3D Video

From Capture to Diffusion

Edited by
Laurent Lucas
Céline Loscos
Yannick Remion

Color Section

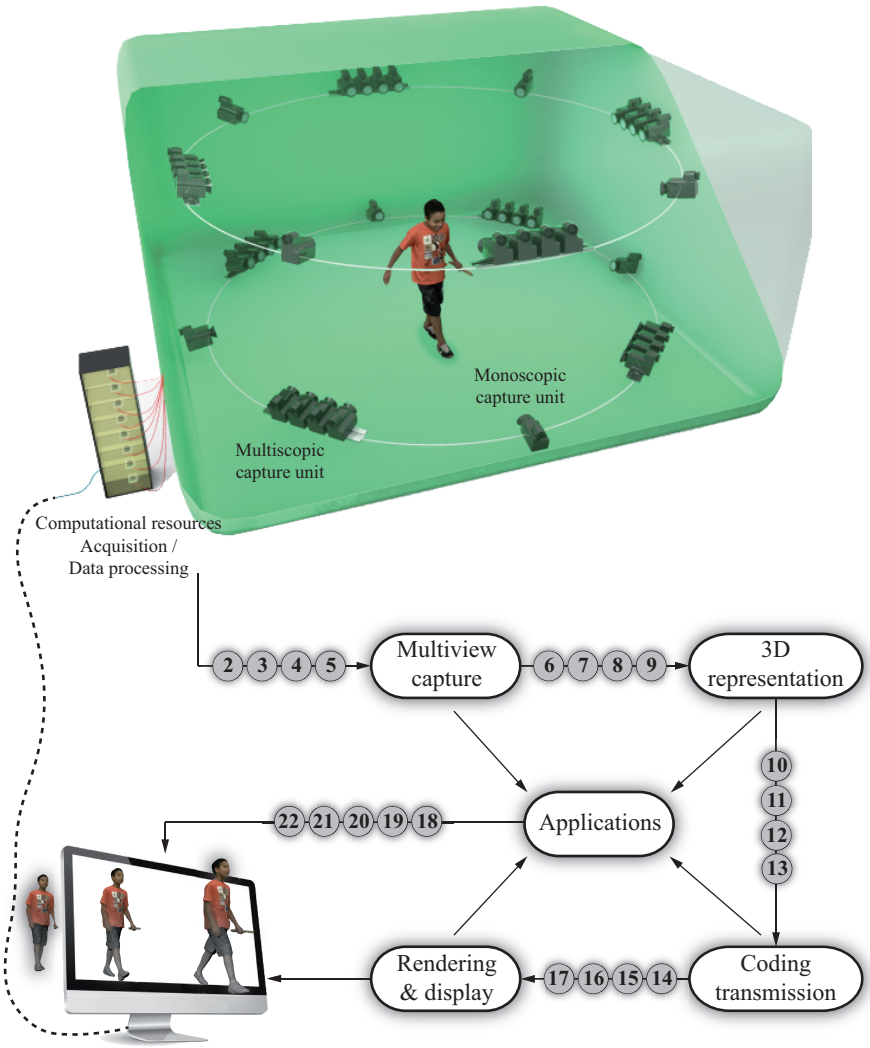


Figure I.1. *Organization of this book: the numbered chips correspond to the different chapters*

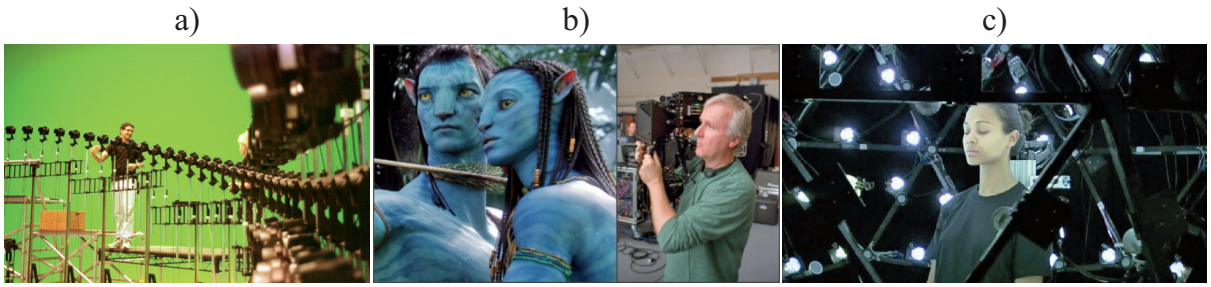


Figure 1.1. *Multiview system used to film The Matrix©Warner Bros. Entertainment Inc. a): 120 still cameras and two video cameras enabling time slicing (bullet time effect); b): stereoscopic filming; c): omnidirectional 3D capture for Avatar©20th Century Fox by James Cameron*

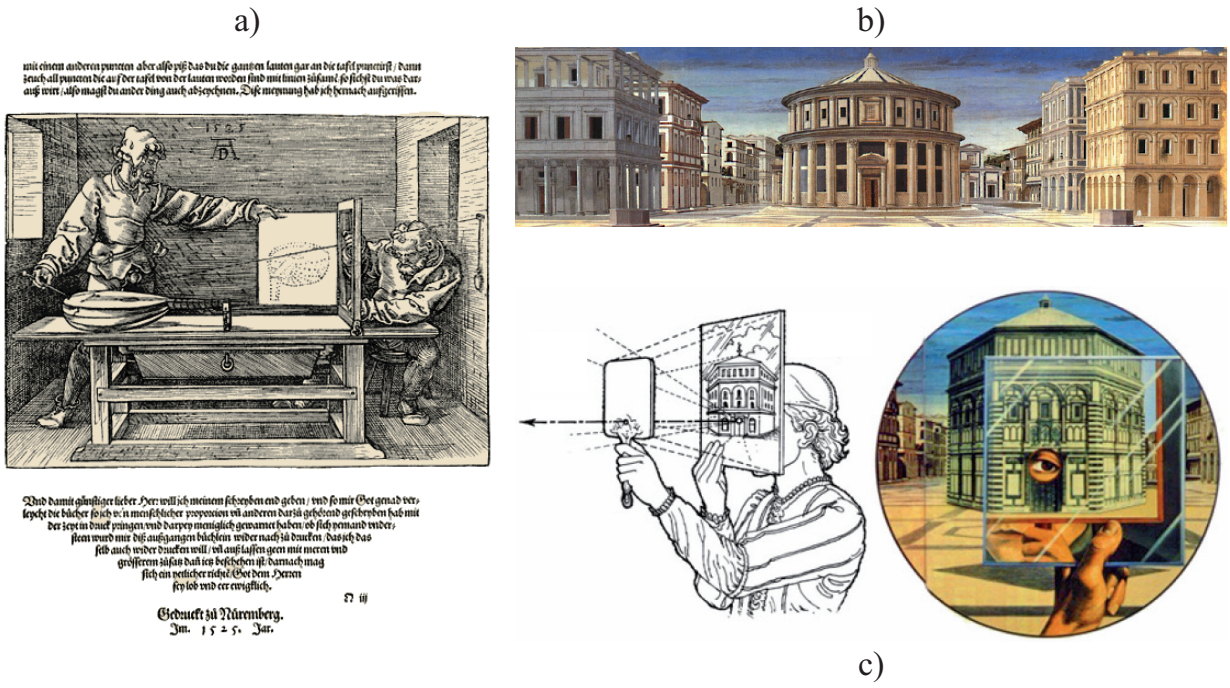


Figure 1.2. *a): the Dürer perspectograph; b): the ideal city (1475) from Piero della Francesca, c): Brunelleschi experiment*

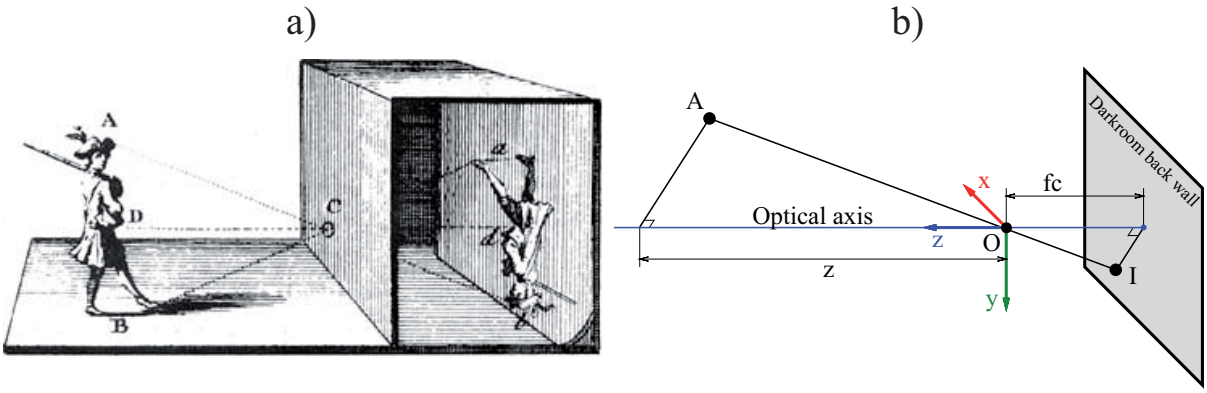


Figure 1.3. *A pinhole camera (camera obscura): a): illustration from The Encyclopedia of Diderot & d'Alembert; b): geometric model of the central projection involved*

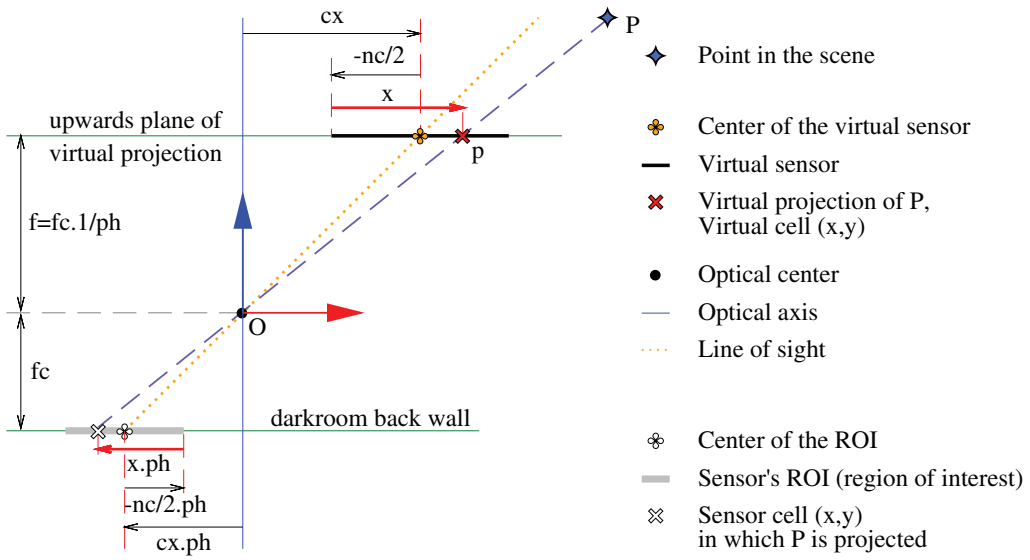


Figure 1.4. *Decentered and normalized reverse pinhole: from above (according to the y-axis), the projective geometries of the real sensor and its normalized virtual representation*

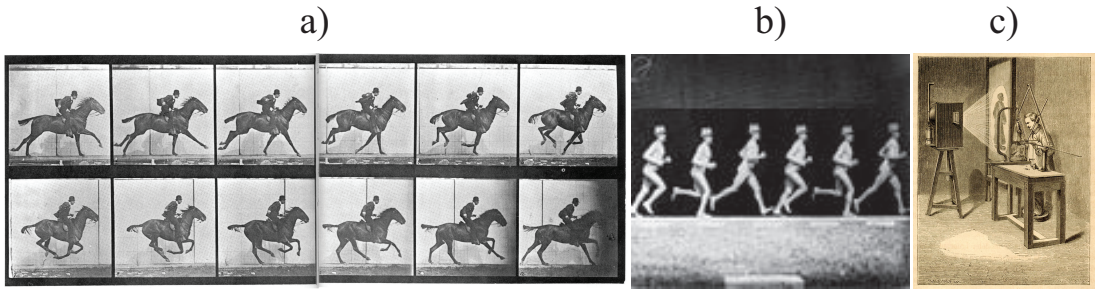


Figure 1.6. *a): Initial sequences of images with Muybridge's multiview systems; b): Marey's device superposing successive shots by a single camera; and c) the photosculpture procedure projecting 24 images on a screen connected to a pantograph*

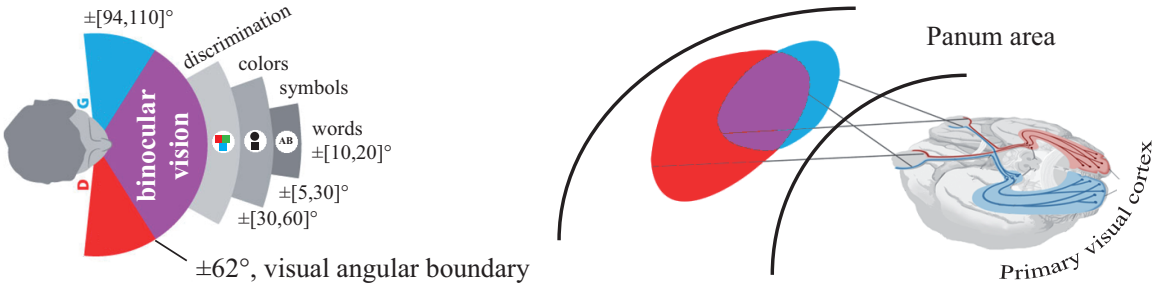


Figure 1.7. *Physiological limits and description of the human visual field; Panum's area indicates the area in which two images are fused to provide a single perception*

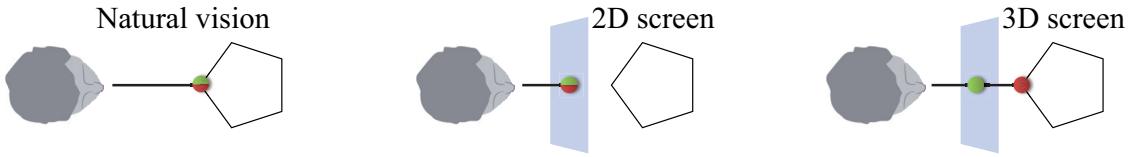


Figure 1.9. *Visual exploration using the convergence-focus reflex (• the focus point, • the convergence point)*

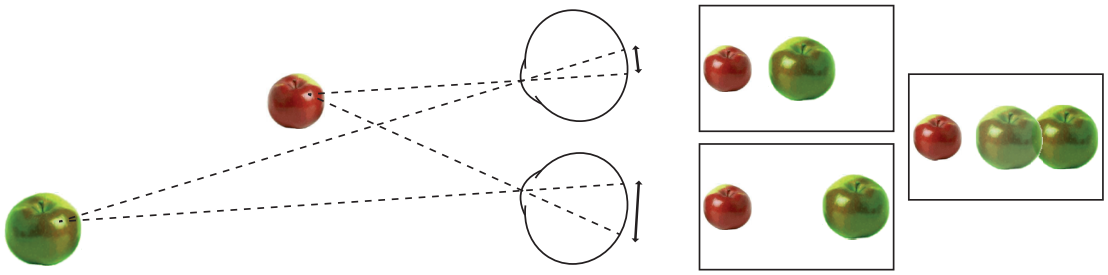


Figure 1.10. *Fusion and disparity in retinal images. Disparity accounts for the fact that an image is projected onto different places on the two retinas. More than being a mere stimulus to vergence movements, the disparity between images from the two eyes provides indications about the depth of objects in the scene*

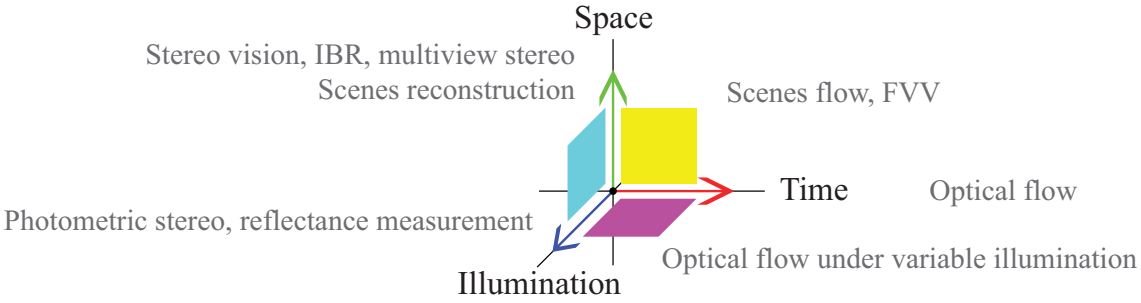


Figure 1.11. *Thematic classification in the Vision Field Calculating Space (according to [DAI 10])*

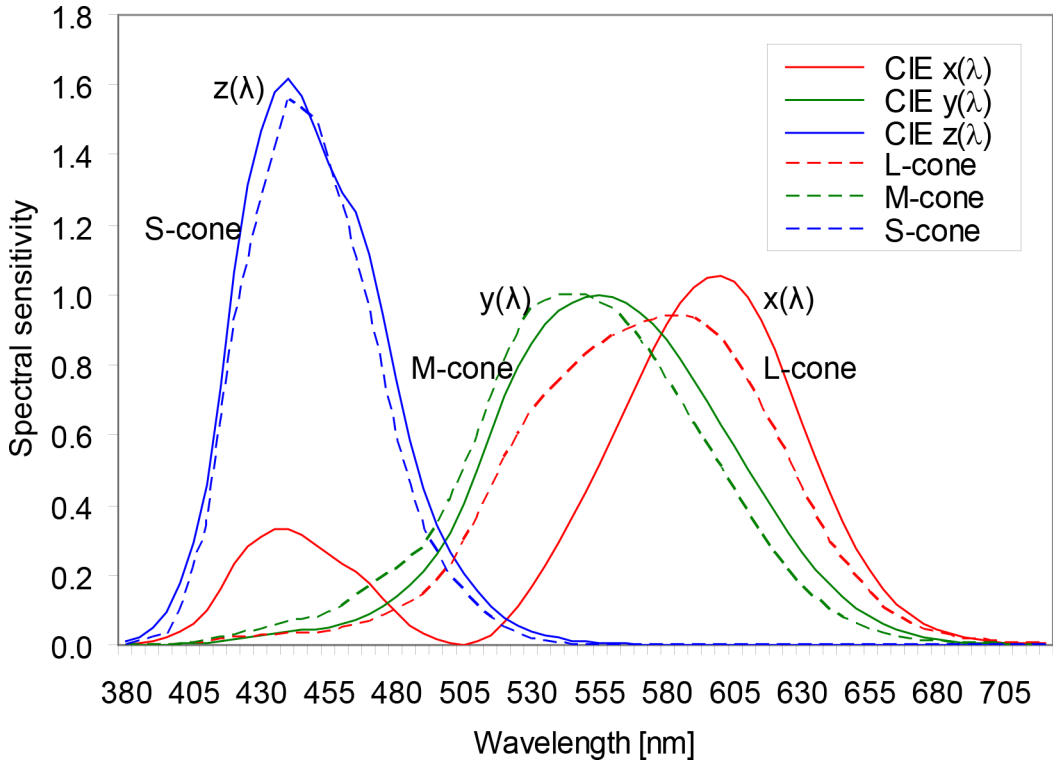


Figure 2.1. *The CIE color-matching functions (CMFs) versus the spectral sensitivities of the human cone: the solid lines represent the CIE 1931 CMFs (version: modification [VOS 78]), the dotted lines show the different physiological responses of the cones [EST 79]. In particular, the responses to the L-cone appear to be significantly different in relation to the CIE function x(λ)*



Figure 2.2. *Illustration of fall-off (digital effect accentuated by an improvement in contrast) and distortion in the image created by the lens used*

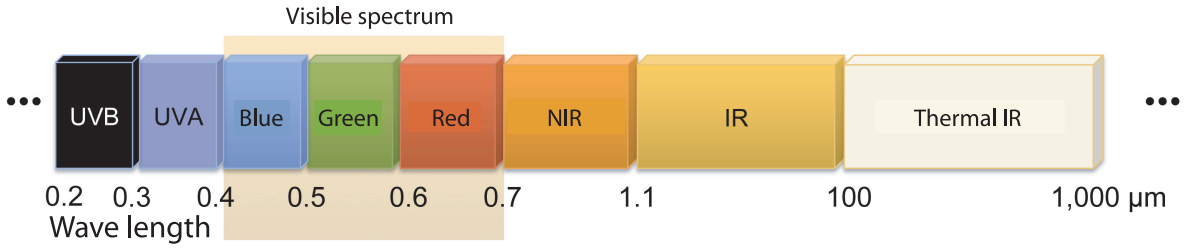


Figure 2.3. *Range represented by hyperspectral imaging*



a) b)

Figure 3.2. *Examples of rigs: a) Binocle Brigger III in a studio configuration, a robotized rig for 3DTV; b) a heliborne rig with four cameras used by Binocle for the film La France entre ciel et mer*



a) **b)**

Figure 3.3. *Examples of integrated cameras: a) a Cam-Box prototype camera with eight integrated perspectives developed by 3DTV Solutions and the University of Reims and b) the Lytro plenoptic camera*

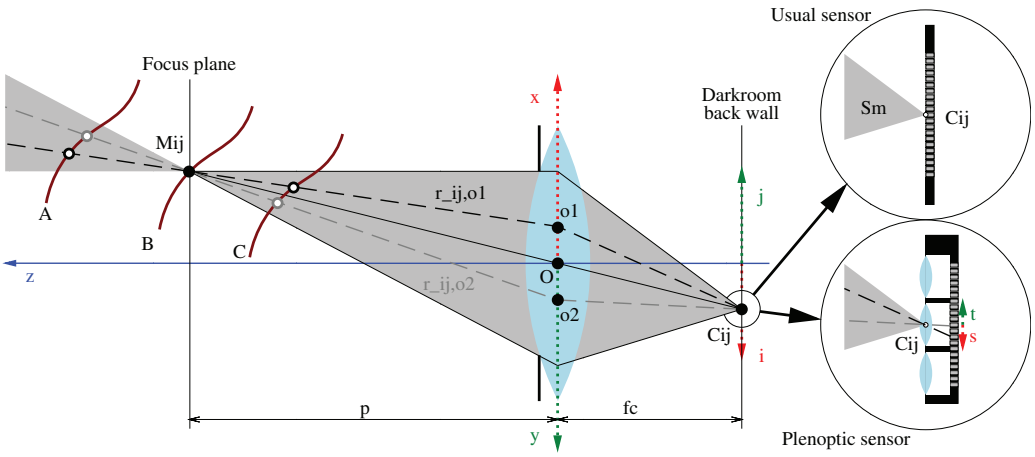


Figure 3.4. *Differences between standard and plenoptic cameras: from above (axes x, j, s) or the side (axes y, j, t) the rays converging as a single point at the back wall of the darkroom are summed in the first and differentiated by refraction and sampling in the second*

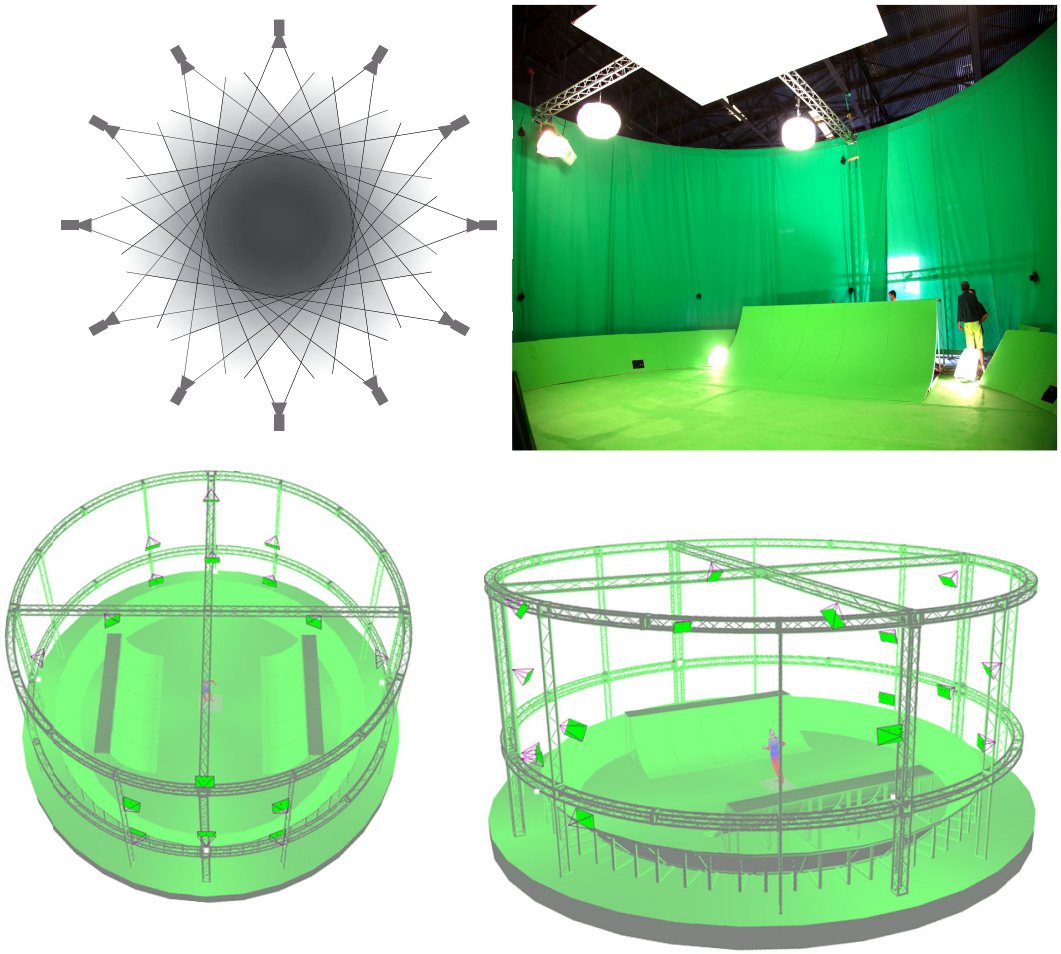
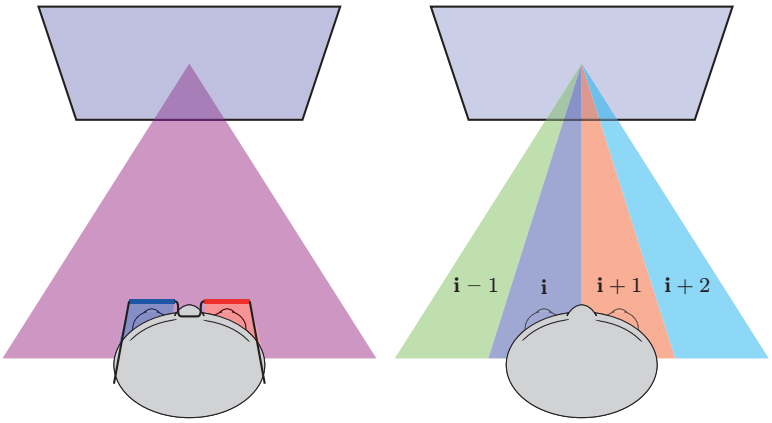


Figure 3.5. *Examples of 3D video studios: from top left, circular arrangement of 12 cameras showing the scenic space used as an intersection of camera field depth zones (in light gray); top right and below, the studio of the Recover3d project*



a) A stereoscopic system: a single beam transporting two images that are physically separated by glass

b) An autostereoscopic system: several distinct optical beams, each transporting an image

Figure 4.1. *Image transport according to the technology used: glasses-based stereoscopy and autostereoscopy*

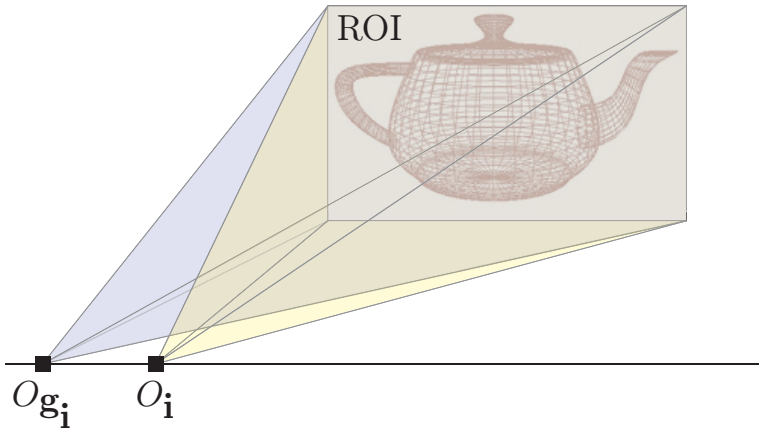


Figure 4.2. *Image viewing pyramids: the shared base is the device'sROI and the apices are the users' eyes O_i and O_{gi}*

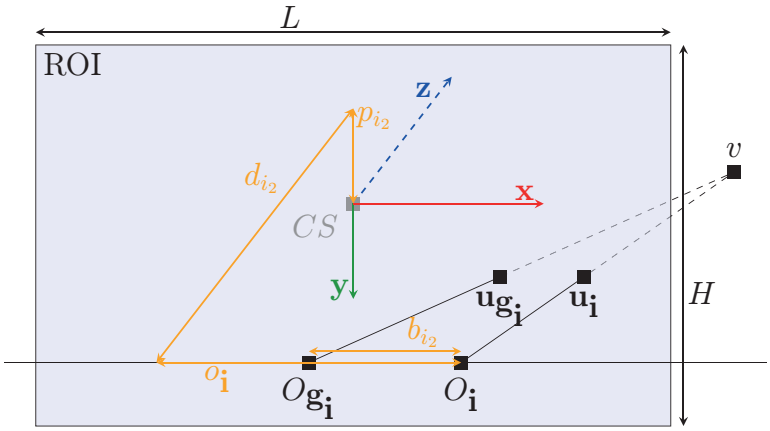


Figure 4.3. *Characterization of geometry of the 3D multiscopic display using colocalized planar mixing*

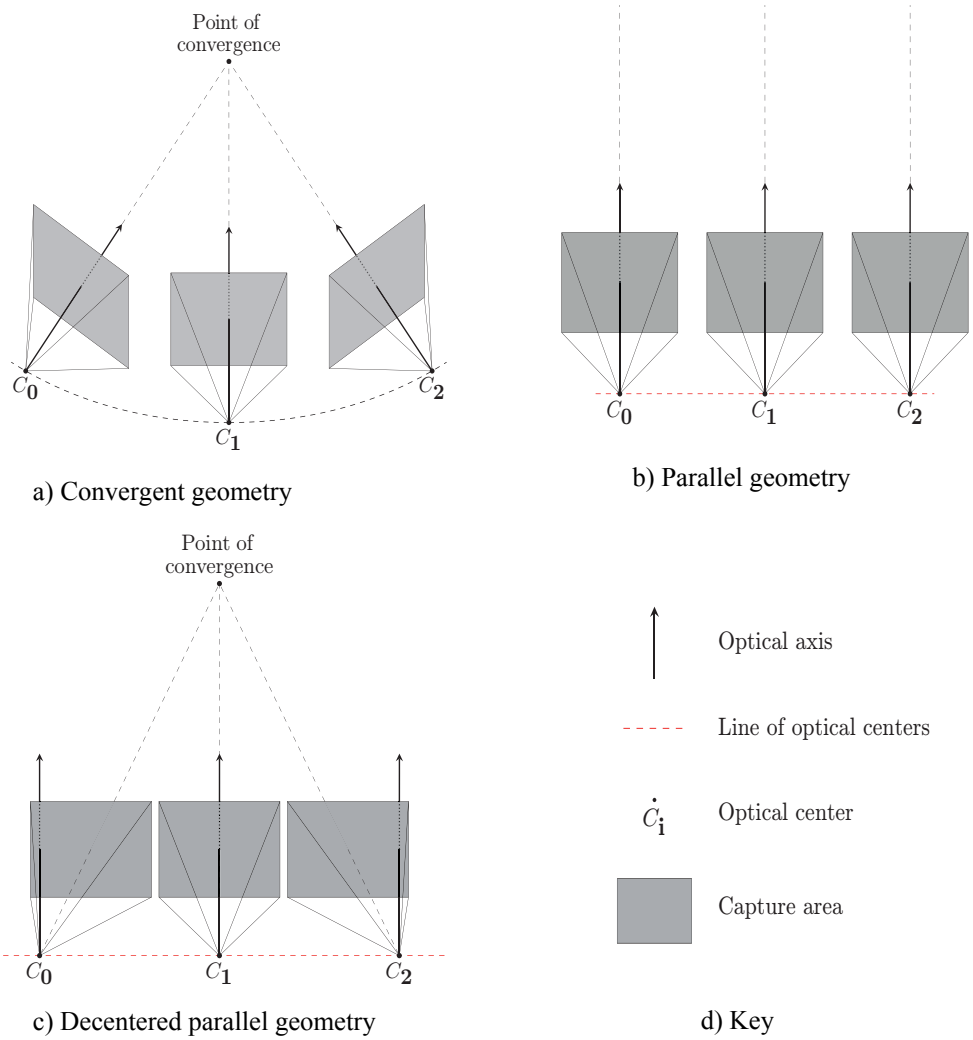


Figure 4.4. *The different shooting geometries represented in reversed pinhole model*

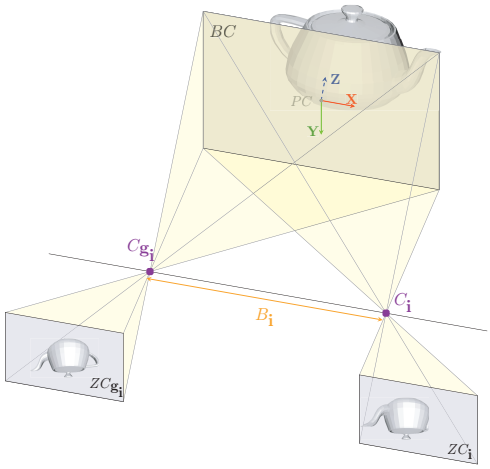
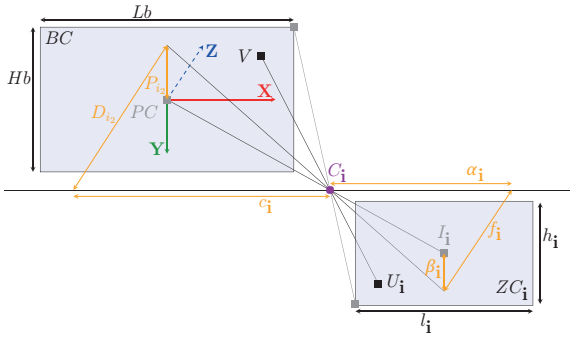
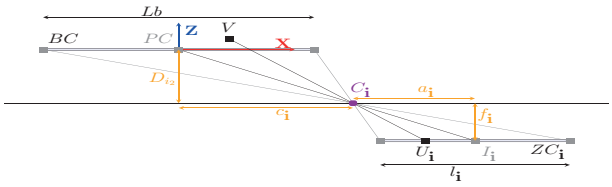


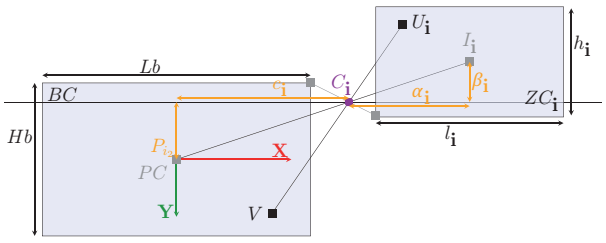
Figure 4.5. *Generic description of a decentered parallel geometry*



a) Perspective view



b) View from above



c) Front view

Figure 4.6. *Characterization of the decentered parallel shooting geometry*

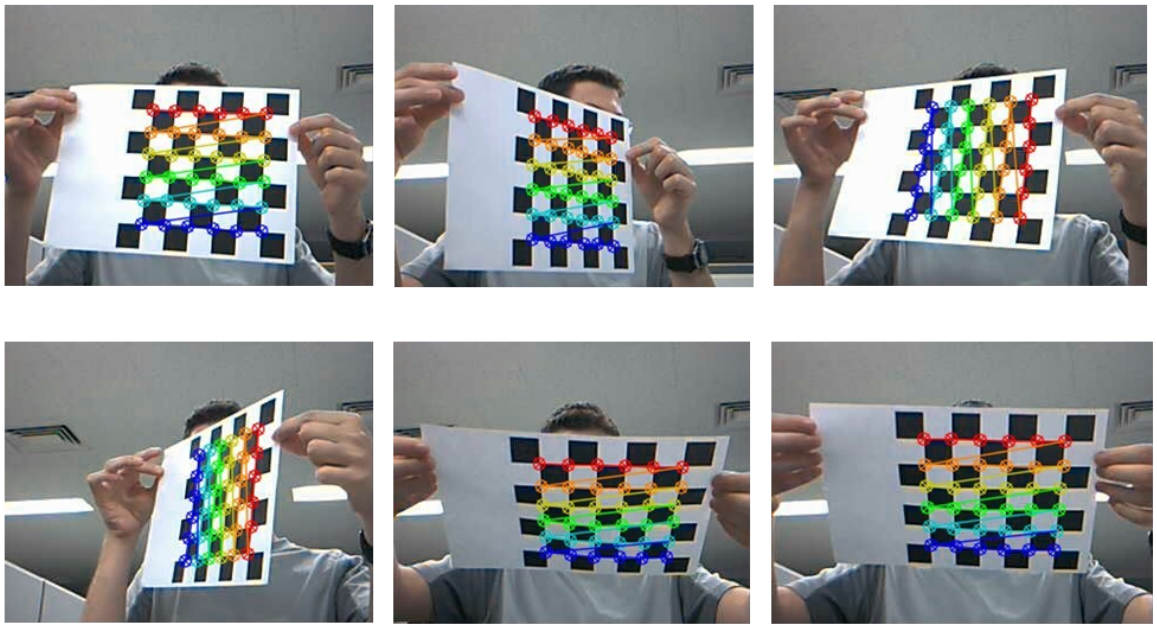


Figure 5.1. *Calibration with a chessboard photographed from several angles to show the leak points generated by the camera's perspective. The chessboard is automatically detected by the OpenCV library*



Figure 5.3. *Correction of radial distortion*



a)



b)

Figure 5.4. *Rectification of two images. a) The original images. b) The corrected images. Each pixel in an image is aligned horizontally with its counterpart in the other*



Figure 5.6. *There are infinite solutions to the image correction problem*



Figure 5.7. *Correction of several images*

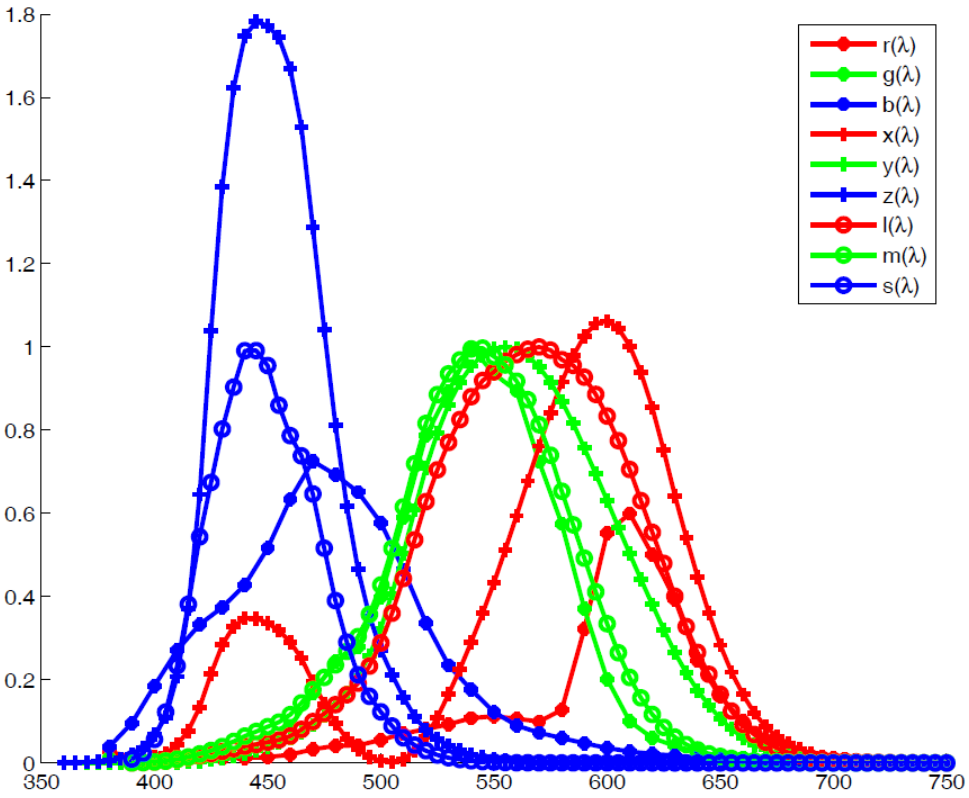
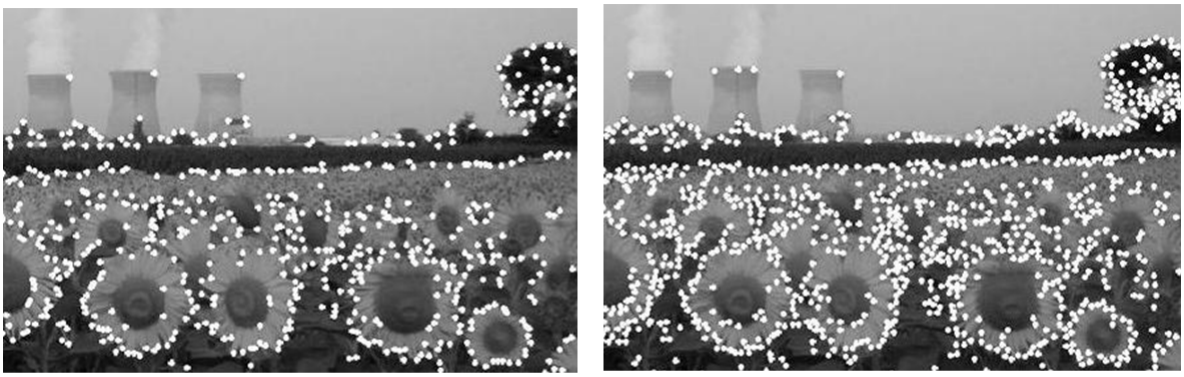
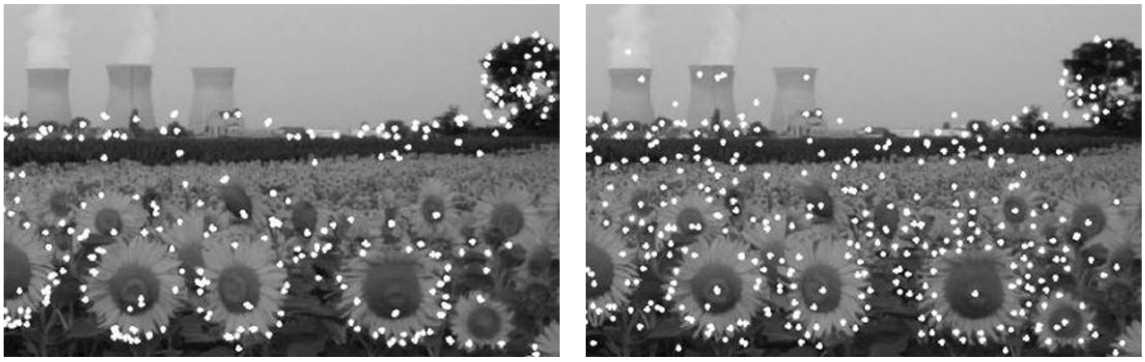


Figure 5.8. *Spectral sensitivities of cones (LMS, normalized by M), of primaries defined by the International Commission on Illumination (CIE), $\overline{x}, \overline{y}, \overline{z}$ and of a camera r, g, b (Sinarback 54 Camera [DAY 03])*



a) Harris b) Hessian

Figure 6.1. *Feature points based on derivatives*



a) SIFT: Multi-scale b) SURF

Figure 6.2. *Scale invariant detection*

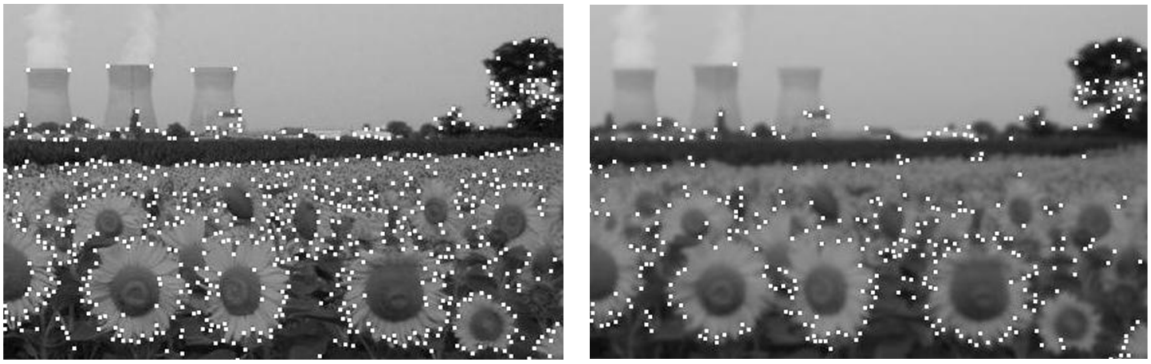
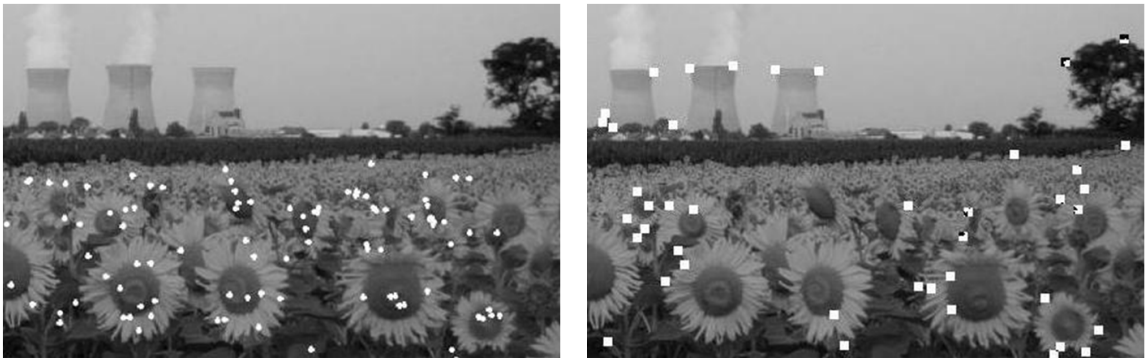


Figure 6.3. *The SUSAN detector*



a) EBR b) CSS

Figure 6.4. *Edge-based operators*



a) IBR b) MSER

Figure 6.5. *Stable region based operators*

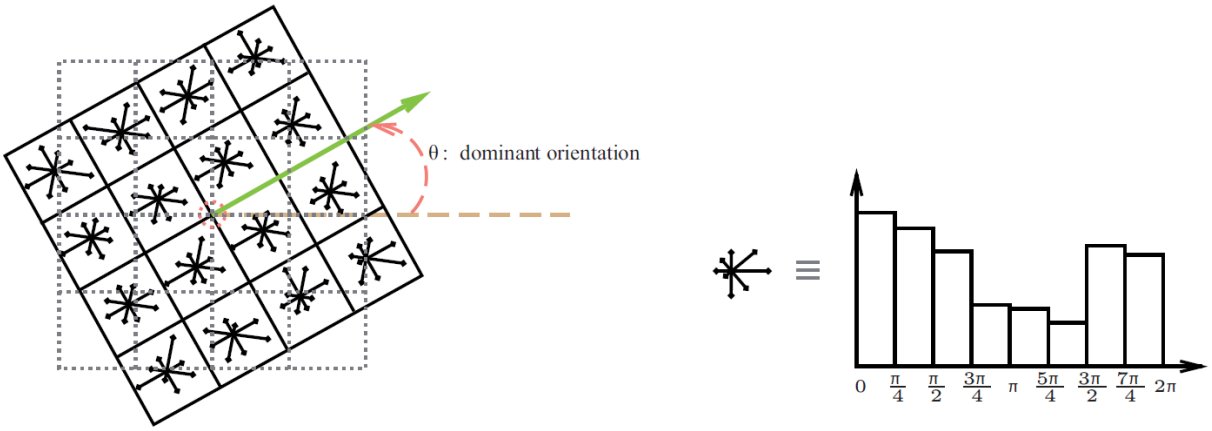


Figure 6.6. *The structure of a SIFT descriptor*

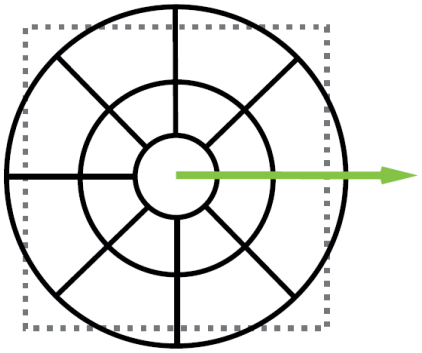


Figure 6.7. *The structure of the GLOH descriptor*

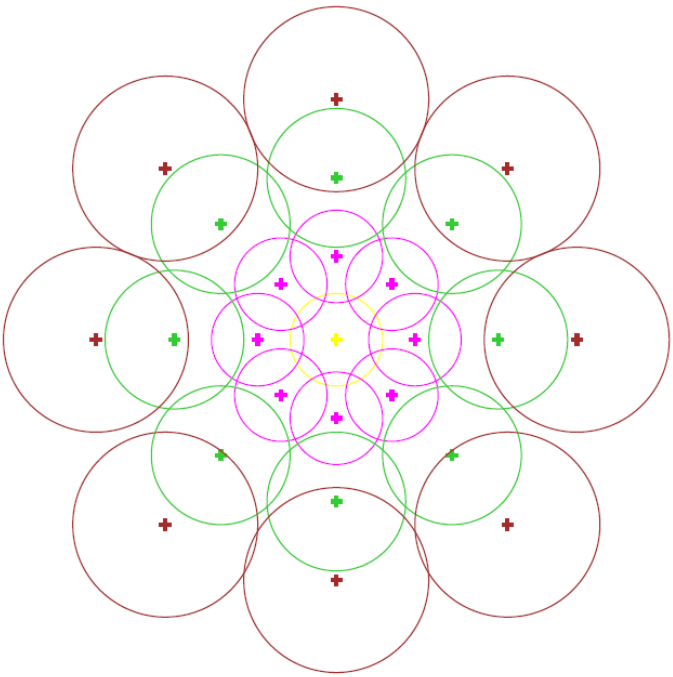


Figure 6.8. *The structure of the DAISY descriptor*

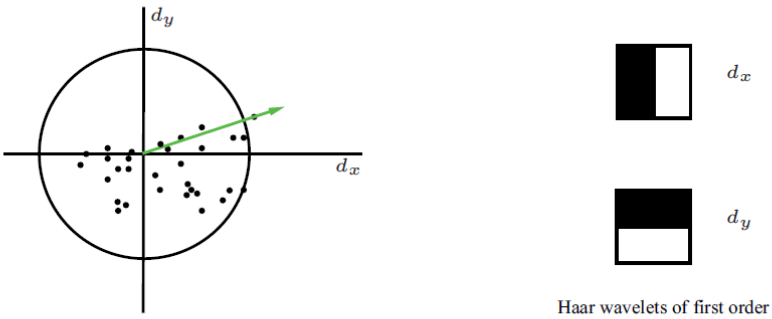


Figure 6.9. *Structure of the SURF descriptor*

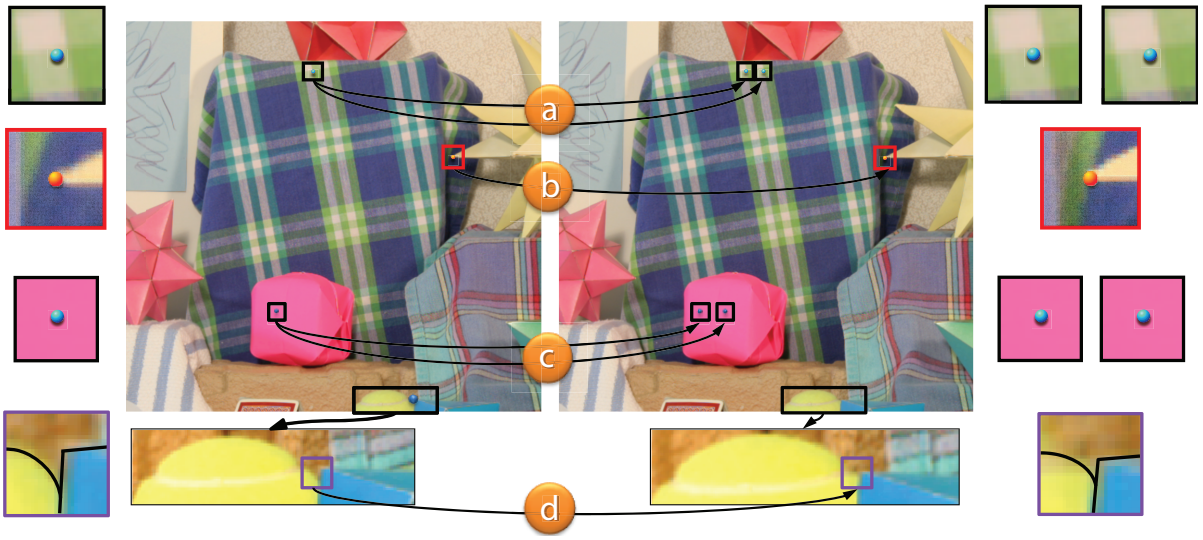


Figure 7.1. *Difficulties in matching. a) Repetitive textures, b) discontinuity, c) homogeneous areas and d) occlusions*

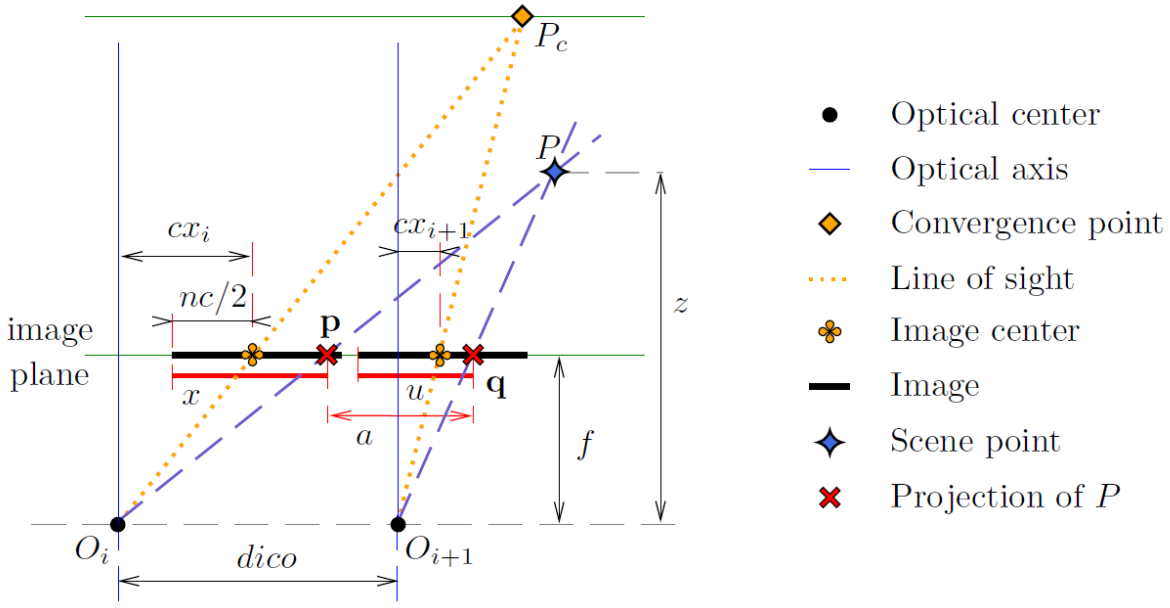


Figure 7.2. *Relation between disparity and depth*

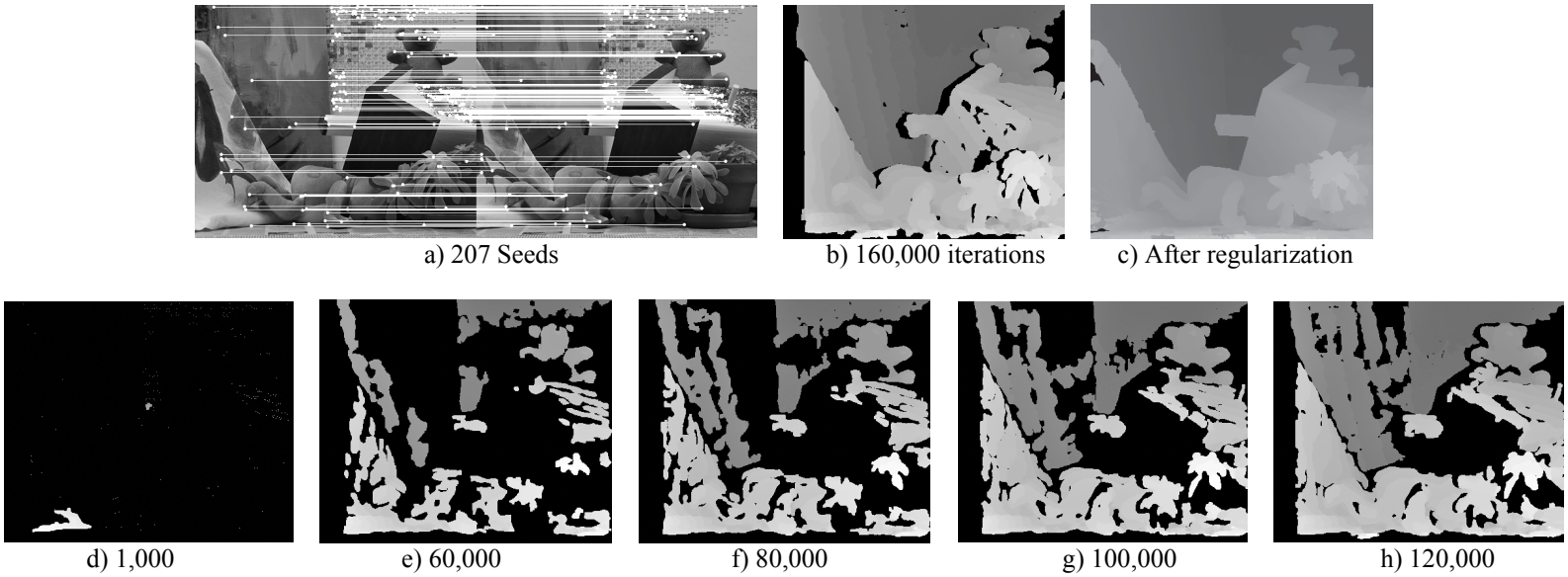


Figure 7.3. *Initial set of seeds a) and disparity maps obtained from different iterations (b, d, e, f, g and h) during sequential propagation. The image c) shows the result obtained after regularization*

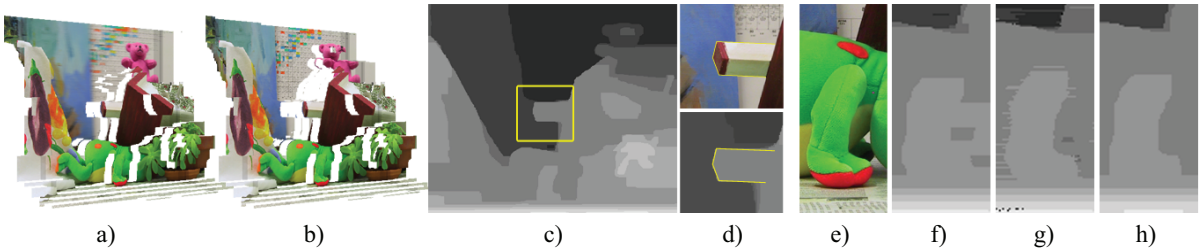
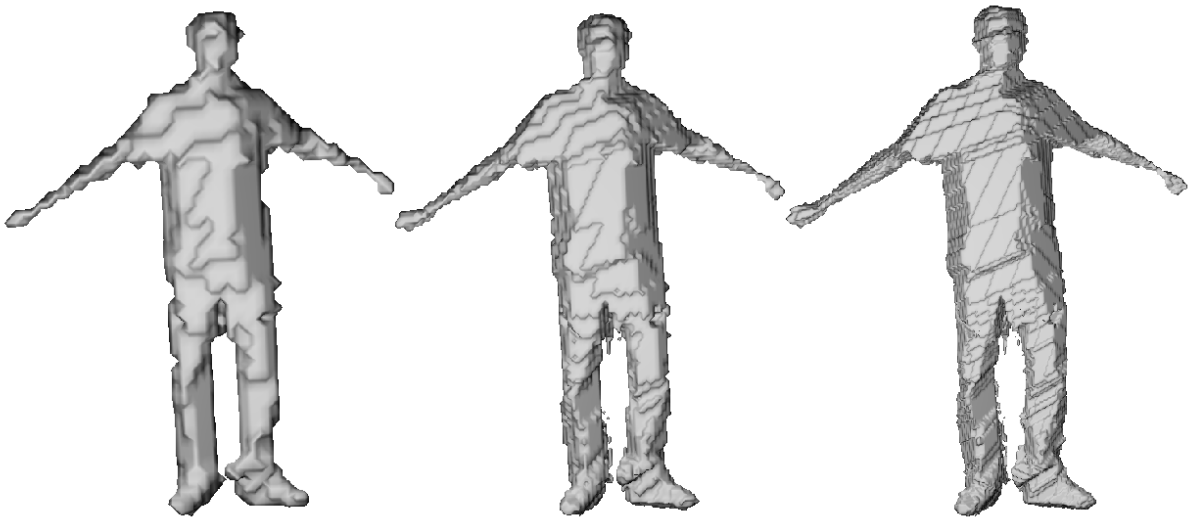
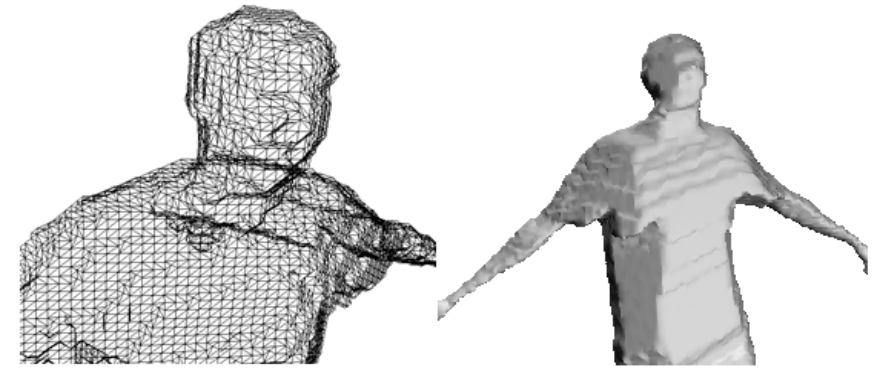


Figure 7.4. *Different stages in reconstruction (a and b), the disparity map and occlusion zoom in with K_{occ} = 100 and K_{liss} = 20 (c and d), increasing from (e) with K_{occ} = 200 (f), K_{liss} = 0 (g) and K_{liss} = 40 (h)*

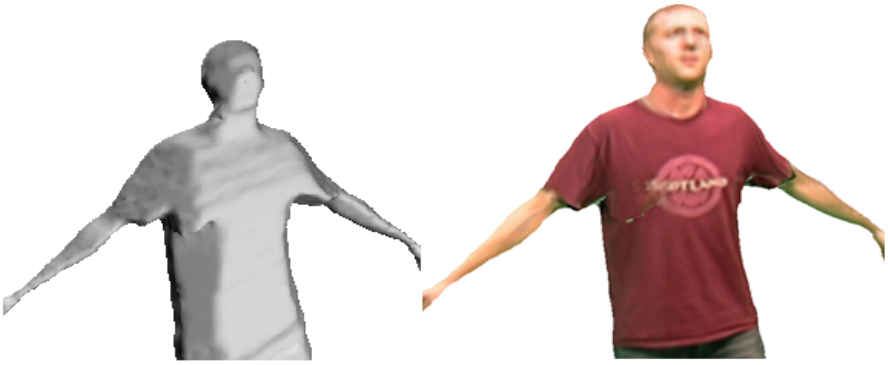


a) 64^3 voxels b) 128^3 voxels c) 256^3 voxels

Figure 8.5. *Models of voxel grids with three different solutions (© 2013 XD Productions)*



a) Marching cubes b) Refinement



c) Relaxation d) Texturing

Figure 8.6. *Stages in the graphic pipeline after reconstruction (©2013 XD Productions)*

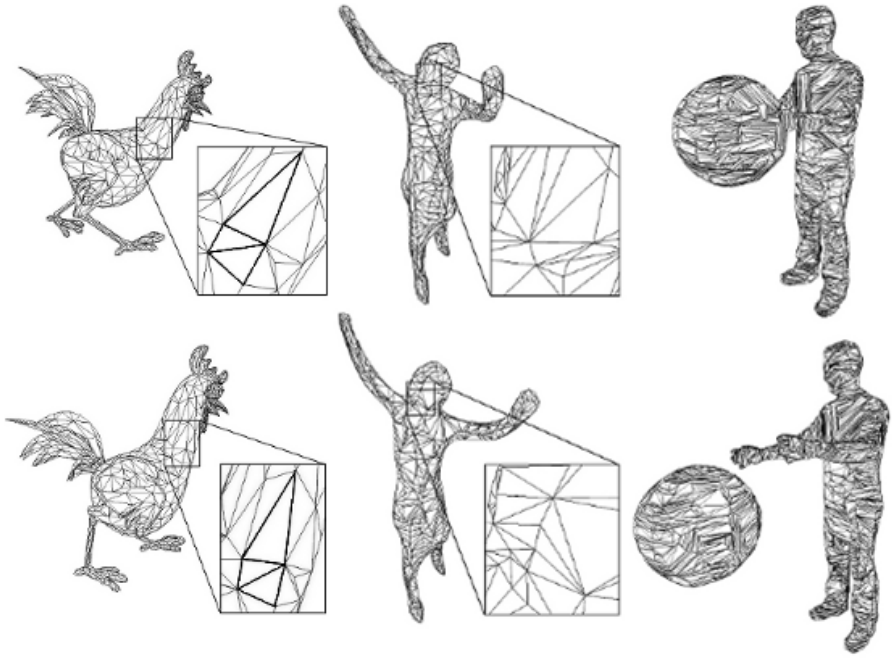


Figure 8.7. *Different types of mesh sequences. a) © 1996 Microsoft Corporation; b) and c) source: GRImage INRIA Rhône-Alps & 4DView Solutions, http://4drepository.inrialpes.fr*

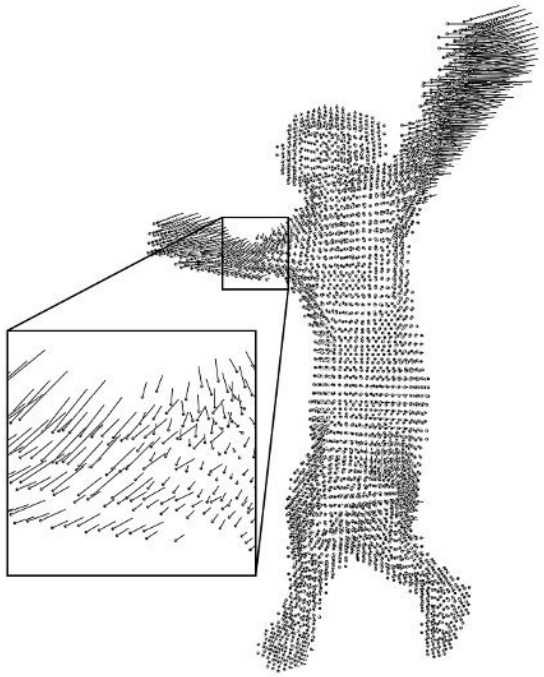


Figure 8.8. *Example of the scene flow*

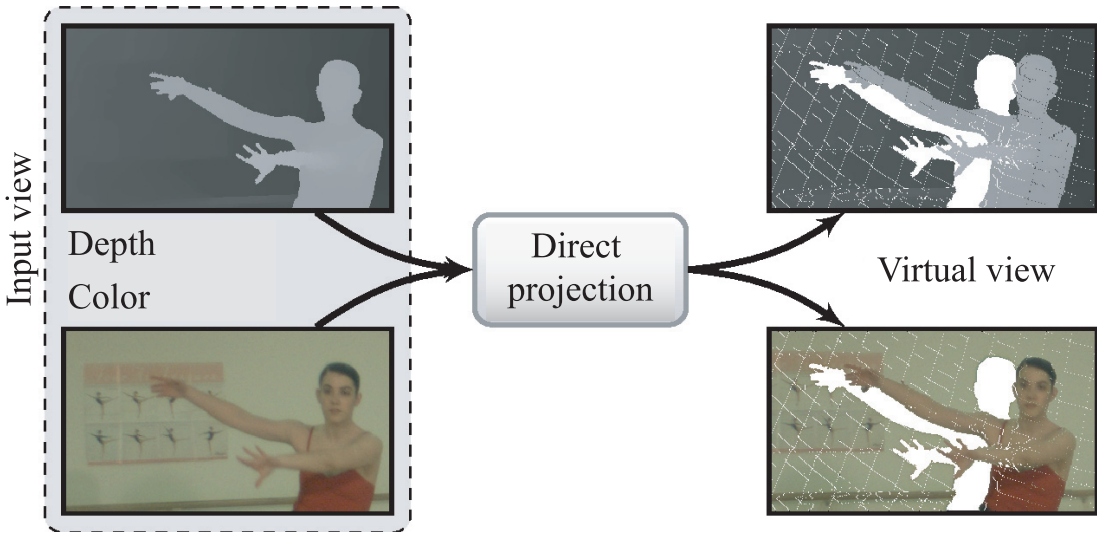


Figure 9.1. *Direct projection*



a) uncovered areas b) cracks c) ghost effect

Figure 9.2. *Distortions commonly associated with direct projection*

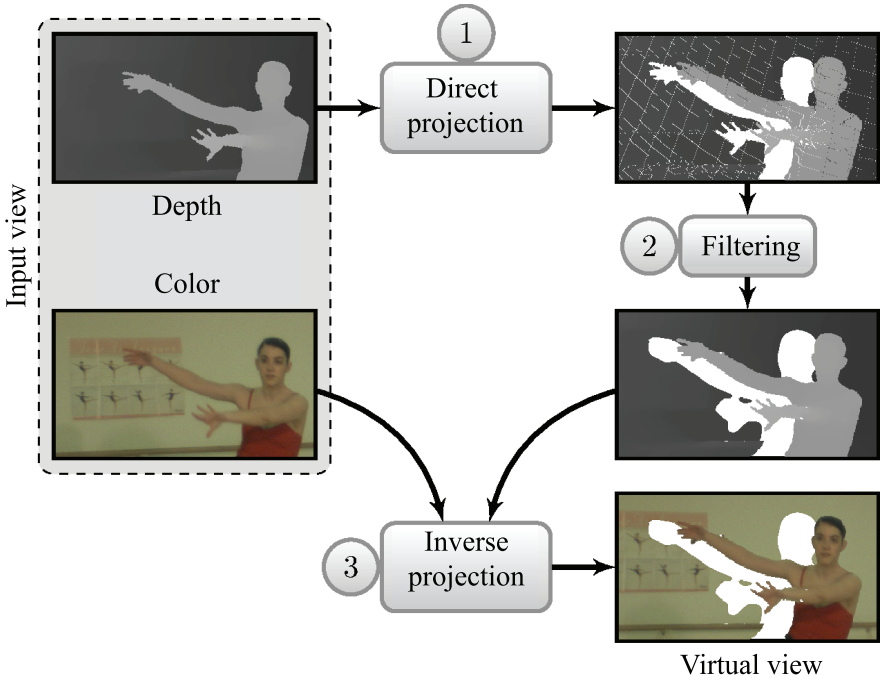


Figure 9.3. *Inverse projection*

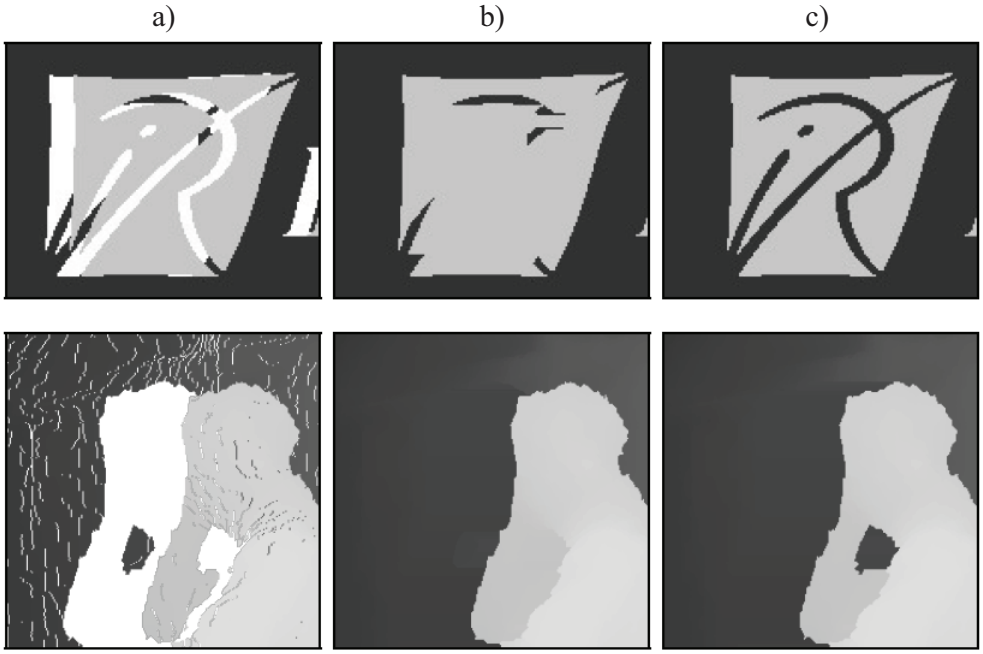


Figure 9.4. *Depth map obtained by direct projection using different methods: a) point-based basic direct projection; b) median filtering and directional "inpainting" [NGU 09]; and c) JPF with a path order [JAN 12]*

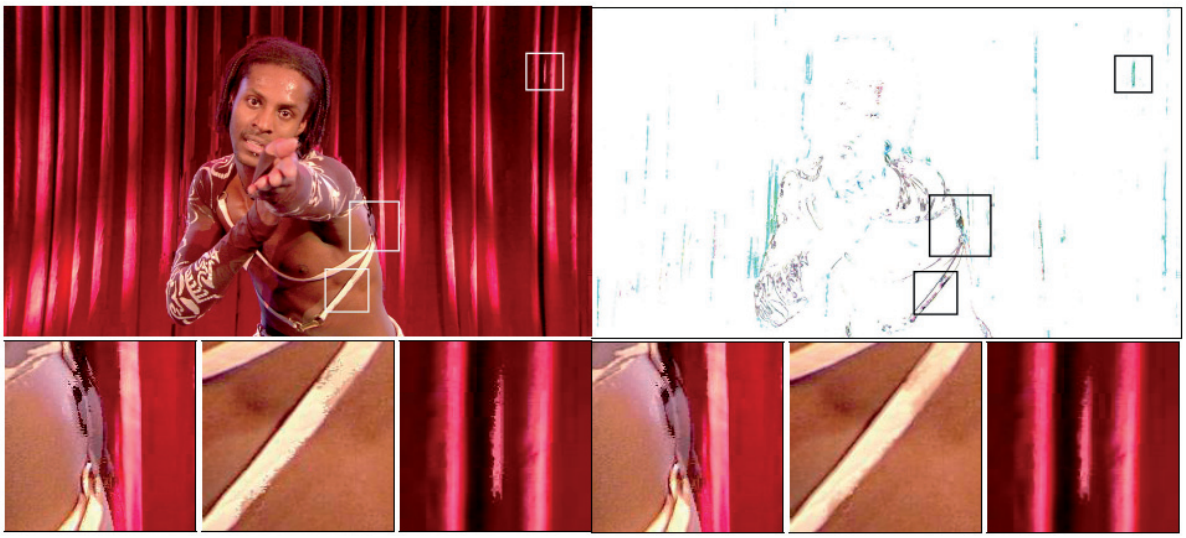


Figure 9.5. *Detecting and suppressing synthesis distortions using anisotropic smoothing [DEV 11]*

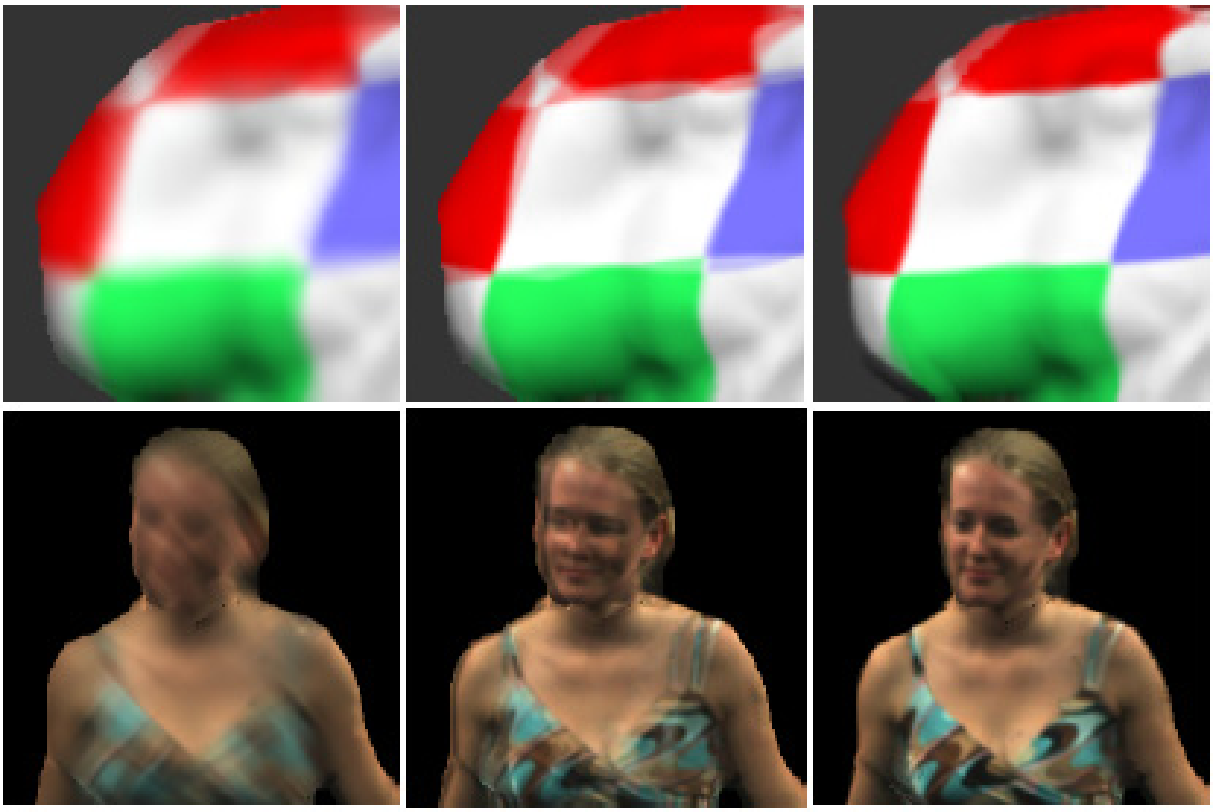


Figure 9.6. *Removing blurring when fusing textures using floating textures [EIS 08]*

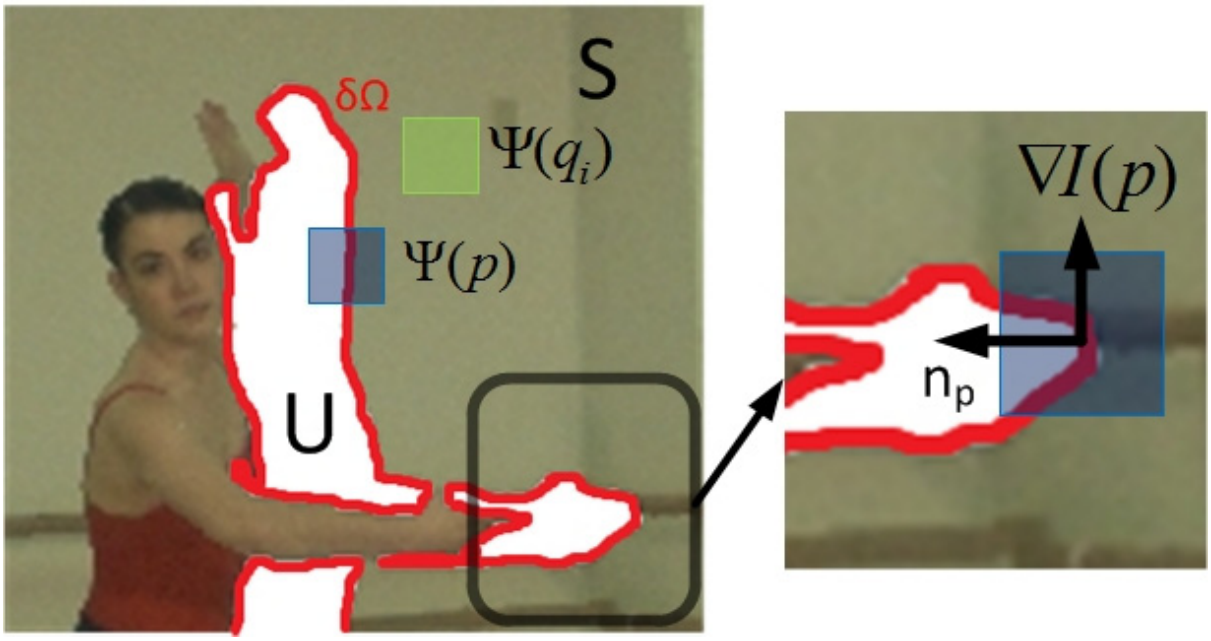


Figure 9.7. *Notations used for 2D inpainting*

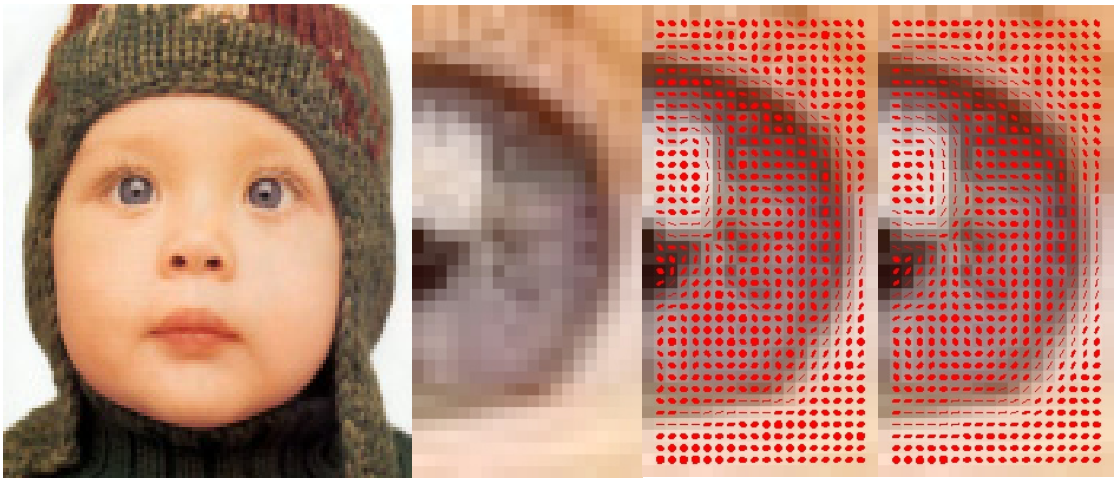
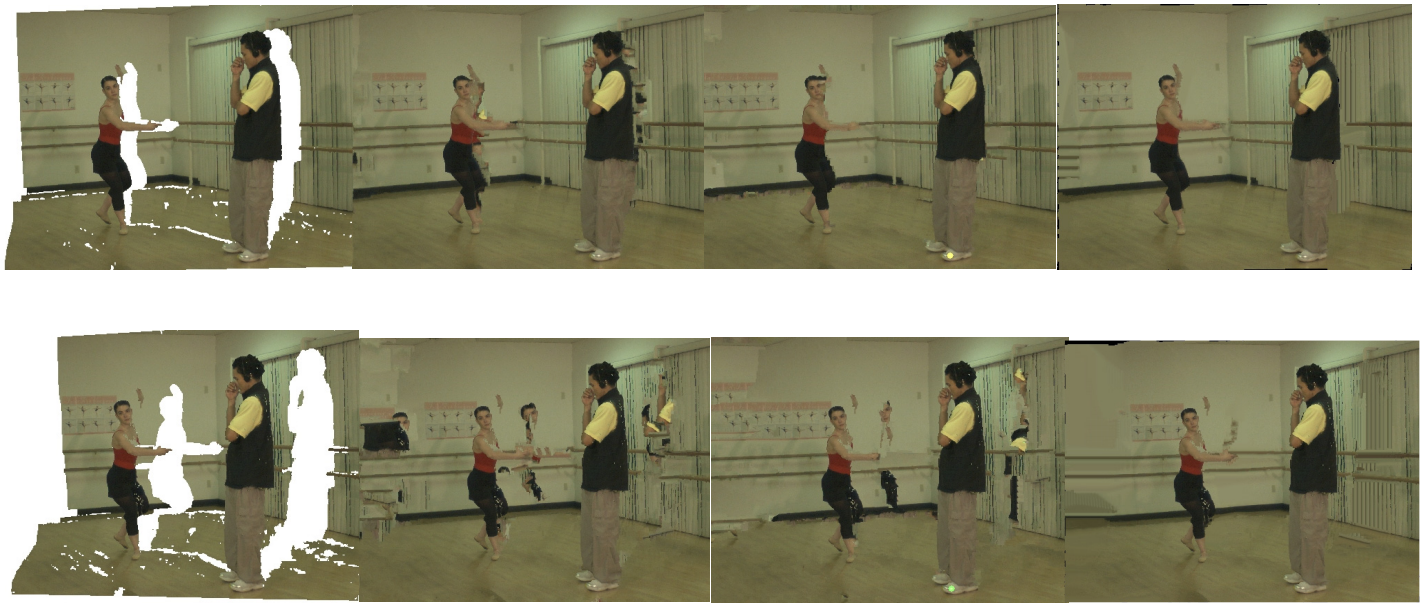


Figure 9.8. *Illustration of a structure tensor field: a) the original image; b) close-up of part of the eye; c) a non-regularized tensor field; and d) a tensor field regularized by Gaussian smoothing*



a) b) c) d)

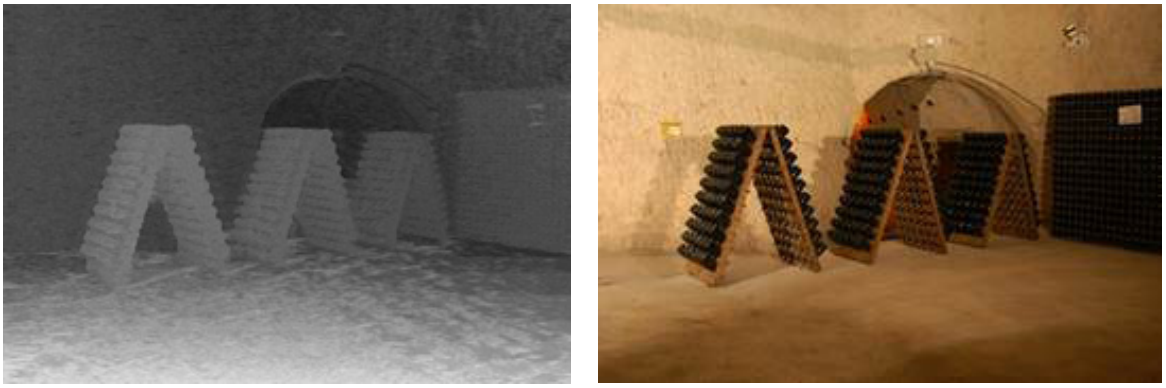
Figure 9.9. *Examples of "inpainting": a) original point of view projected in two camera referentials, uncovered zones are shown in white; b) examples of inpainting by Criminisi et al. [CRI 04] c), Daribo and Pesquet-Popescu [DAR 10], and Gautier d) et al. [GAU 11]*



a) b)

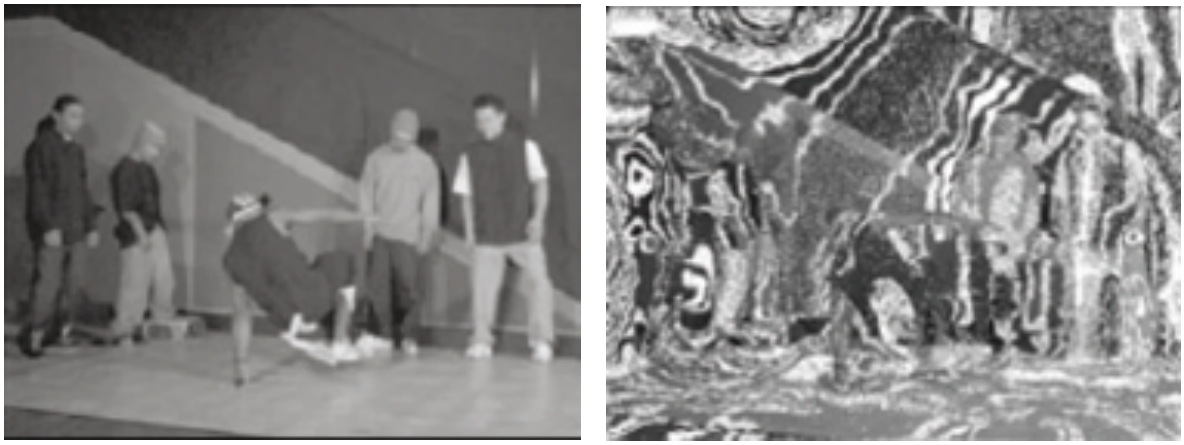
c) d)

Figure 9.10. *Examples of inpainting by [SUN 12]: a) and c) virtual viewpoints with inpainting; b) and d) result (taken from [SUN 12])*

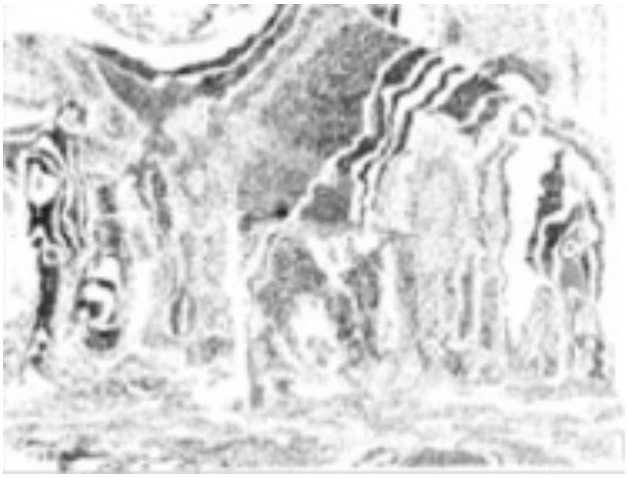


a) 2d view b) Associated depth map

Figure 10.4. *The 2D-plus-depth format*



a) b)



c)

Figure 10.10. *Three layers taken from the LDI generated for the break dancers sequence using the method proposed by [YOO 07]: a)* $\mathcal{L}$ [0], *b)* $\mathcal{L}$ [2] *and c)* $\mathcal{L}$ [4]

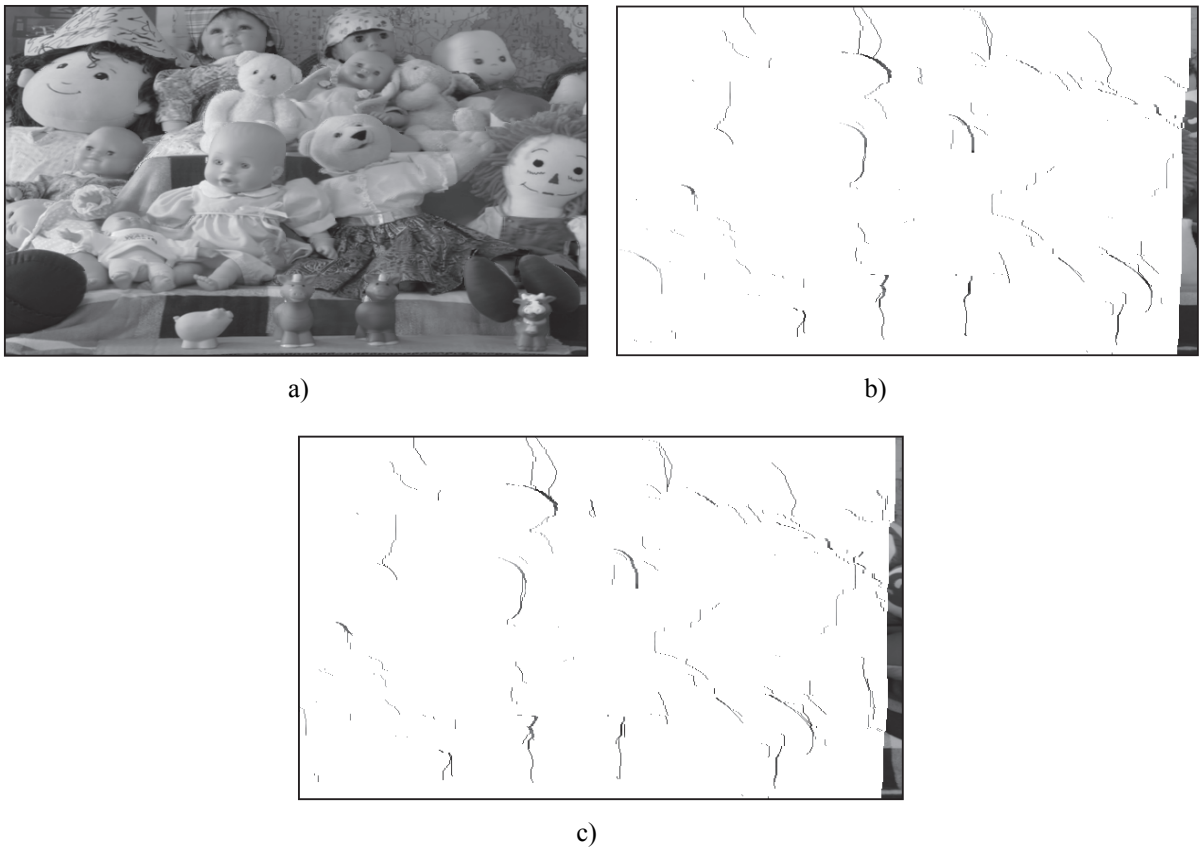


Figure 10.11. *Three layers taken from the LDI generated for the doll sequence using the method proposed by [BAT 11]: a) $\mathcal{L}$ [0], b) $\mathcal{L}$ [2] and c) $\mathcal{L}$ [4]*



Figure 10.12. *DES format (figure taken from [SMO 09]): a) LDI associated with the first reference view, and b) LDI associated with the second reference view*

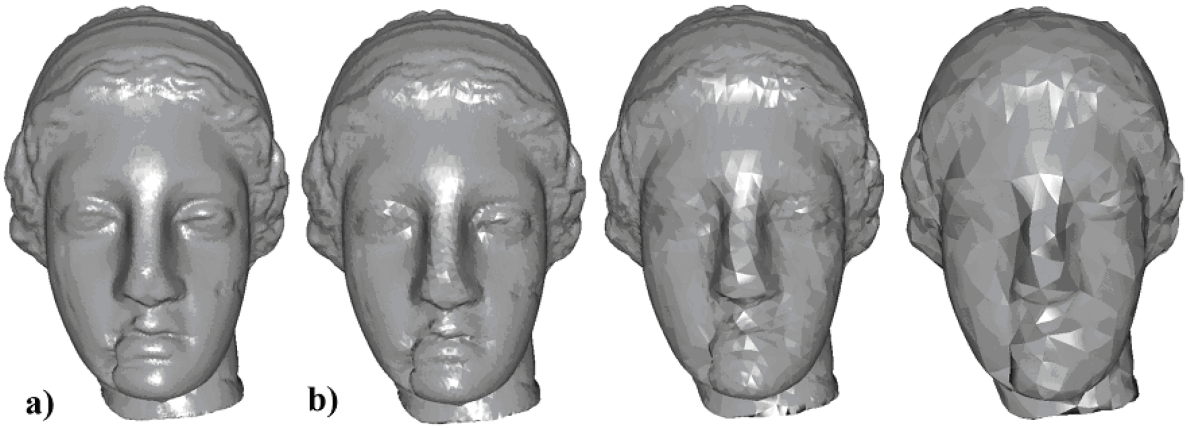


Figure 11.2. *Image of Venus: example of multiresolution meshes: a) original mesh and b) different resolutions of approximation. The information lost between two resolution levels is contained in the sub-images of wavelet coeffecients*

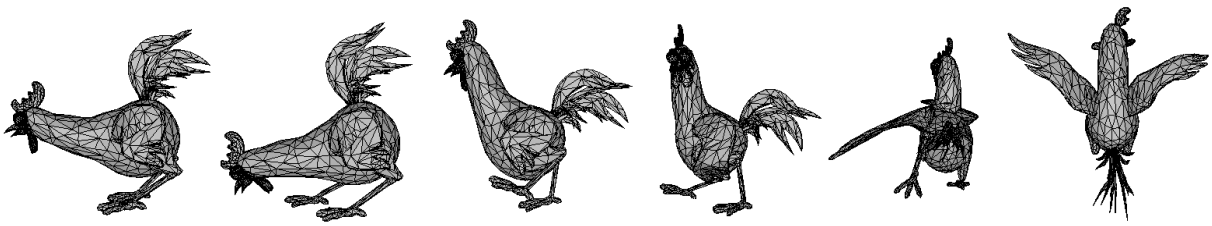


Figure 11.4. *Chicken crossing sequence (©Copyright 1996, Microsoft Corporation)*

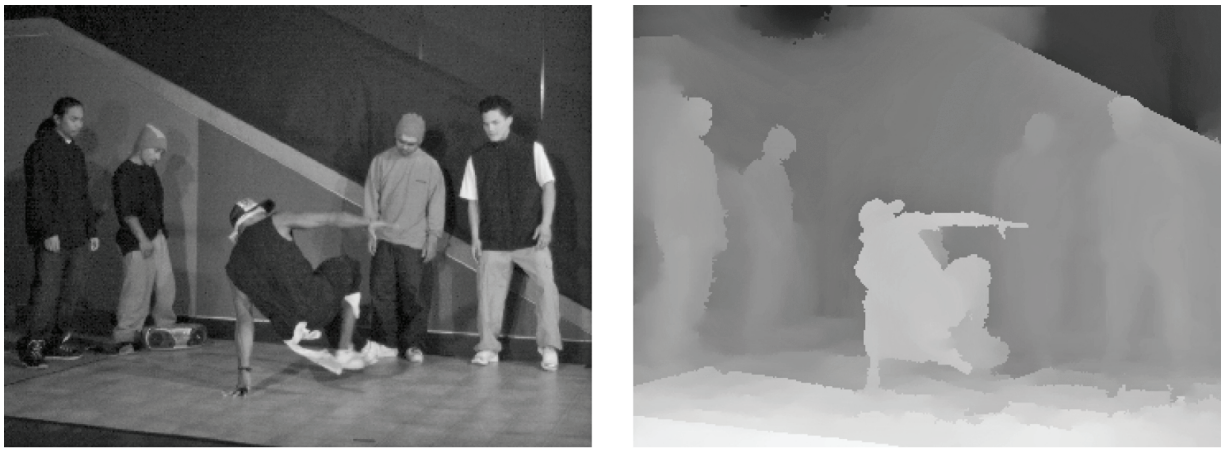


Figure 12.1. *Texture frame and associated depth map*

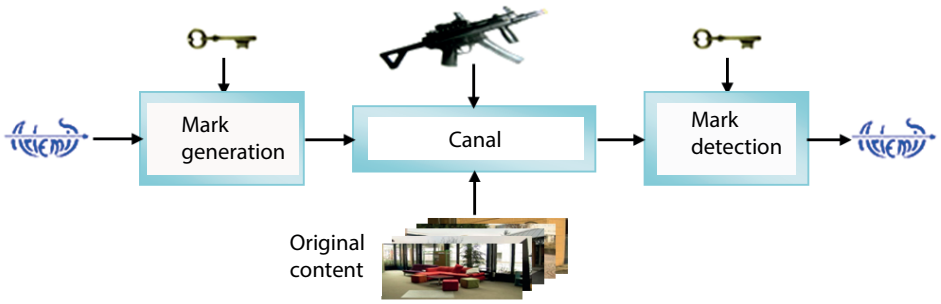


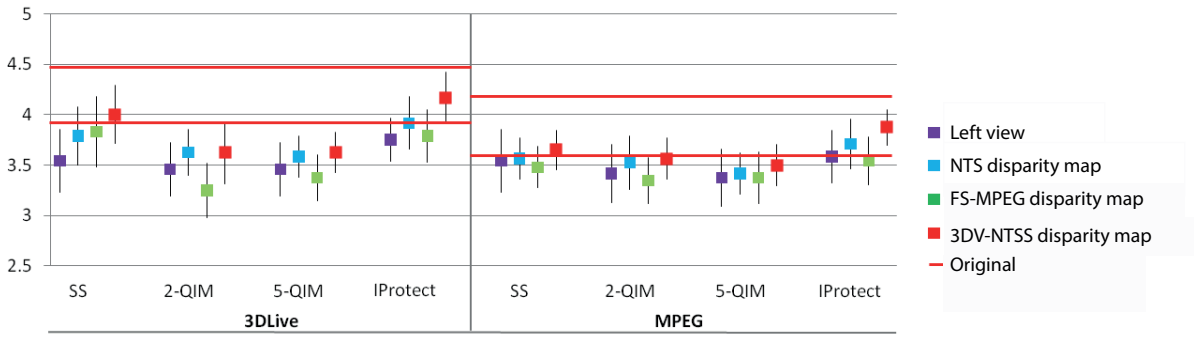
Figure 13.1. *Theoretical model of watermarking*



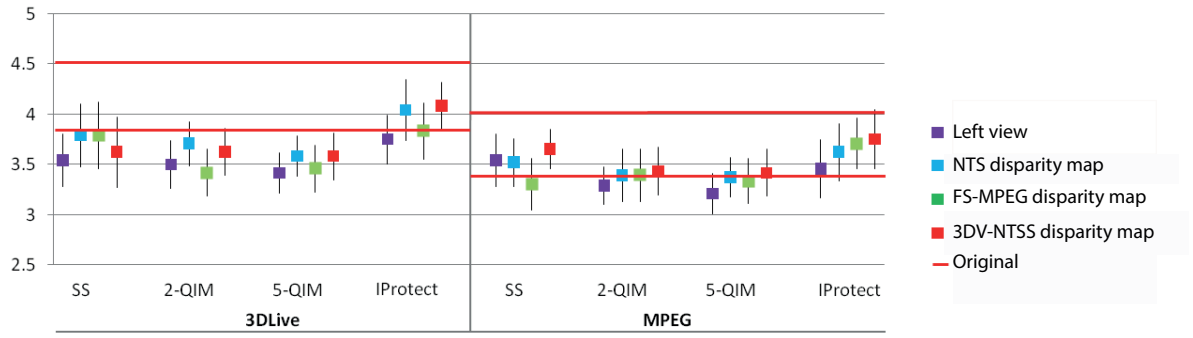
Figure 13.2. *Examples of left and right views sampled from the 3DLive corpus: rugby, volleyball, dance, theatre paly and rock concert sequences*



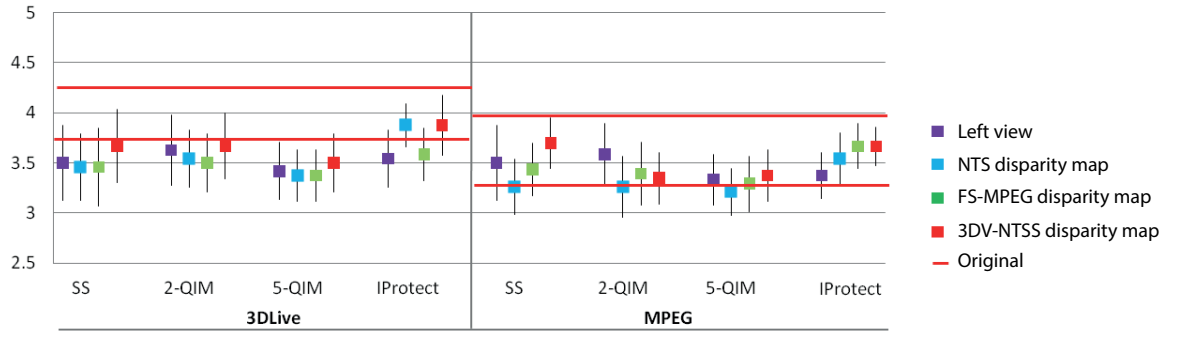
Figure 13.3. *Examples of left and right views sampled from the MPEG corpus: roller, office, city tour and cartoon sequences. The frames are represented at the same size, although their original sizes differ*



a) Subjective evaluation of image quality

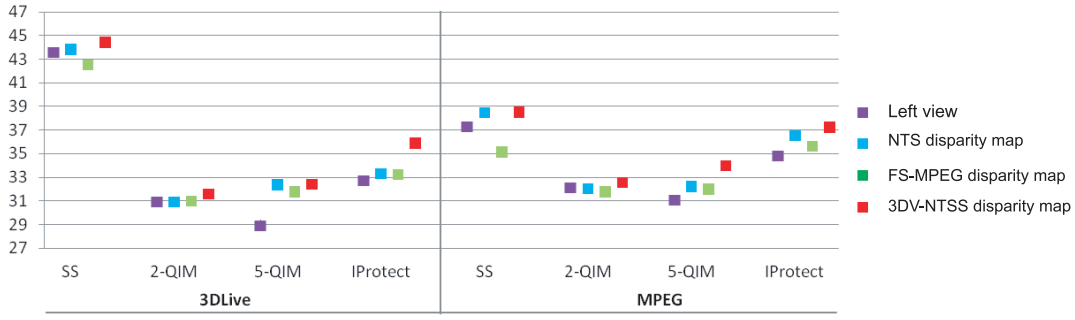


b) Subjective evaluation of depth perception

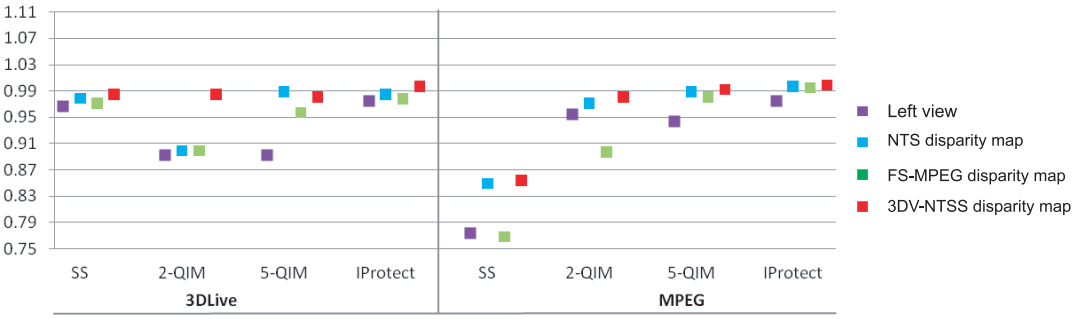


c) Subjective evaluation of visual comfort

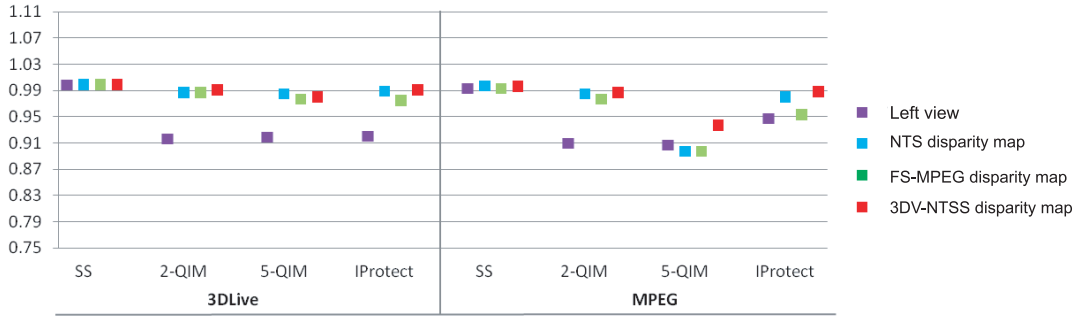
Figure 13.5. *Subjective evaluation: a) image quality, b) depth perception and c) visual comfort. The squares represent MOS values for each insertion technique/domain (from left to right: left view, NTSS, FS-MPEG and 3DV-NTSS), and the vertical lines centered on the squares show the 95% confidence intervals for watermarked content. The bold horizontal lines show the 95% confidence intervals for subjective evaluation of the original content*



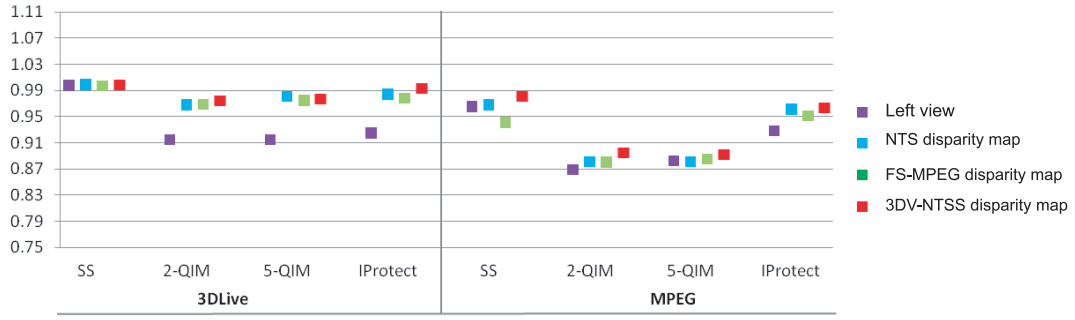
a) Mean PSNR values



b) Mean IF values

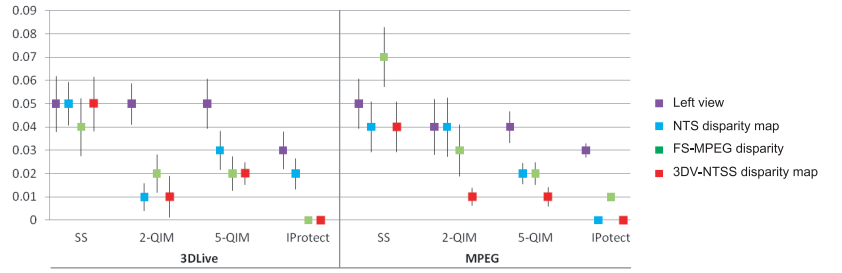


c) Mean NCC values

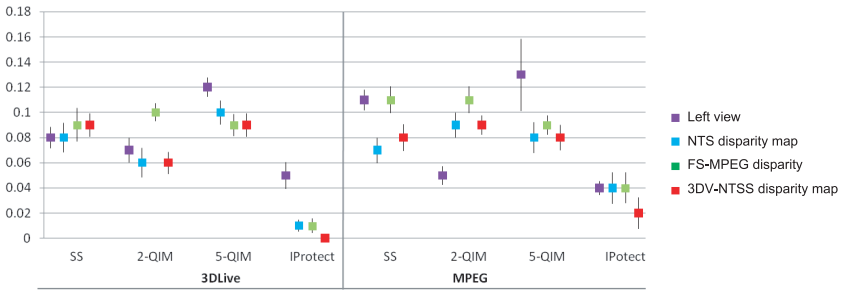


d) Mean SSIM values

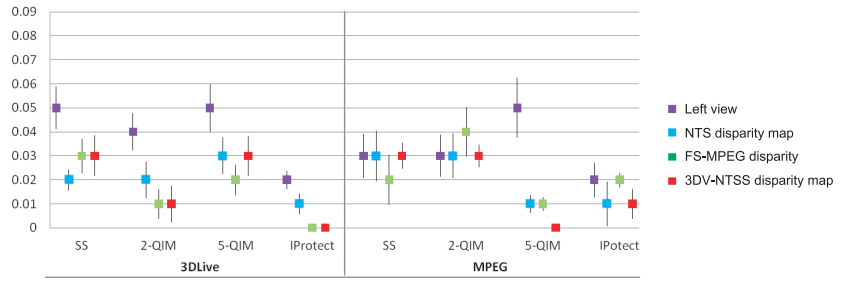
Figure 13.6. *Objective evaluation of visual quality of watermarked content using different objective quality metrics*



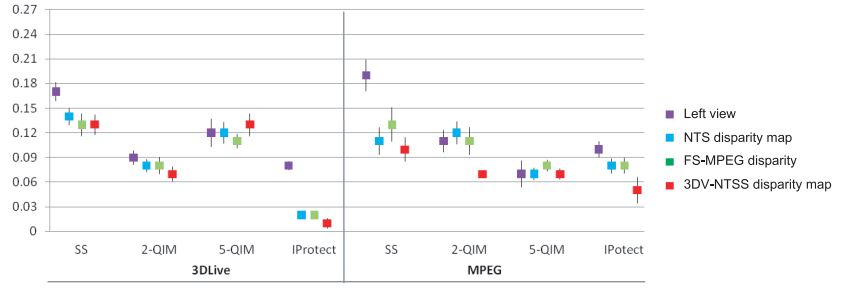
a) Attack by Gaussian filtering



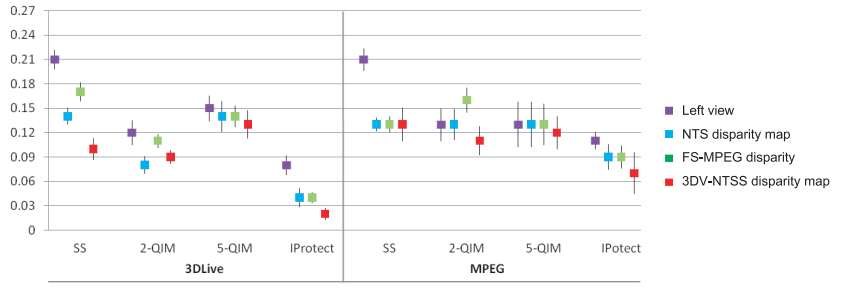
b) Attack by contour enhancement



c) Attack by HJPEG compression



d) Attack by small rotations



e) StirMark random bending attack

Figure 13.7. *Robustness of watermarking methods to different attacks: mean BER values and associated 95% confidence intervals*

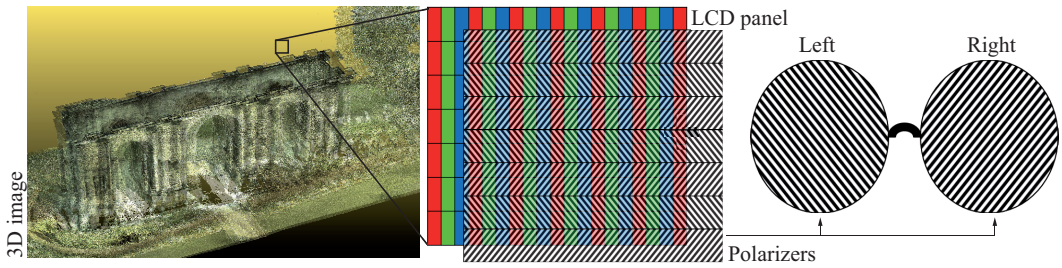


Figure 14.2. *Light polarization principle*

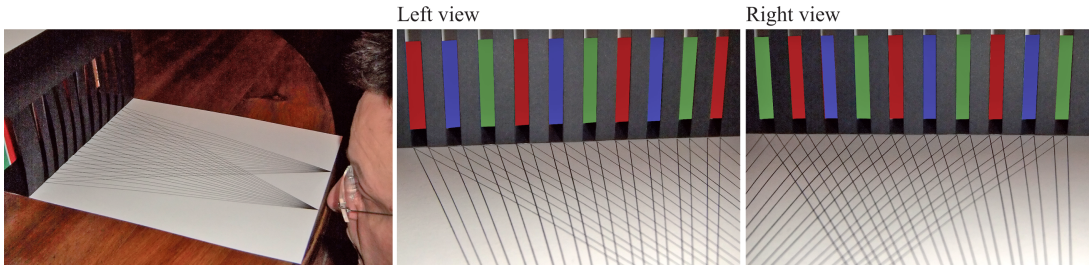


Figure 14.3. *Operating principle of a parallax barrier*

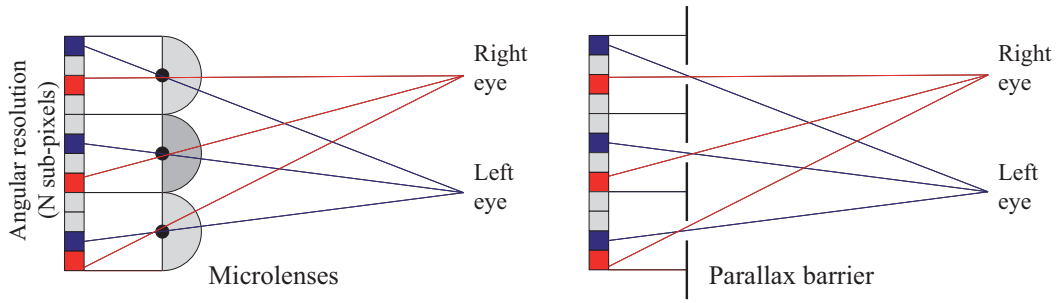


Figure 14.4. *Optical filters*

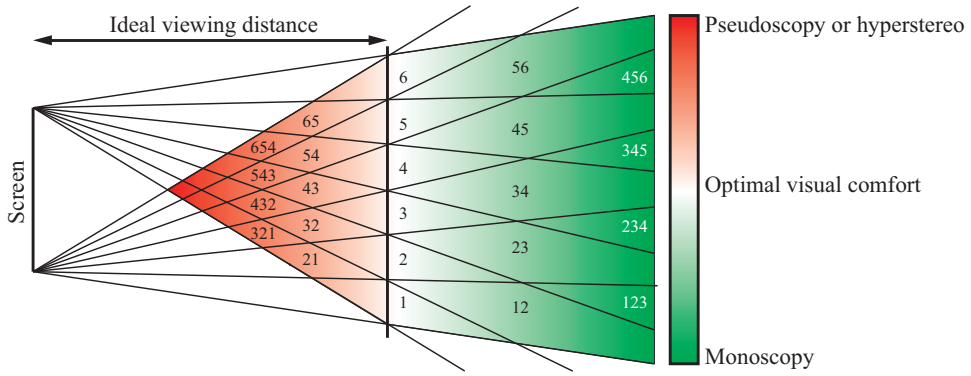


Figure 14.5. *Observation windows produced by different optical beams. The optimal comfort zones for good 3D vision (stereopsy) correspond to zones 1 – 6*

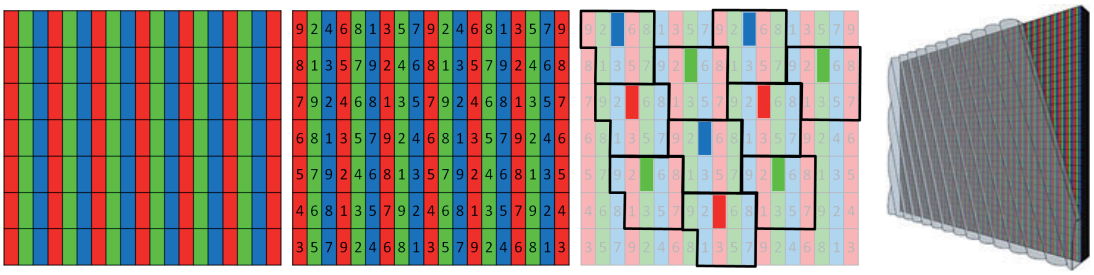


Figure 14.6. *Multiview representation of a pixel (nine views)*

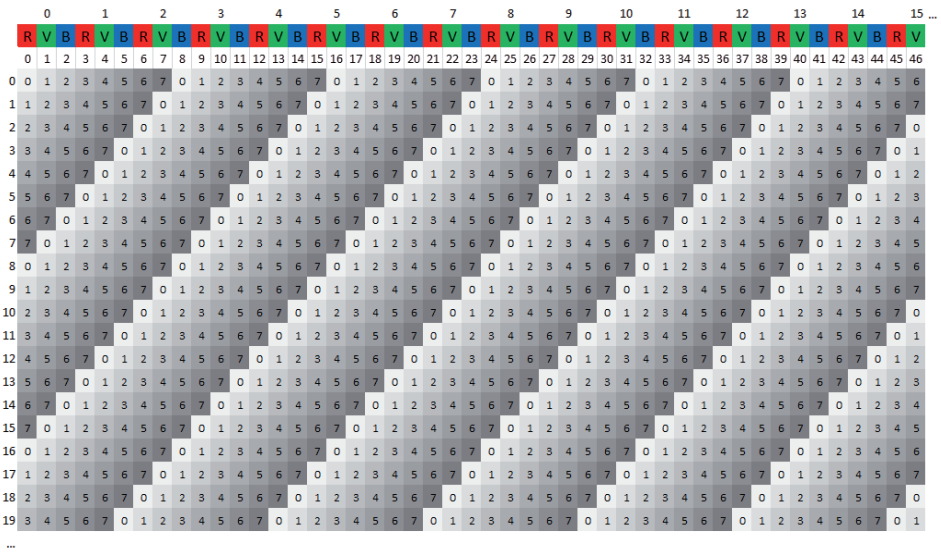


Figure 14.7. *Mixing filters for 4D-view screens with eight views* ($i \in [0.46]$, $j \in [0.19]$, $n = 8$, $q_a = 1$, $q_b = 1$, $q_x = 1$, $q_y = 1$)

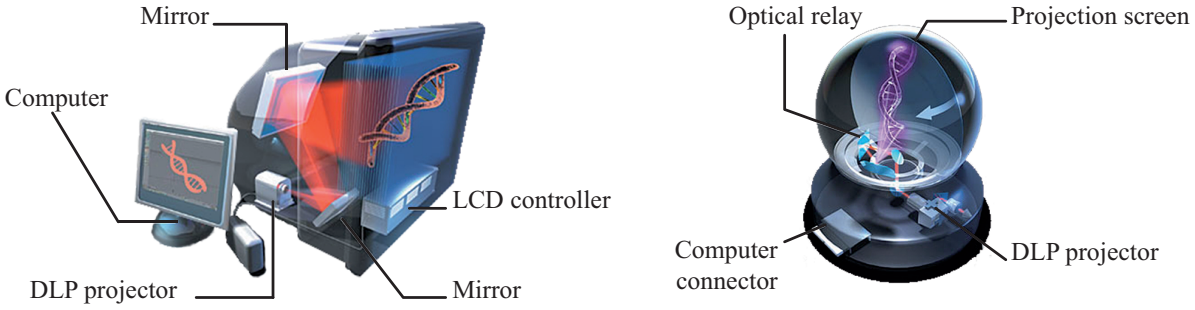


Figure 14.8. *DepthCube and Perspecta volumetric screens, Actuality Systems Inc. [FAV 02]*

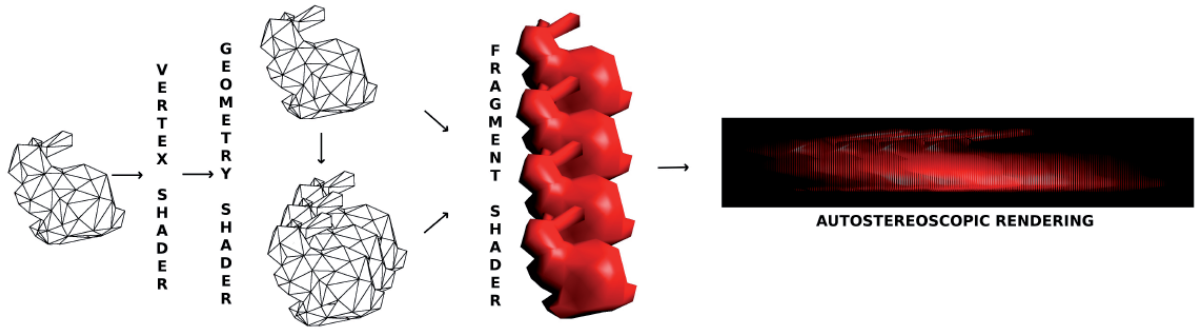


Figure 14.9. *Overview of our multiview stereoscopic GPU rendering method*

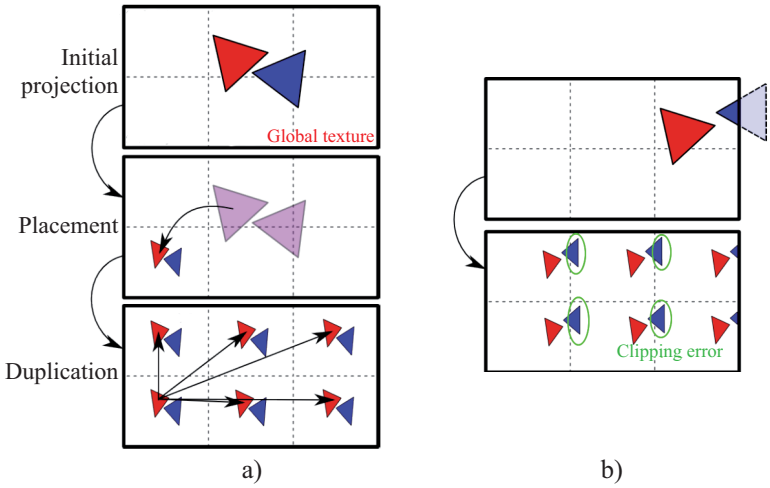


Figure 14.10. *a) Use of a texture for multiview rendering; b) clipping issue for this technique*

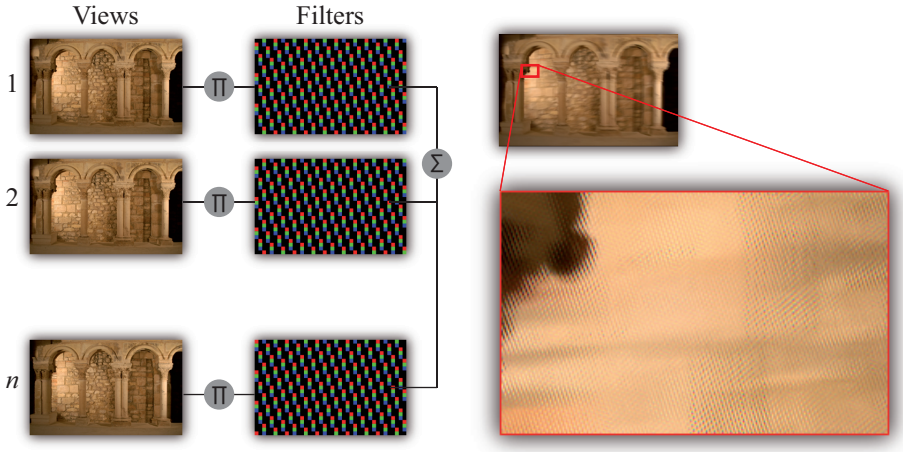


Figure 14.11. *View interleaving principle with mixing filters*

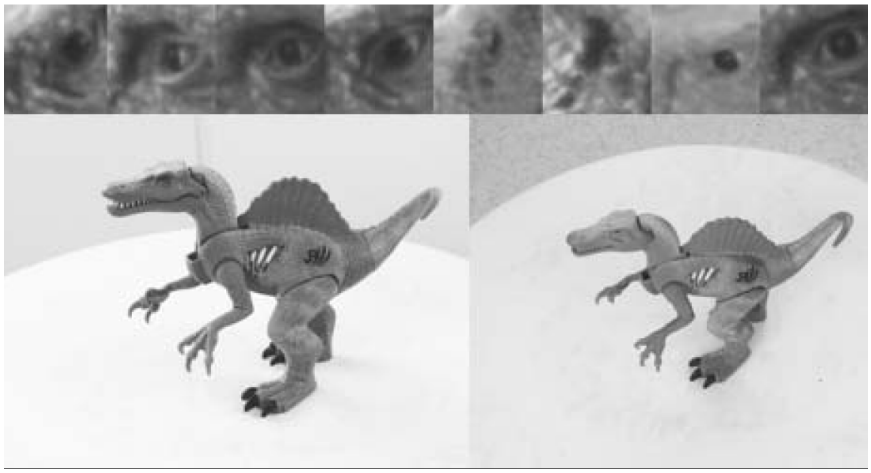


Figure 15.1. *Representation of a point by a collection of its aspects: the eye of the dinosaur is represented by the various appearances which it may take in a set of learning images*

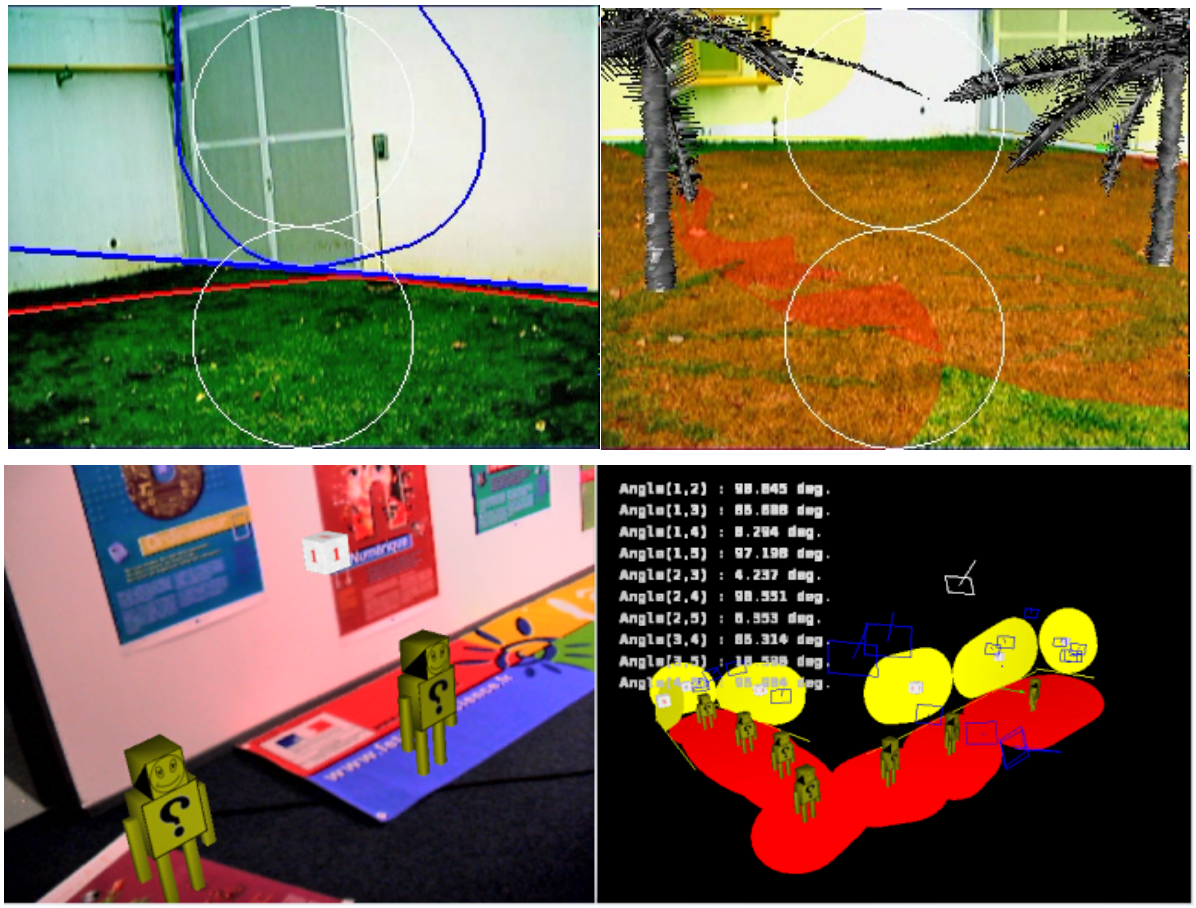
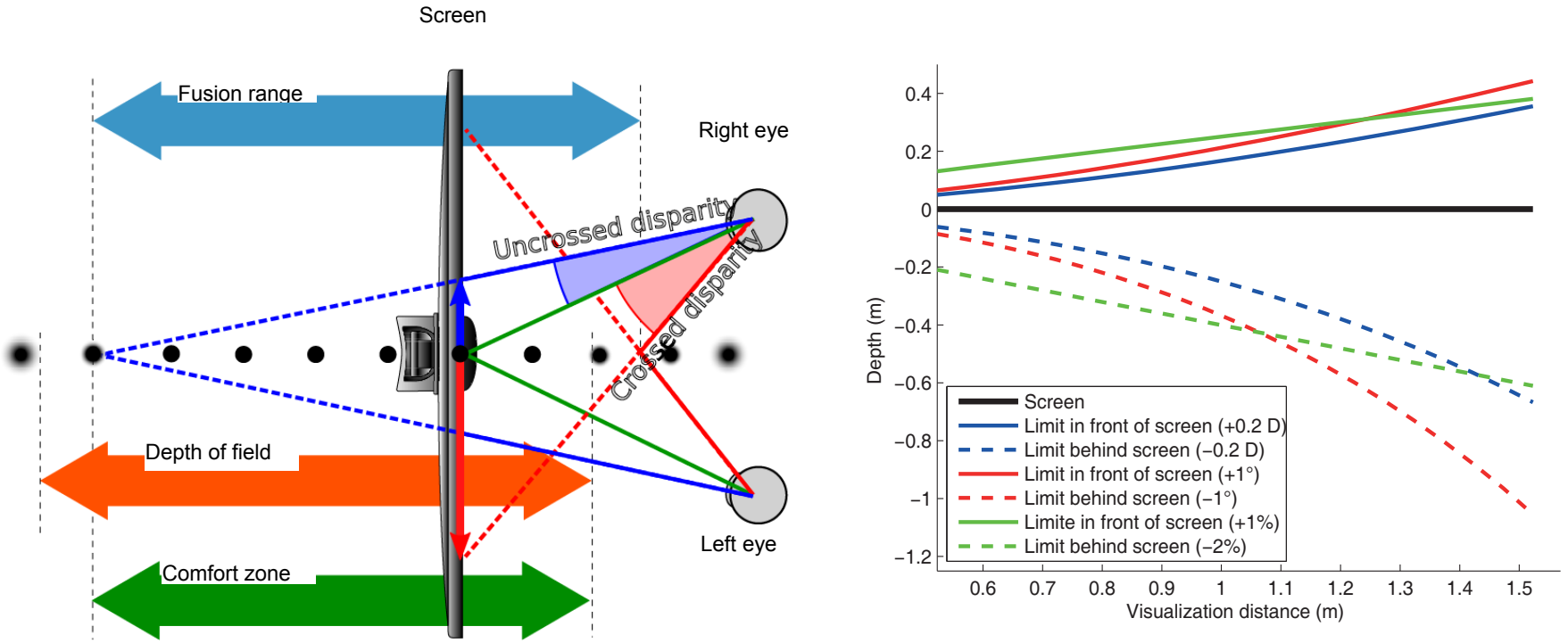


Figure 15.4. *In situ environment modeling using a method developed by the MAGRIT team. a) The walls and floor are targeted and captured through the camera, then reconstructed in real time using simple interactions. After a few seconds, a virtual scene (of simple palm trees, in this case) may be integrated into the modeled environment. A video showing these operations is available at www.loria.fr/~gsimon/vc/blobs_outdoor.avi. b) Example of blobs obtained in the corridors at the Loria. These blobs can be used to easily position virtual objects in relation to the real scene*



a) Depth of field, fusion and comfort zone

b) Limits of the comfort zone for a 42” screen with a 16/9 display ratio

Figure 16.5. *The comfort zone a) and its definitions b)*

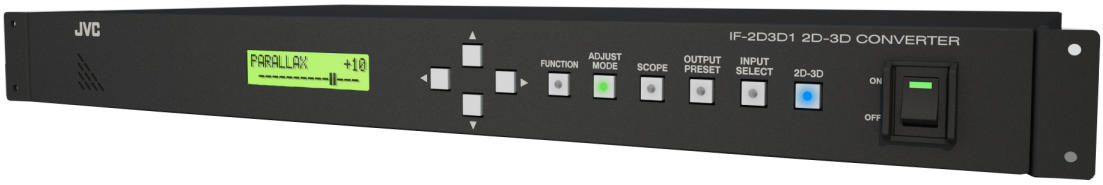


Figure 17.1. *The professional JVC IF-2D3D1 package performs automatic 2D–3D conversion in addition to other stereoscopic processes*



a)



b) c)

Figure 17.2. *a) Starting image of a depth map conversion; b) depth map with two segmented elements; c) the same depth map after depth matching between two points at the same distance from the two elements*

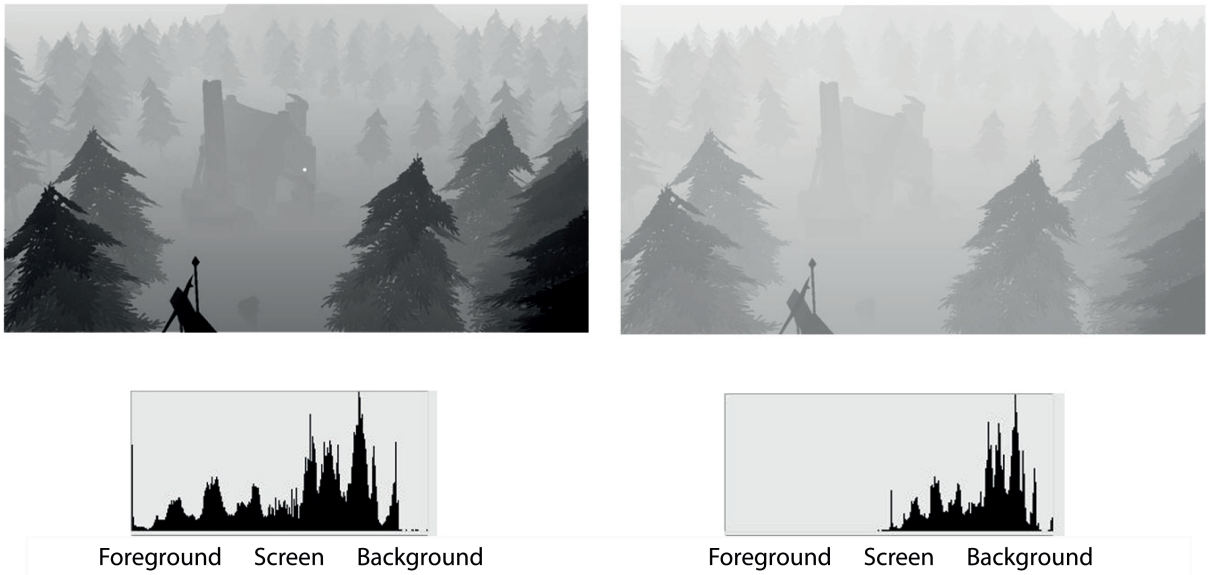


Figure 17.3. *Left: a depth map using the whole depth budget. Right: a depth map where the whole scene has been relegated behind the screen. The histogram of the depth map is shown under each image*

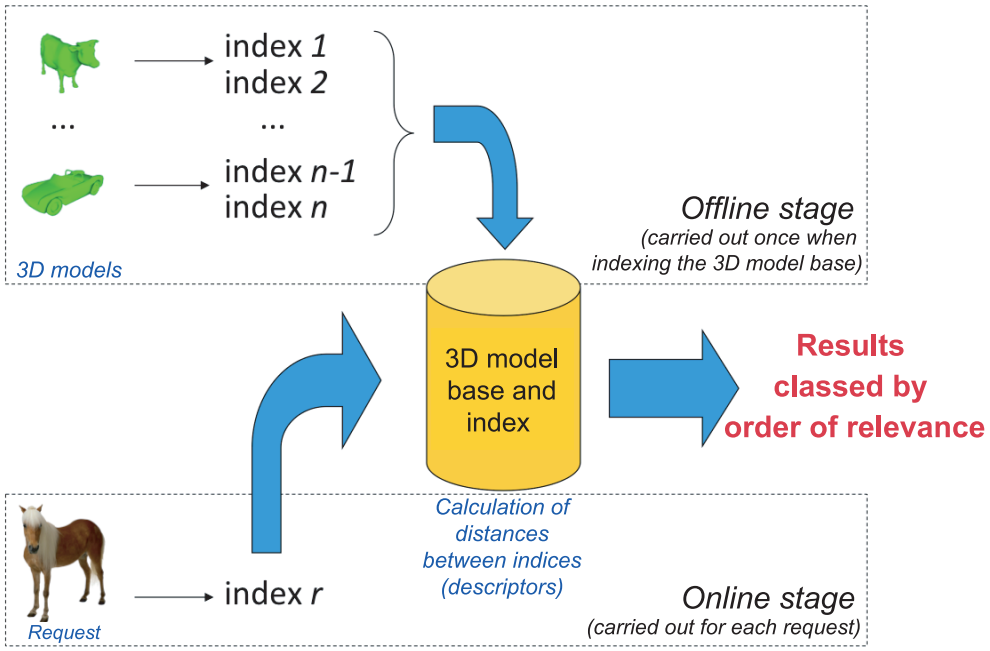


Figure 18.1. *The two stages of 3D model indexing. The offline stage is carried out once, whereas the online stage is carried out for each request*

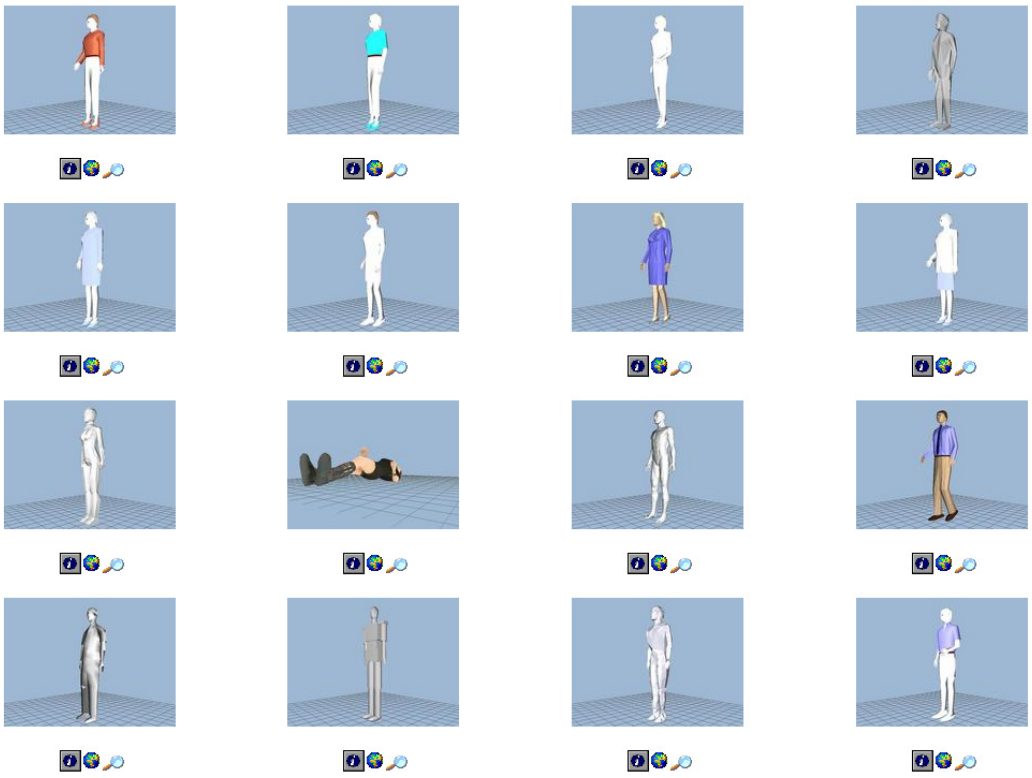


Figure 18.2. *Example of results obtained using the 2D view oriented approach developed by the MIIRE team [FIL 07]*

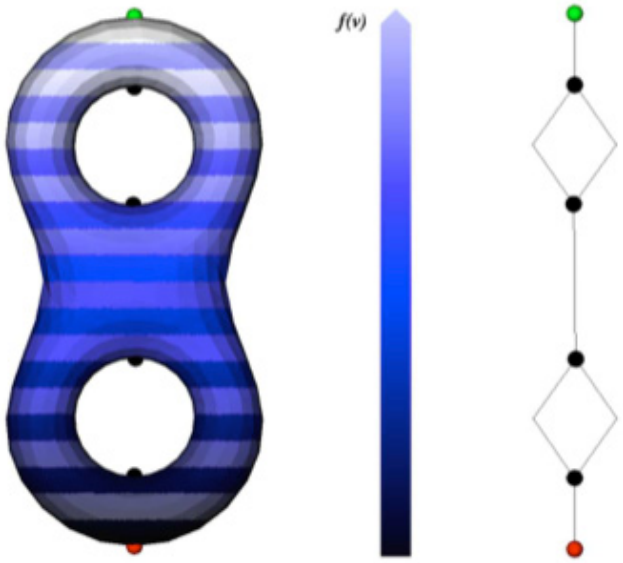


Figure 18.3. *Example of a Reeb graph obtained by the Morse height function*

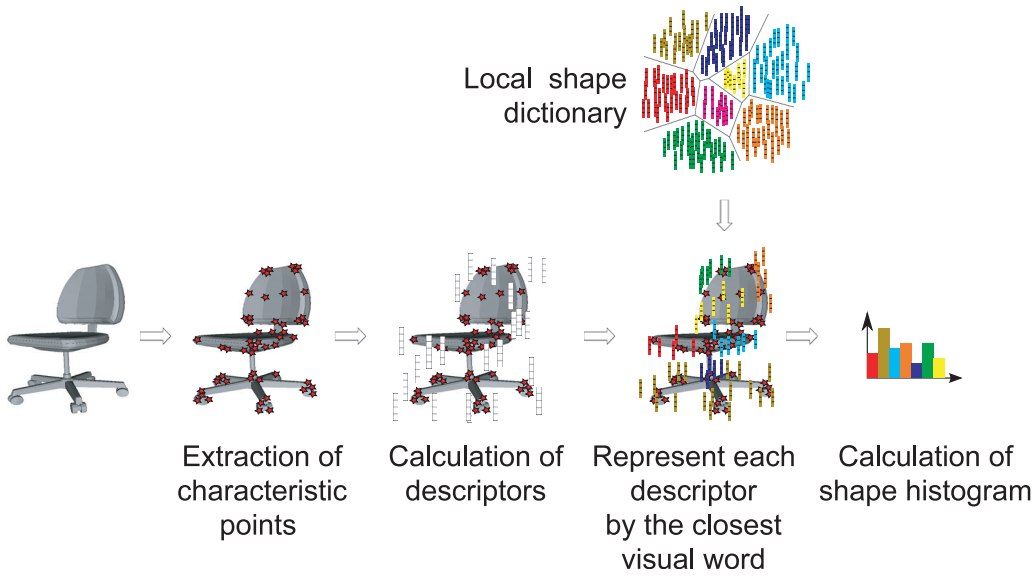


Figure 18.4. *Bag-of-words approach. A set of characteristic points is first extracted from the 3D object. Next, a descriptor is calculated for each characterized point; each descriptor is then represented by the closest visual word in the shape dictionary. Finally, a histogram is calculated showing the number of occurrences of each visual word*



Figure 18.5. *Examples of 3D meshes extracted from a 3D video sequence*

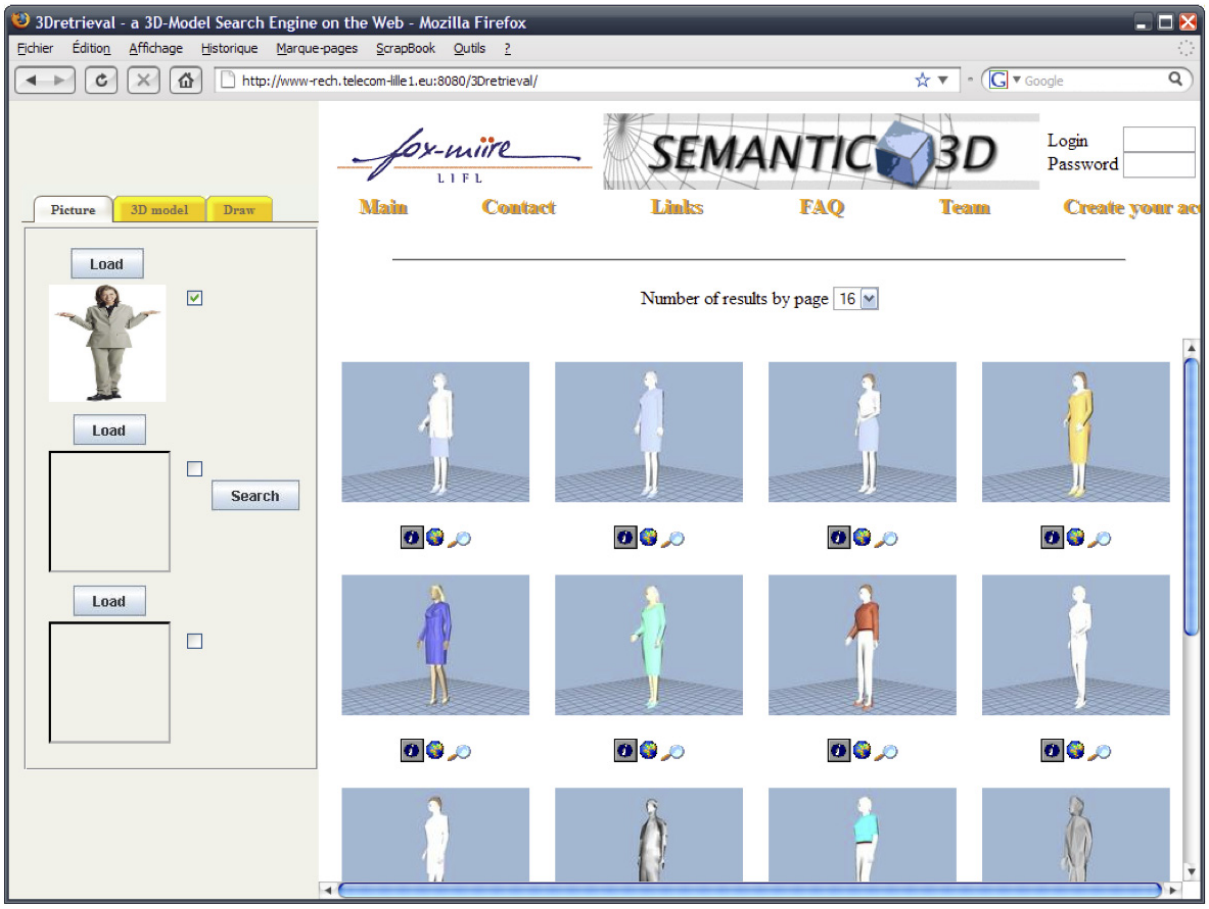


Figure 18.6. *Web interface of the MIIRE team's 3D search engine [FIL 07]*

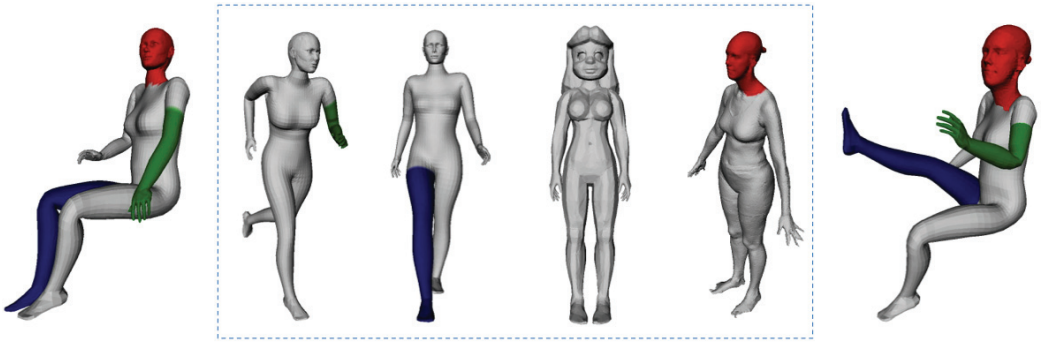


Figure 18.7. *Example modeling. a) Initial request; b) results produced by the search engine; and c) use of parts of results to create a new 3D object*


a) Underexposed

b) Intermediate


c) Overexposed

Figure 19.3. *Images of different exposures acquired using the autobracketing function on a photographic camera*

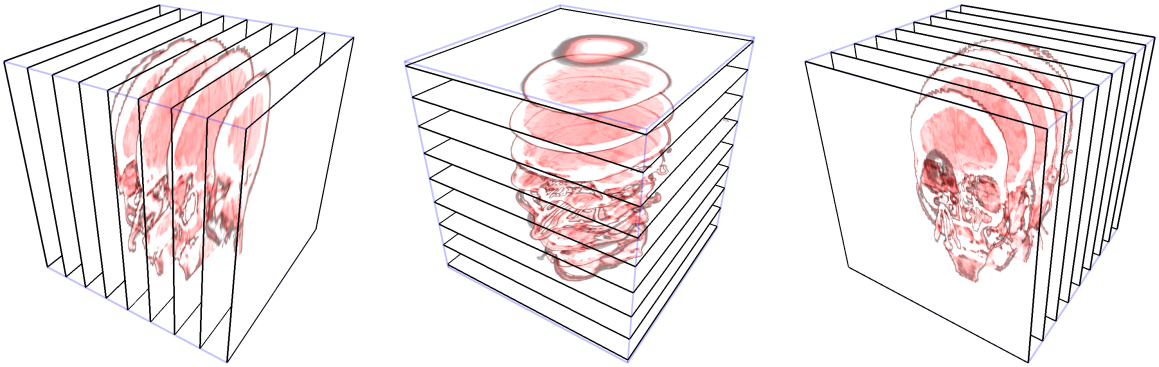


Figure 20.1. *Axis-aligned slices requiring several 2D textures (or one 3D texture)*

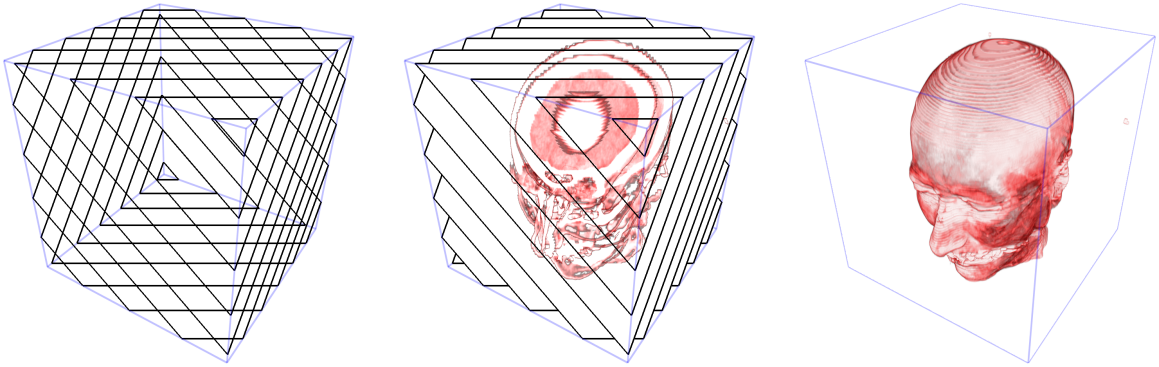
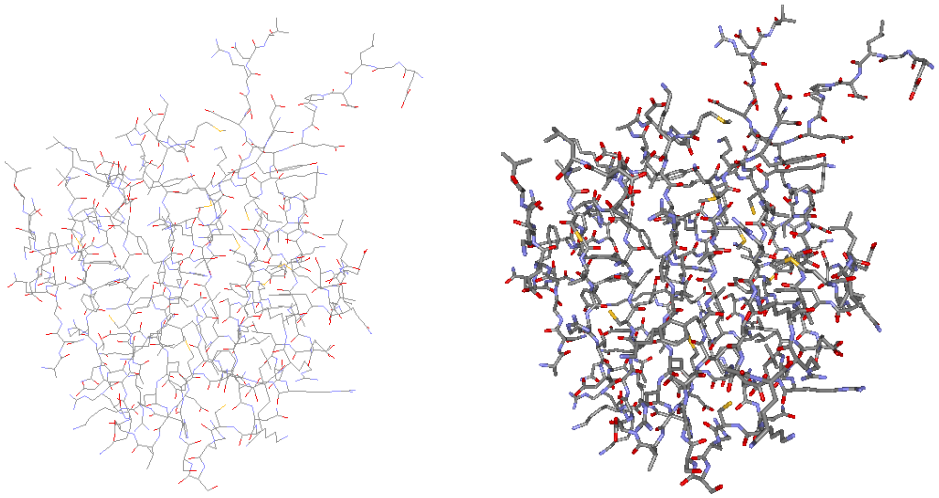
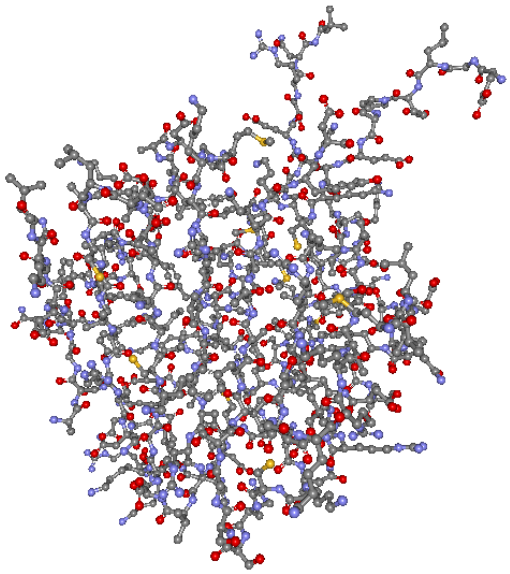


Figure 20.2. *View-aligned slicing requiring a single 3D texture*



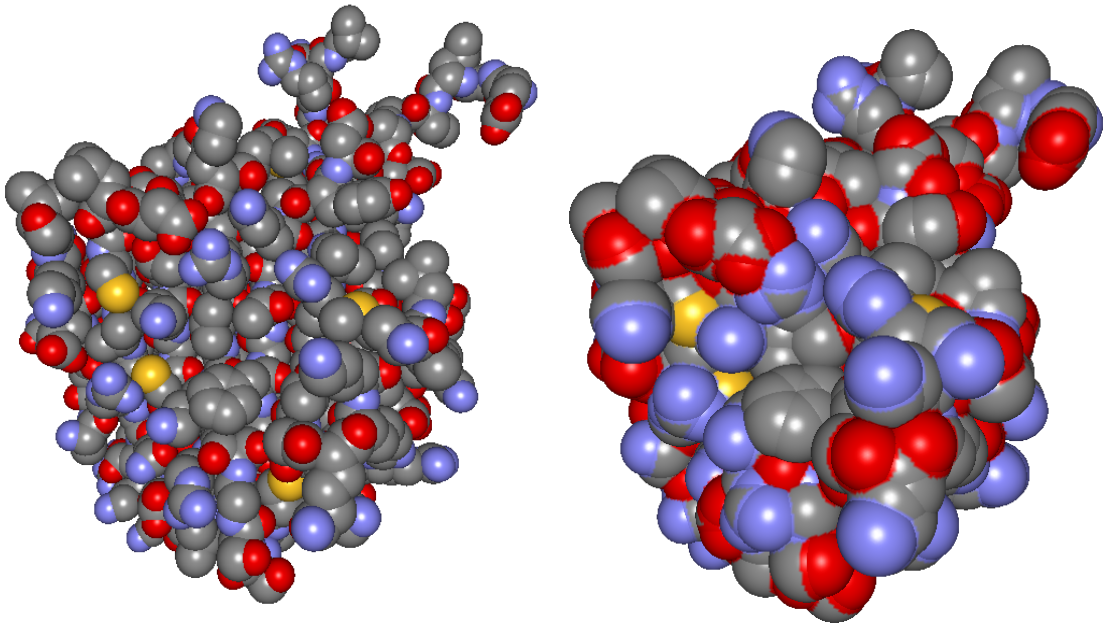
a) Wireframe b) Licorice

Figure 20.3. *Basic visualization modes*



c) Ball and stick model

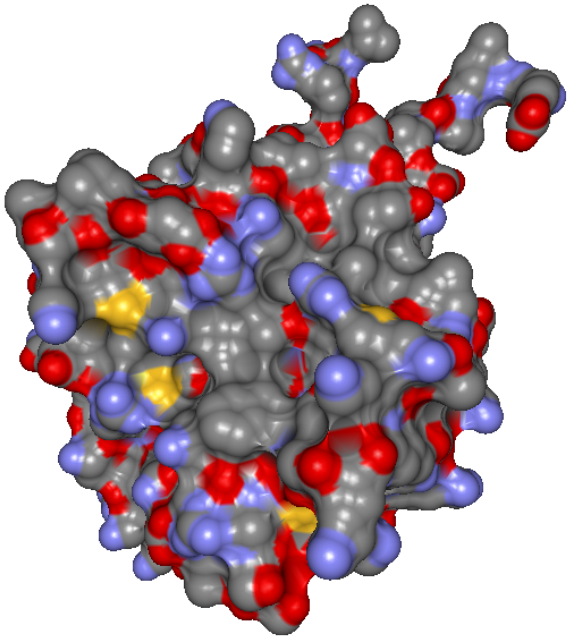
Figure 20.3. *(continued). Basic visualization modes*



a) Van der Waals surface

b) Solvent accessible surface

Figure 20.4. *Surfacic visualization modes*



c) Solvent excluded surface

Figure 20.4. *(continued).Surfacic visualization modes*

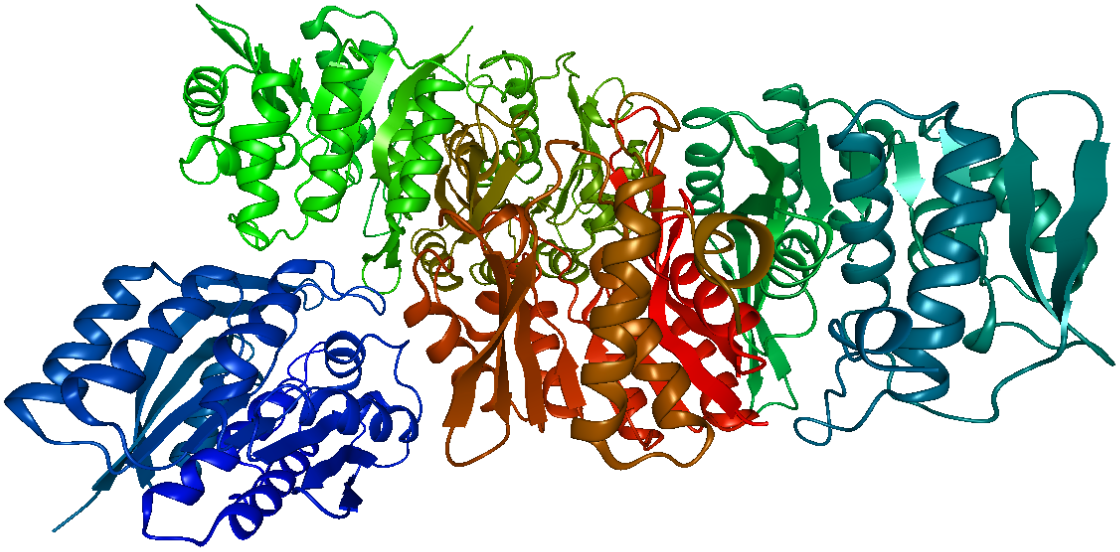


Figure 20.5. *Cartoon representation*

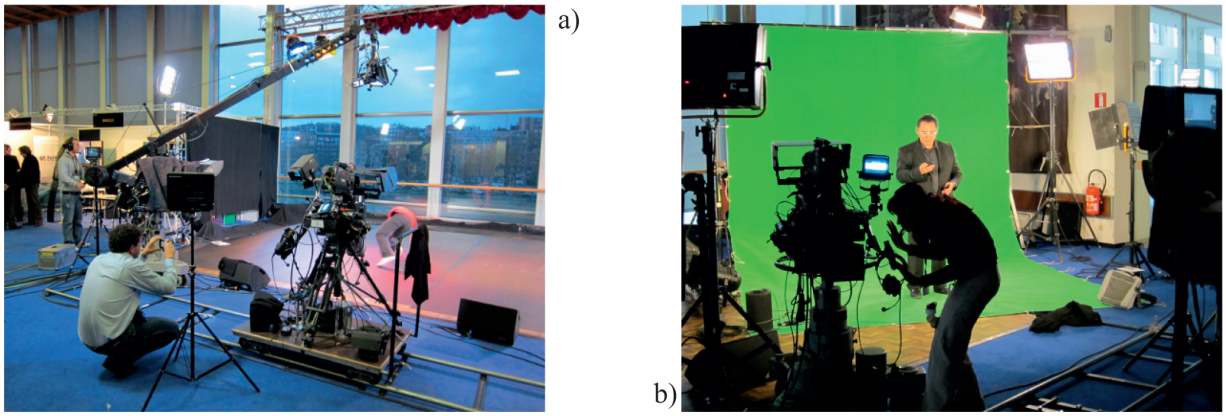


Figure 22.1. *Capture of a show in live S-3D in a hall of the Palais des Congrès of Liège. a) The artists evolved on a special surface laid onto the carpeting. To the left, one sees exhibition stands. b) The green key used to insert a real person into a virtual set. The wooden floor and the tracks show the relation to the previous image*

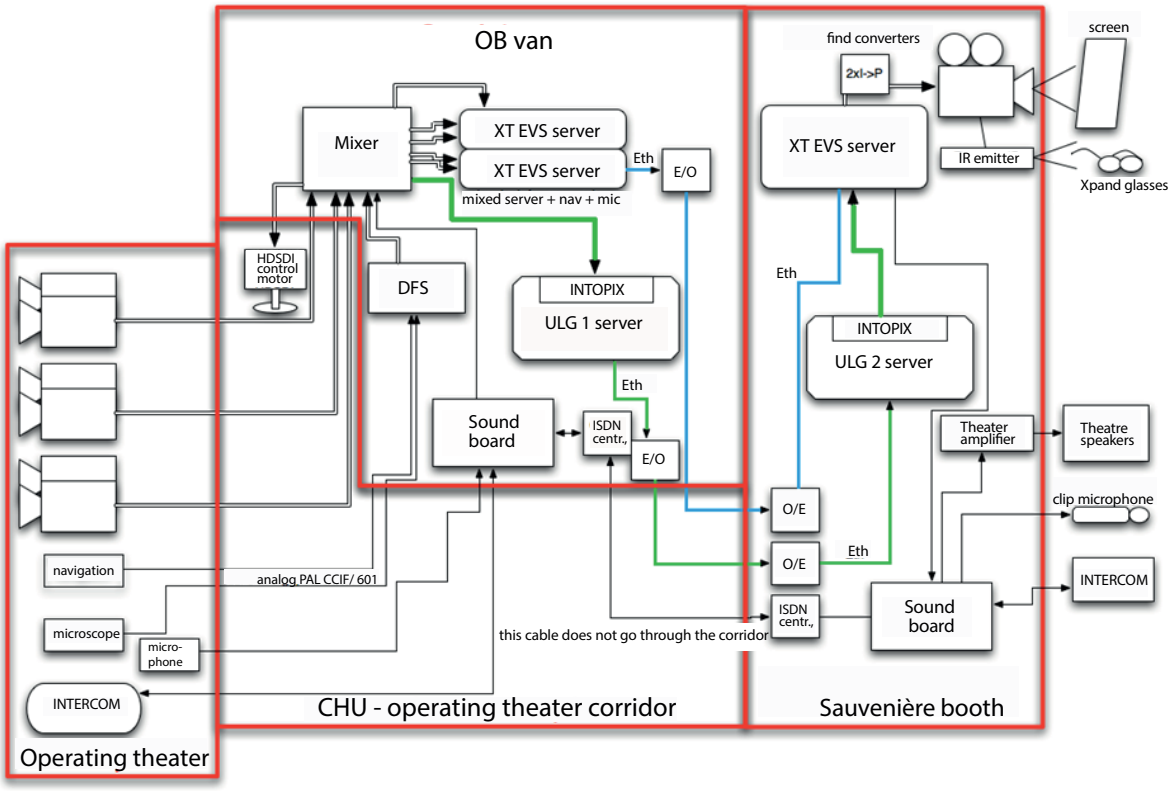


Figure 22.2. *Block diagram of the full retransmission chain, including capture, transmission, and projection*

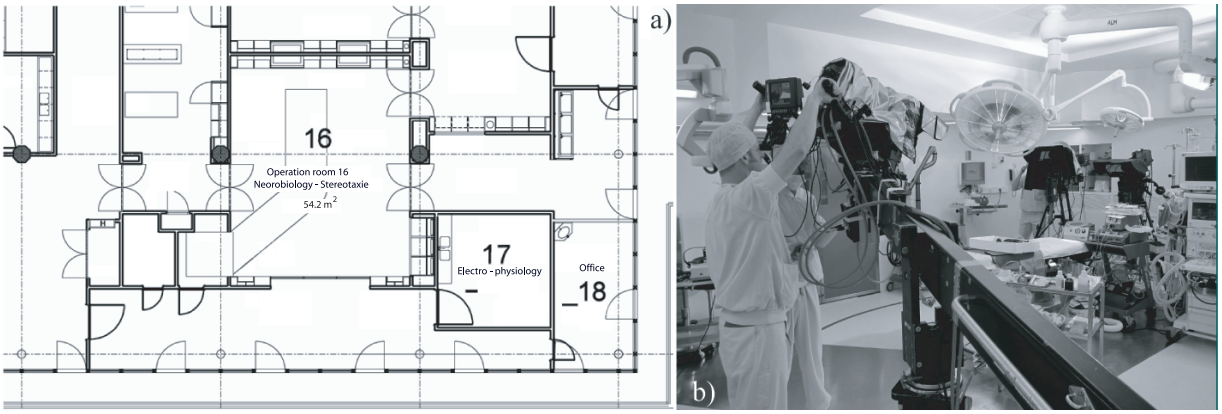


Figure 22.3. *a) Plan of part of the operating room of interest at the CHU of Liège, with, at bottom right, the balcony and its wall that various cables were drawn over. b) Arrangement of the three S-3D rigs around the operating table. The capture operators wore standard operating room clothing*

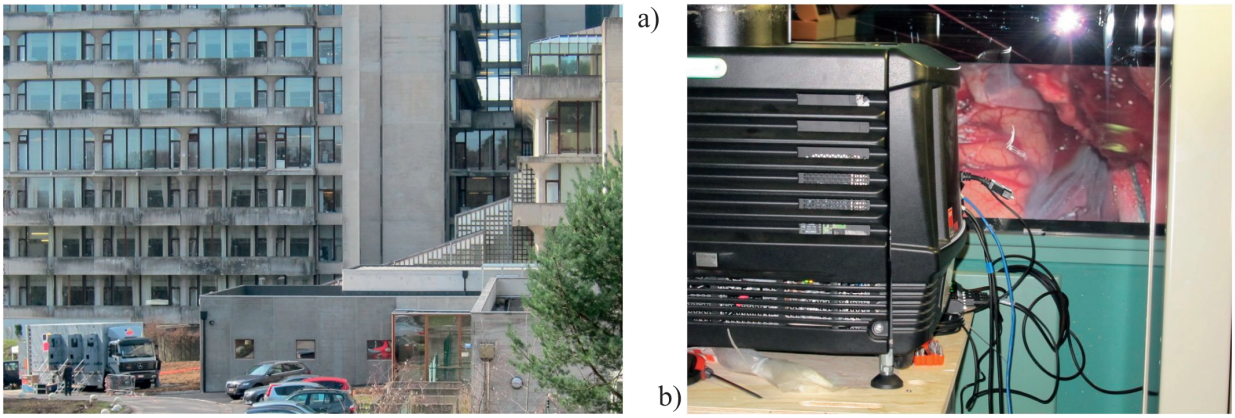


Figure 22.4. *a) OB van in the parking at the base of the building (far right) where the operating room was located (on the top floor). b) View from the projection booth at the Cinéma Sauvenière, showing the DP2000 projector that we had installed for the occasion and a partial view of an image projected onto the screen (showing the operating field), through the fireproof window of the booth*

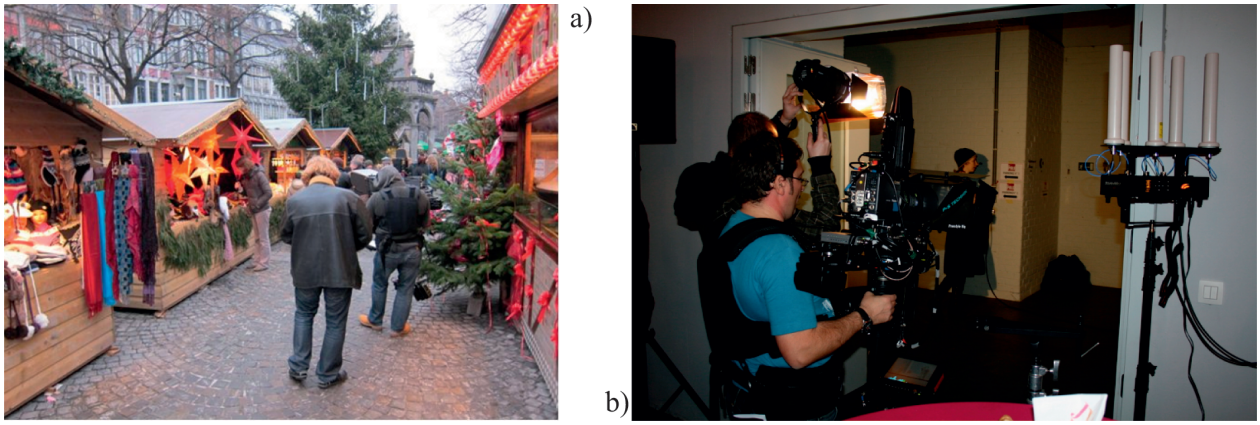


Figure 22.5. *a) First steadicam capture location in Liège: the "Marché de Noël". b) Third steadicam capture location in Liège: the Pôle Image de Liège (PIL) (courtesy of C. Iland from the PIL)*

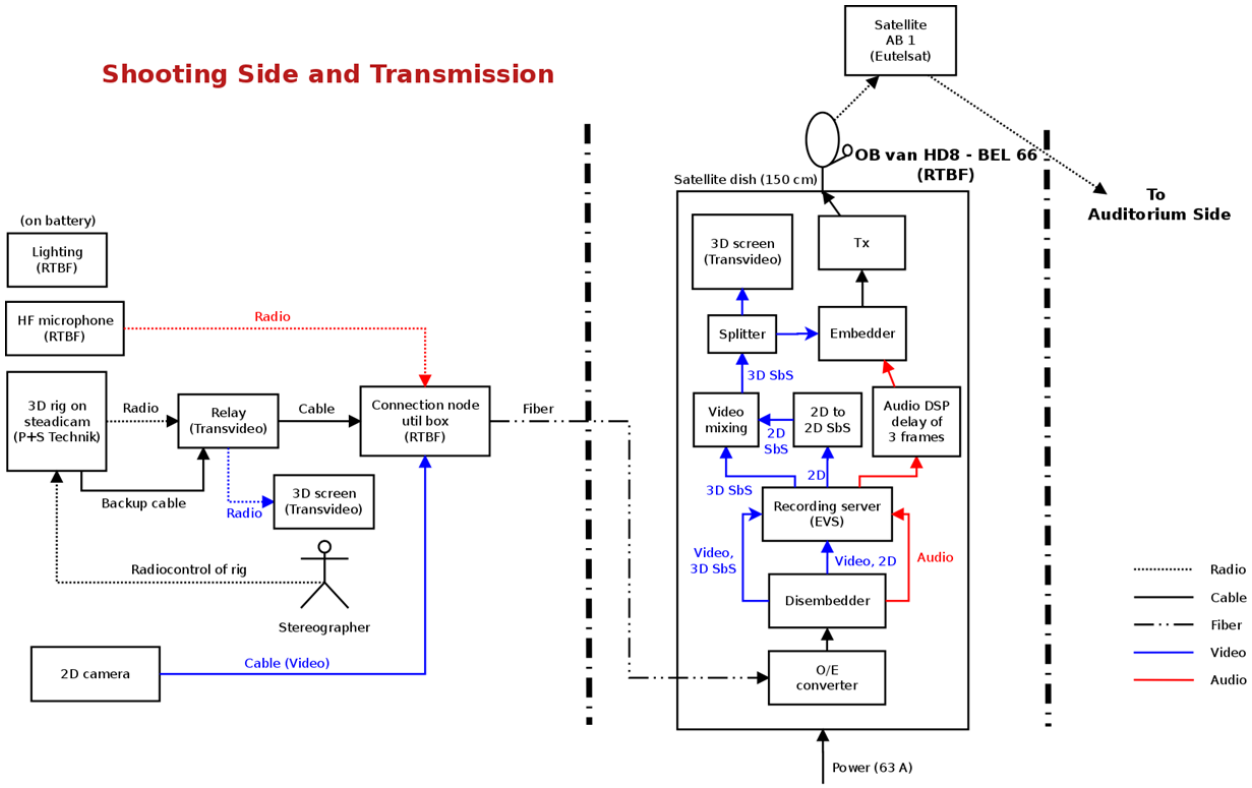


Figure 22.6. *Architecture of capture and of part of transmission*

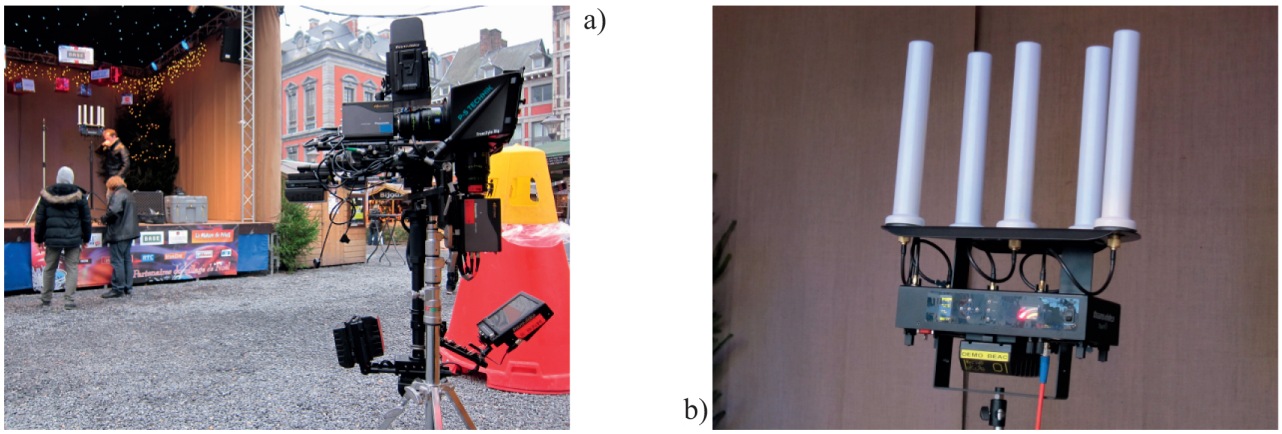


Figure 22.7. *a) View of the steadicam on its base and of the fixed radio reception/transmission system (with "candle" antennas). b) Transvideo reception/transmission system. The system allowed one to receive images from the S-3D rig and to send them directly to the control unit of the stereographer. The system was also able to send images via coaxial cables, in this case to the RTBF connection unit*



Figure 22.8. *a) The S-3D steadicam in operation. The steadicammer is followed by the stereographer, who controls the rig by radio remote control. b) One of the exhibitors interviewed during the retransmission, who quickly realized that brandishing a sausage at the client was an excellent way of demonstrating the "jumping out of the sceen" effect*

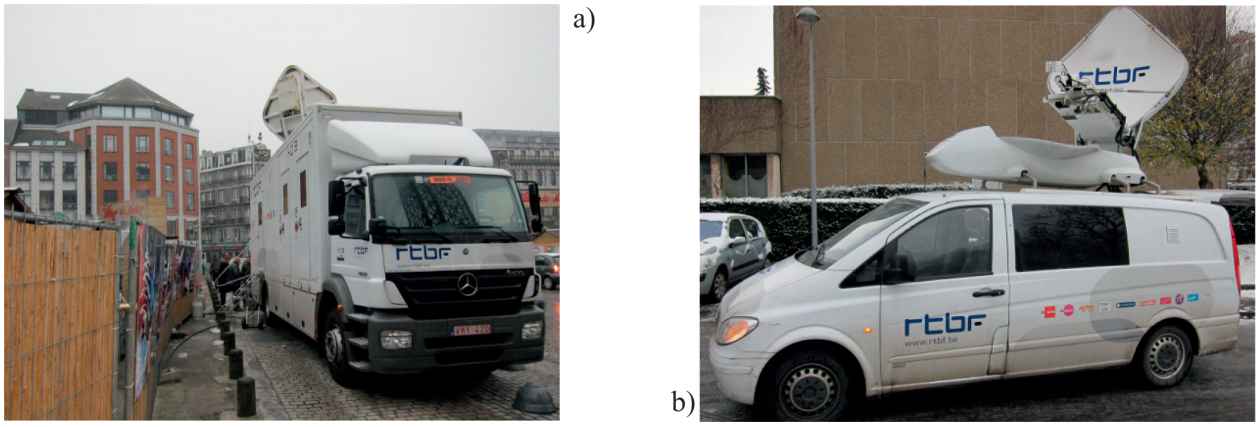


Figure 22.9. *a) RTBF HD8 OB van used to mix images and to send signals to the satellite (uplink). b) RTBF BEL69 satellite van used to receive signals from the satellite (downlink)*

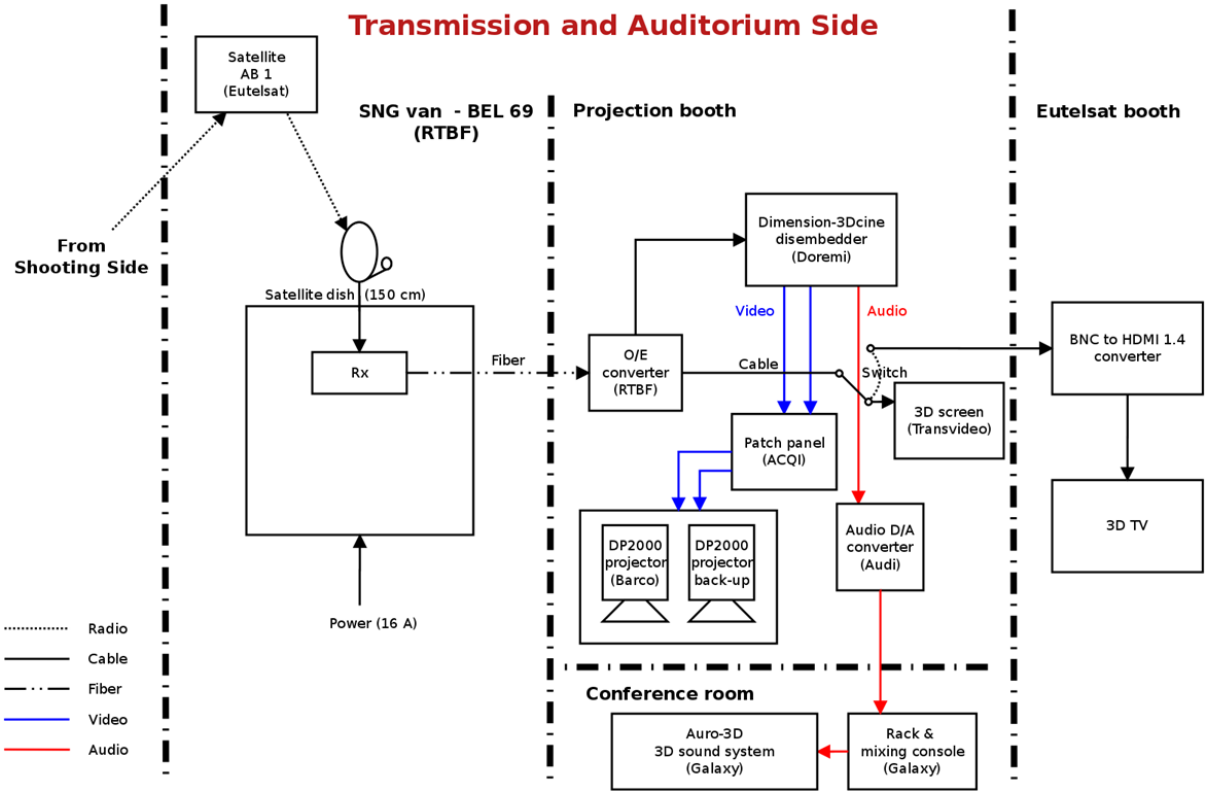


Figure 22.10. *Architecture of part of the transmission system and of the visualization system*

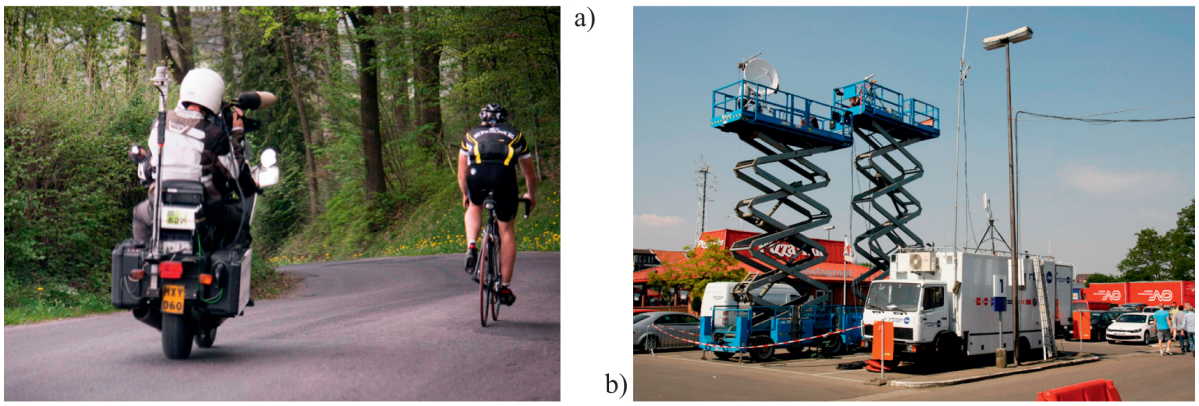


Figure 22.11. *a) View of the traditional radio transmission equipment on the back of the S-3D capture motorcycle. The side-by-side Davio converter is not visible. b) A variety of radio reception equipment on the finish line at Ans during the 2011 Liège-Bastogne-Liège race*