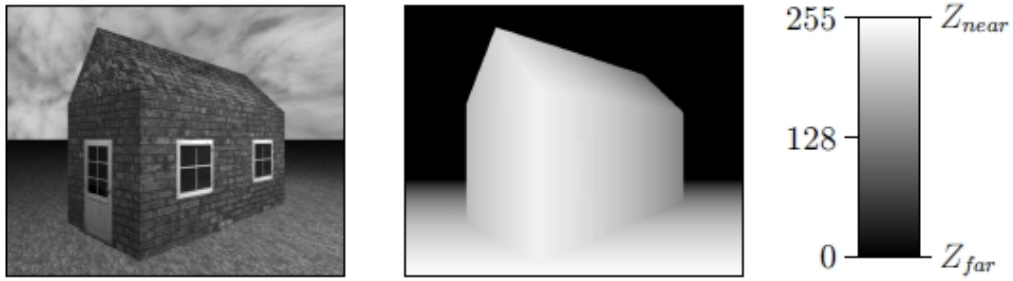
- Render Novel Views using View Synthesis:
- Generate virtual viewpoint images

Depth data is necessary for view synthesis

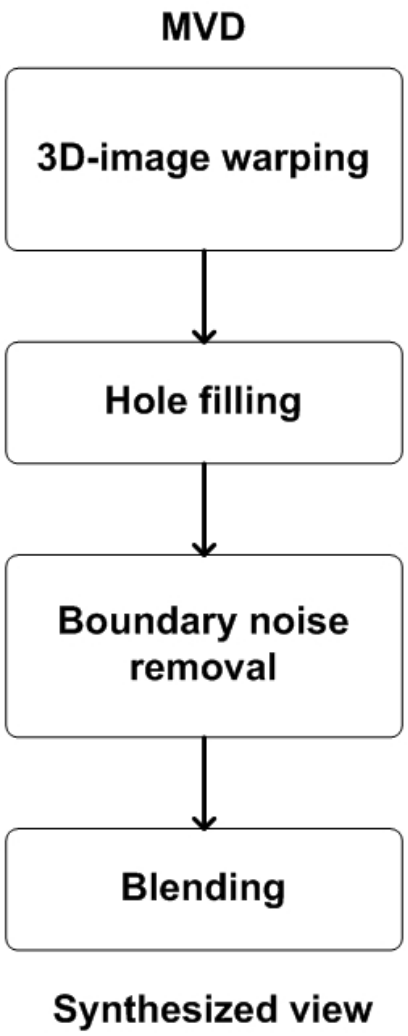
Conceptually, DIBR can be understood as a two-step process:

1. 3D image warping
2. reconstruction and re-sampling (view blending) (determination of pixel sample values in the synthesized image).

ATTEST data representation (Advanced Three-Dimensional Television System Technologies).



Synoptic

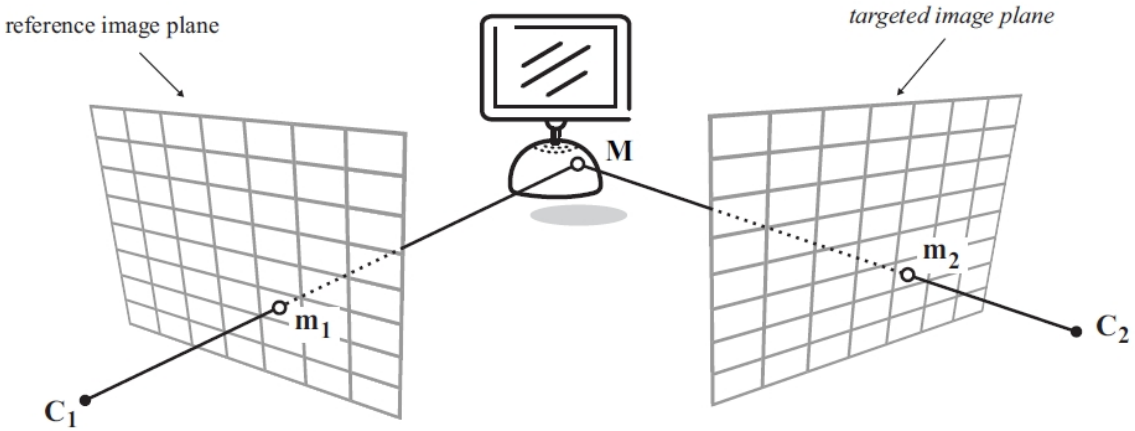


- Pre-processing of depth map
- 2D-to-3D followed by 3D-to-2D
- Inpainting
- Ghosts and cracks

3D-image warping

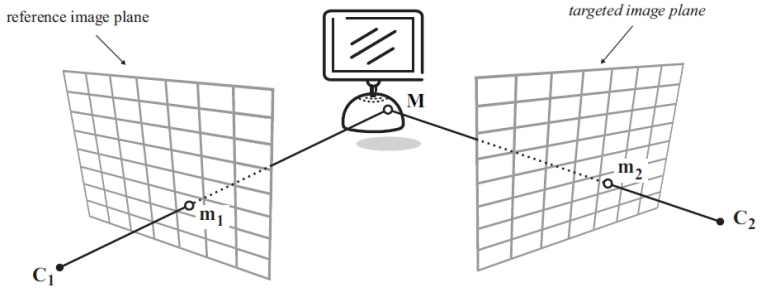
Definition (3D-image warping)

3D warping = projection (2D-to-3D) and subsequent projection (3D-to-2D).



- Intensity of the reference view image $I_1(u_1, v_1)$ at pixel coordiantes (u_1, v_1)
- Synthesize the second view $I_2(u_2, v_2)$ with the depth data $Z(u_1, v_1)$

3D-image warping



2D to 3D:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = R_1^{-1} K_1^{-1} \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} \lambda_1 - R_1^{-1} t_1$$

K_1, R_1 and t_1 denote the 3×3 intrinsic matrix, the 3×3 orthogonal rotation matrix and the the 3×1 translation vector of the reference view I_1.

3D-image warping

3D to 2D (3D-world back-projection):

$$\begin{pmatrix} u_2' \\ v_2' \\ w_2' \end{pmatrix} = K_2 R_2 \begin{pmatrix} x \\ y \\ z \end{pmatrix} + K_2 t_2$$

K_2, R_2 and t_2 are the targeted camera quantities.

For rectified cameras:

- if the camera are identical, same intrinsic parameters $K = K_1 = K_2$;
- Theire is no rotation $R_2 = I_3$, there is only a translation

$$\begin{pmatrix} u_2' \\ v_2' \\ w_2' \end{pmatrix} = \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} z + K \begin{pmatrix} t_x \\ 1 \\ 0 \end{pmatrix}$$

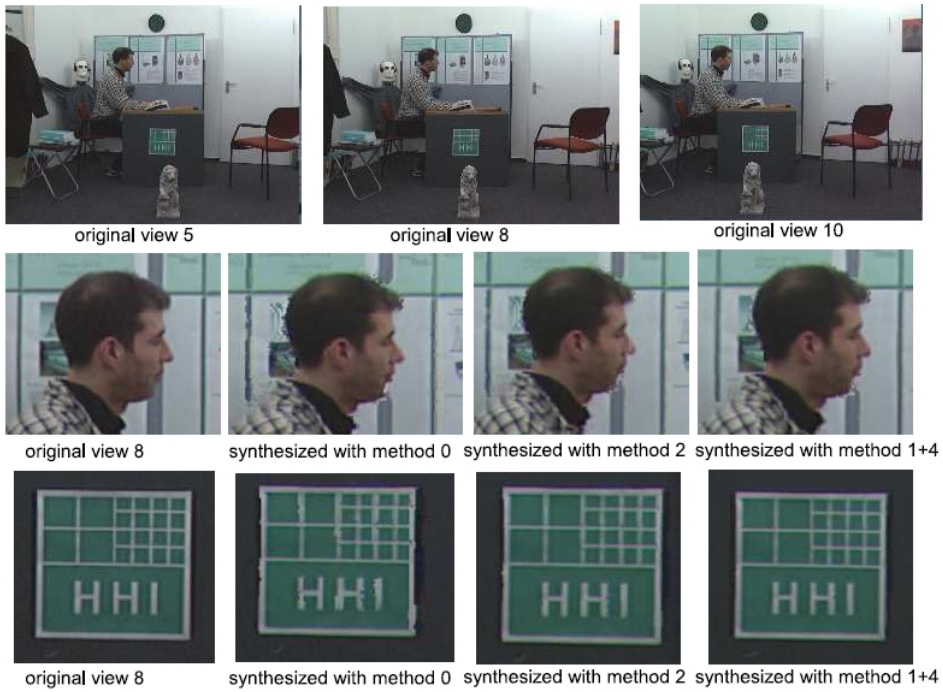
After solving: $u_2 = u_1 + \frac{f \times t_x}{z}$ and $v_2 = v_1$, where $(u_2, v_2) = (u_2'/w_2', v_2'/w_2')$.

Blending

Blend two warped pixels from different reference views to one final image:

1. If the depth difference is within a given threshold ϵ: color values are linearly combined with weights based on camera's positions.
2. Otherwise the color of the pixel closer to the camera is set as the target pixel color

ϵ provides some tolerance to noisy depth.



Sample frames (synthesize view 8 from view 5 and 10 in the "door flower" sequence).

From Z. Ni et al. *IMPROVING THE QUALITY OF DEPTH IMAGE BASED RENDERING FOR 3D VIDEO SYSTEMS*, 2009.

Ghosting

Definition (ghost)

Ghosts are artifacts due to projection of pixels that have background depth and mixed foreground/background color.

Problems due to...

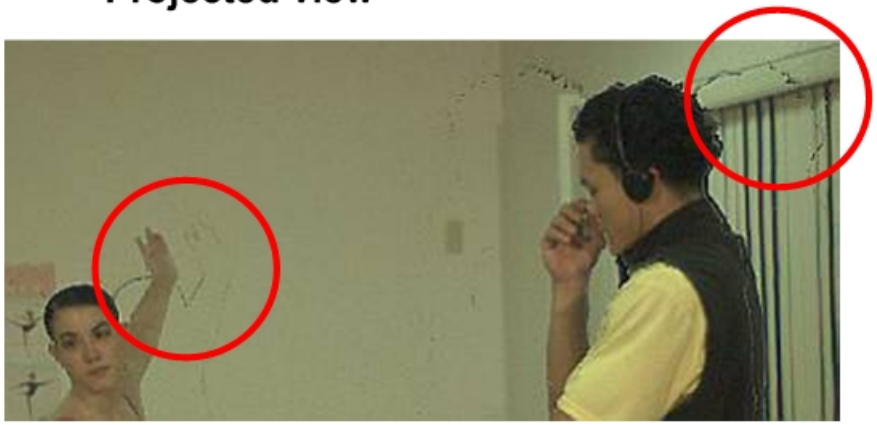
- Accuracy of depth acquisition
- Accuracy of depth estimation
- Depth compression
- etc...

Ghosting

Pixels around depth discontinuities are unreliable.

Projected view

Projected view

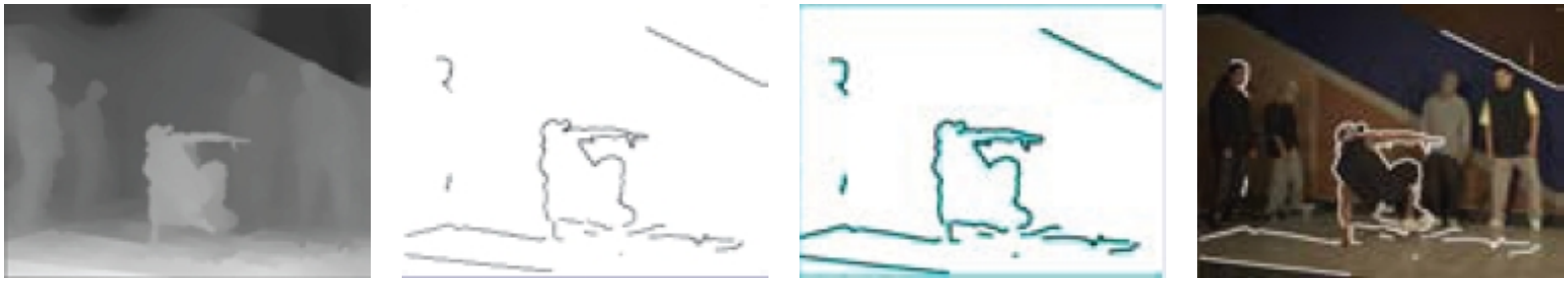


Virtual view

Ghosting removal

There exist a number of methods to remove ghost artifacts:

- by detecting depth discontinuities on the depth map in order to separate pixels near a boundary from pixels far from a boundary.
  - → Detect depth discontinuity [Canny];
  - → Classify background and foreground pixels near each boundaries.
  - → Ignore background blended pixels from data.



Depth map / Canny / Boundaries / Ghost removal

Occlusion - Disocclusion and cracks

Definition (Occlusion - Disocclusion)

- Occlusion : disappeared region, overlapped region
- Disocclusion : revealed region, exposed region

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

Definition (Cracks)

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

$$Z = \frac{1}{\frac{d}{255}\left(\frac{1}{Z_{NEAR}} - \frac{1}{Z_{FAR}}\right) + \frac{1}{Z_{FAR}}}$$

where Z float value and d integer value.

Pre-processing

Pre-processing the depth video allows to reduce the number and the size of the disoccluded regions.

The goal is to reduce the depth data discontinuities.

- General Smoothing : to remove sharp discontinuities from the depth data.
  - → the perceived stereo image quality is improved by increasing the smoothing strength of the depth map [Tam et al.,2004]
- Bilateral filtering: preserves the sharp depth changes in conjunction with the intensity variation in color space.

Post-processing

Some disocclusions may still remain, which require a next stage, consisting in interpolating the missing values.

The process of filling in the disocclusions is also known as hole-filling.

- Multiple reference images: filling the holes by preprocessing depth images and merging two desired images. [Zhan-wei et al., 2007].
- Depth-based inpainting

Depth-based inpainting

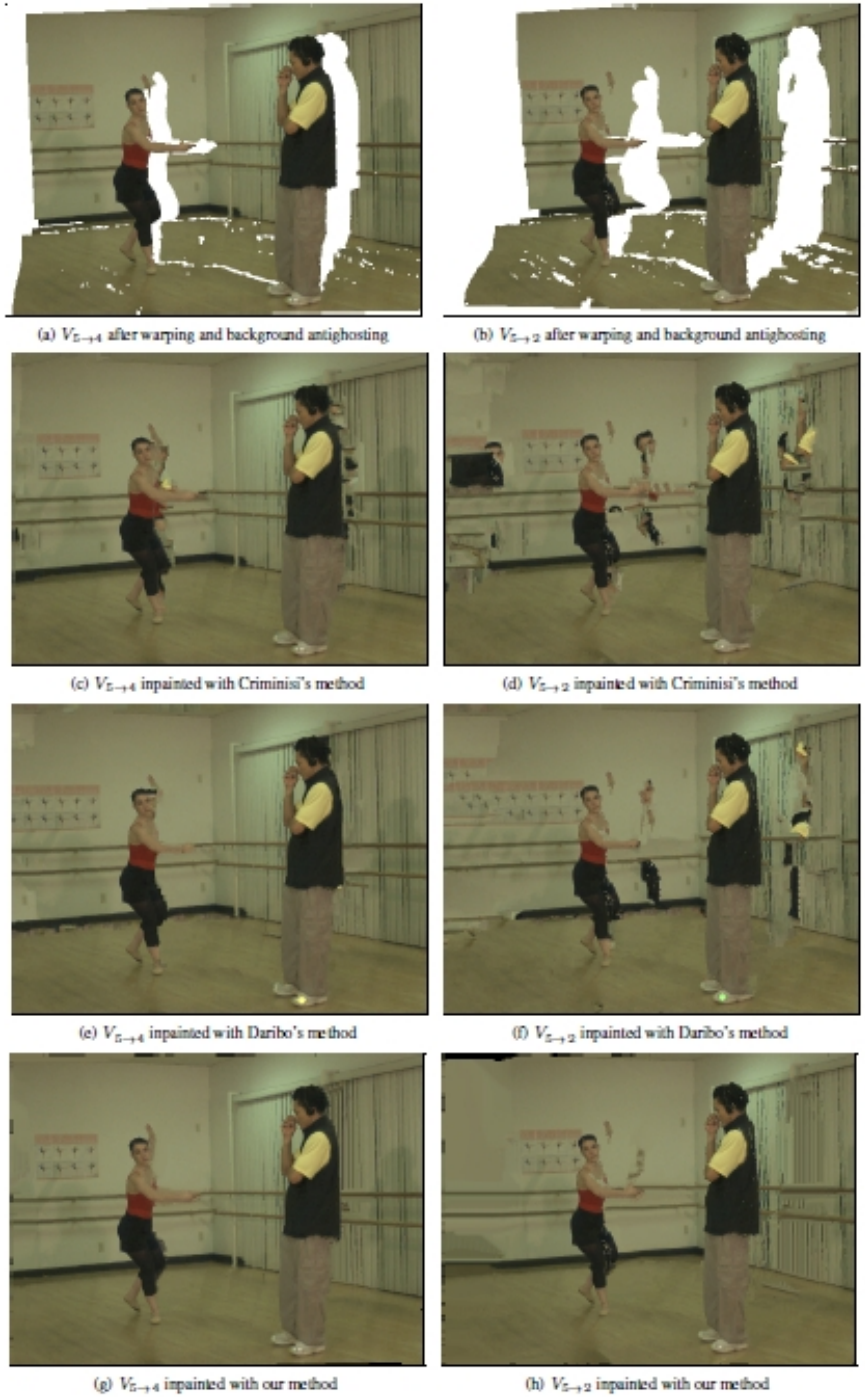
Extension of examplar-based inpainting by using depth map [Daribo & Pesquet, 2010, Gautier et al., 2011].

Criminisi's approach = filling order followed by templated matching

- Modification of the priority computation by using the depth:

$$P(p) = C(p) \times D(p) \times Depth(p)$$

- Modification of the template matching, distance computed on $\{R, G, B, Z\}$.



(a) $V_{5\to4}$ after warping and background antighosting

(b) $V_{5\to2}$ after warping and background antighosting

(c) $V_{5\to4}$ inpainted with Criminisi's method

(d) $V_{5\to2}$ inpainted with Criminisi's method

(e) $V_{5\to4}$ inpainted with Daribo's method

(f) $V_{5\to2}$ inpainted with Daribo's method

(g) $V_{5\to4}$ inpainted with our method

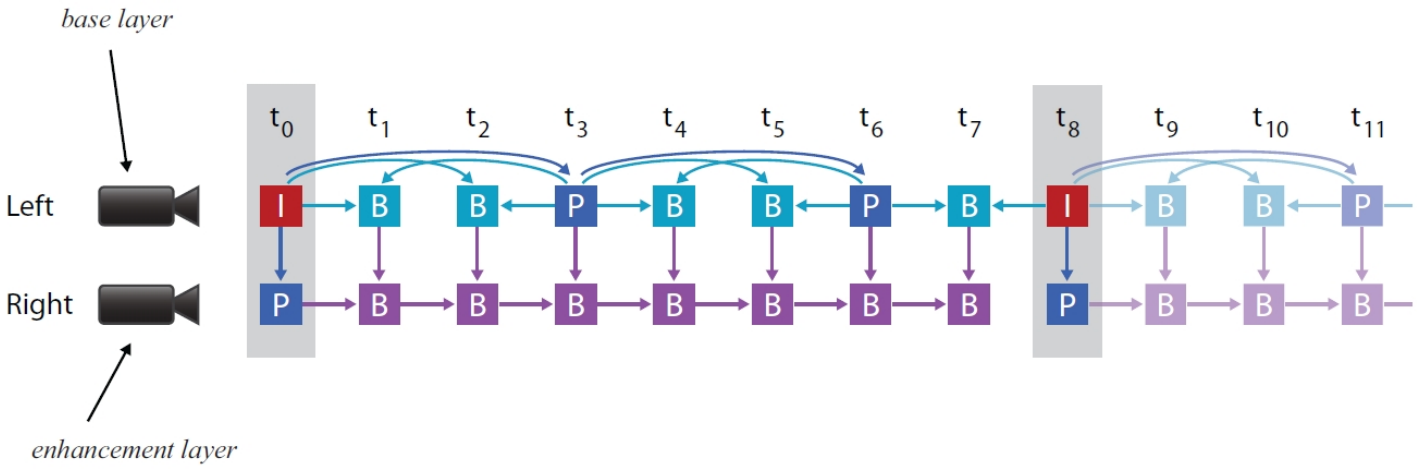
(h) $V_{5\to2}$ inpainted with our method

From [Gautier et al., 2011].

3D Compression

MPEG-2 Multiview profile

- Multiview profile:
  - → based on the scalable coding tools (backward compatibility with the MPEG-2 Main Profil)
  - → conveys the camera parameters (i.e. geometry info, focal length...).
  - → base + enhancement layers (base layer is encoded in conformance with the MP).

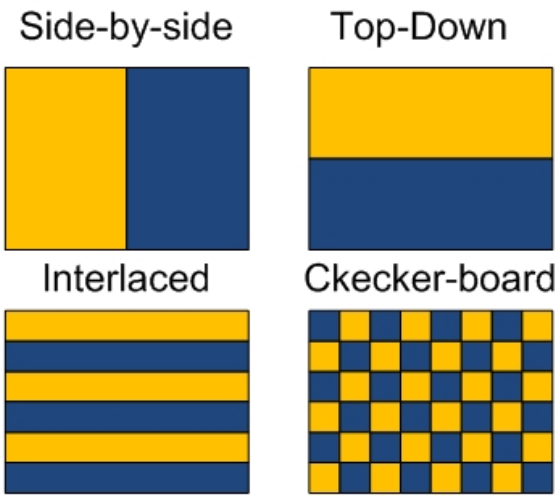


Prediction structure in MPEG-2 MVP using a GOP structure IBBP.

H.264 stereo SEI message

As previously described, a stereo video data may be represented by stereo interleaving techniques :

- time multiplexing;
- spatial multiplexing.



Decoder must know how to decode: SEI = Supplemental Enhancement Information.

The backward compatibility is not supported. If the transmitted bitstream is not demultiplexing, it is not possible for traditional 2D devices to extract, decode and display a 2D version of the 3D video content.

MPEG-C Part 3

Responding to a strong industry demand in 3D video content generation, standardization of the video-plus-depth coding (April 2006):

- encoding of a 3D content inside a conventional MPEG-2 transport stream:
  - → Texture video + depth video + auxiliary data

A number of advantages: display technology independence, backward compatibility, compression efficiency...

The total bandwidth for video-plus-depth data transmission is reduced compared to the stereo video data...
10 to 20% of the total bit rate for the depth coding.

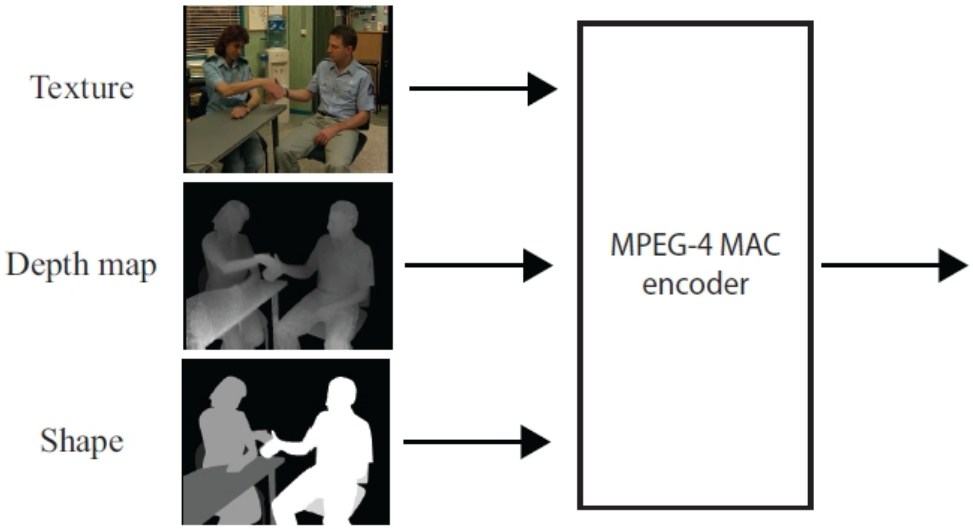
Remark: MPEG-2 can be replaced by H.264.

MPEG-4 MAC

Coding the video-plus-depth by using MPEG-4 MAC (Multiple Auxiliary Components):

- The MAC is the grayscale shape (used to describe transparency)

The idea is to use the MAC channel to describe depth shape, others secondary texture ($\approx$ ALPHA channel).

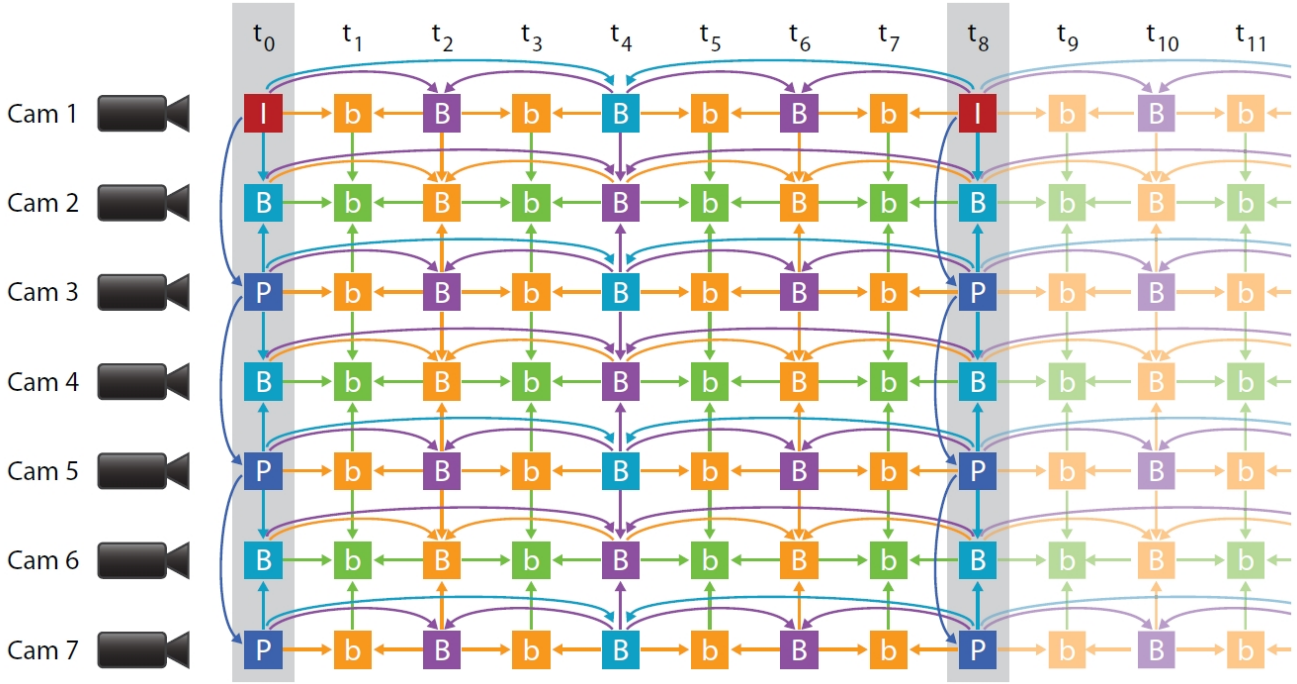


MPEG-4 MAC architecture

Auxiliary components are classically encoded (motion compensation and DCT)

MVC coding

Extension of H.264 to deal with two or more numbers of view (April 2008): inter-view prediction + temporal prediction scalability

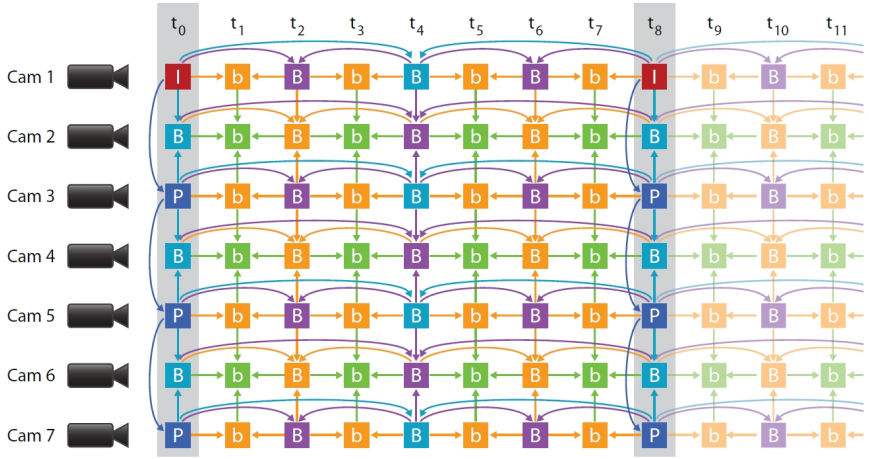


Typical MVC prediction structure.

Remarks:

- the view 4 depends on the views 3 and 5,
- the view 5 depends on the view 3,
- the view 3 depends on the view 1.

MVC coding



- The base view: the first view is independently coded;
- The non-base view: they are encoded by motion and disparity compensation (insertion of temporal key-picture (V-picture) to ease the random access for each view);
- Random access and view switching refers to to the possibility to access a given picture in a required view at a specified time through minimal decoding of other pictures in dependent views. Random acces points
  - → Anchor pictures in the base view
  - → V-pictures in the non-base views
- Temporal and inter-view prediction, to take advantage of the redundancies among the inter-pictures.
- Inter-view SKIP mode, motion information of the current MB is derived from the corresponding MB in the picture at the same temporal index of the neighboring view;
- Compensation of illumination

Suggestion for further reading...

[Daribo & Pesquet, 2010] I. Daribo & B. Pesquet-Popescu. Depth-aided image inpainting for novel view synthesis, in IEEE International Workshop on Multimedia Signal Processing, 2010.

[Dufaux et al.,2007] F. Dufaux, M. Ouaret & T. Ebrahimi, RECENT ADVANCES IN MULTI-VIEW DISTRIBUTED VIDEO CODING, DSS, 2007.

[Gautier et al., 2011] J. Gautier, O. Le Meur & C. Guillemot, Depth-based image completion for view synthesis, 3DTV Conf. 2011.

[Tam et al.,2004] W.J. Tam, G. Alain, L. Zhang, T. Martin & R. Renaud. Smoothing depth maps for improved steroscopic image quality. In Bahram Javidi and Fumio Okano, editors, Three-Dimensional TV, Video, and Display III, volume 5599, pages 162-172, 2004.

[Zhan-wei et al., 2007] L. Zhan-wei, A. Ping, L. Su-xing & Z. Zhao-yang. Arbitrary view generation based on DIBR. In Proc. of the International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), pages 168-171, 2007.