Trends in IVMSP

Patrick Wolfe, Harvard
Lina Karam, Arizona State Univ.
Moderator: Gaurav Sharma, TC Chair

Global Trends

- Applications Side
  - Mobile/distributed (cloud)
  - Streaming Video (this in some sense is already happening)
  - Stereo and 3-D Imaging
  - Computed (as opposed to captured) imaging
- Theory/Algorithms/Techniques
  - increasing cross-over between what were traditionally separate disciplines
  - significant overlaps with computer vision, communications, psychology, machine learning and AI, in addition to the traditional disciplines of mathematics and statistics.

Trends in Image, Video, and Multidimensional Signal Processing

IVMSP Technical Committee

Moderator: Gaurav Sharma

Presentation by

Patrick Wolfe, Harvard University, USA
Lina Karam, Arizona State University, USA

Trends in Image, Video, and Multidimensional Signal Processing: 3D Video Processing

IVMSP Technical Committee

Moderator: Gaurav Sharma

Presented by

Lina Karam
Arizona State University
Tempe, Arizona, USA
karam@asu.edu

2D to 3D
Temporal and spatial

2xFull HD (2x2M pixels)
60 to 240 Hz
24 Giga bits/sec

Full HD: 1920x1080 (2M pixels)
60 to 120 frames per sec (Hz)
6 Giga bits/sec

Spatial resolution

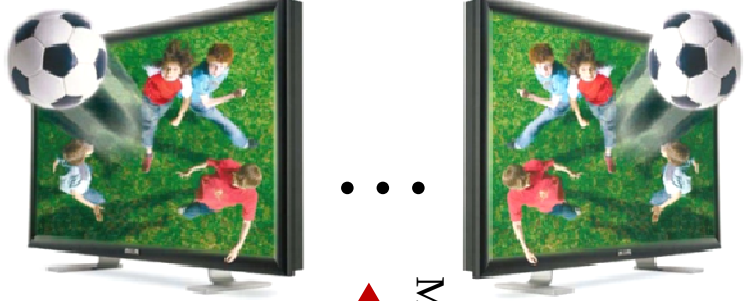
HD: 1280x720 (922K pixels)
60 to 120 frames per sec (Hz)
3 Giga bits/sec

temporal

SD: 720x480 or 720x576 (346K to 415K pixels)
60 to 120 frames per sec (Hz); 1.2 Giga bits/sec

spectral

SD: 720x480 or 720x576 (346K to 415K pixels)
30 to 60 frames per sec (Hz)

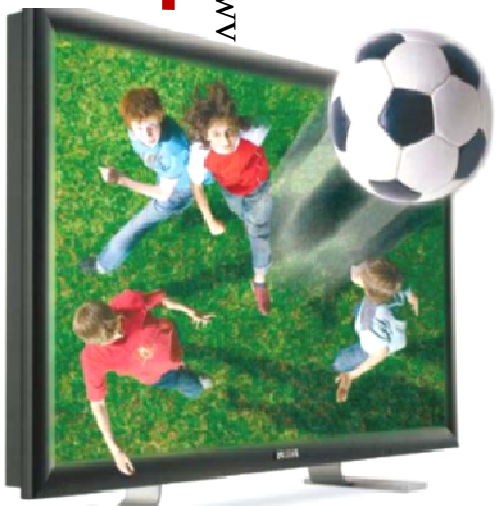


Multi-Viewv

2D to 3D
Temporal and spatial

Multi-Viewv

Full HD: 1920x1080 (2M pixels)
60 to 120 frames per sec (Hz)
6 Giga bits/sec

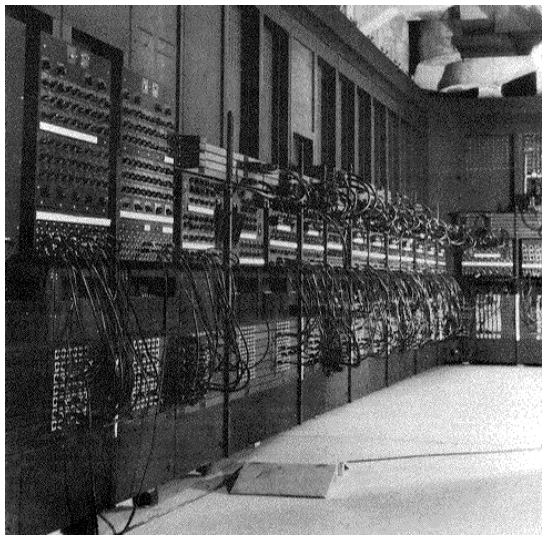


2xFull HD (2x2M pixels)
60 to 240 Hz
24 Giga bits/sec

Spatial resolution

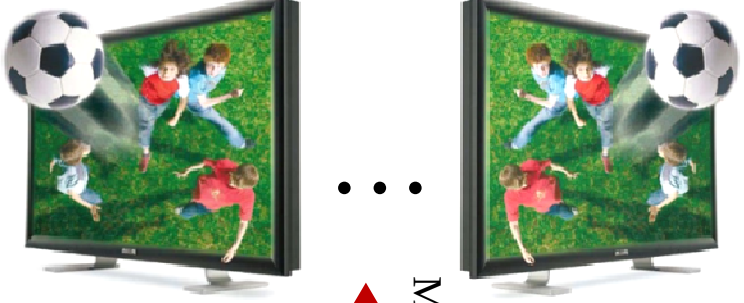
HD: 1280x720 (922K pixels)
60 to 120 frames per sec (Hz)
3 Giga bits/sec

SD: 720x480 or 720x576 (346K to 415K pixels)
60 to 120 frames per sec (Hz); 1.2 Giga bits/sec



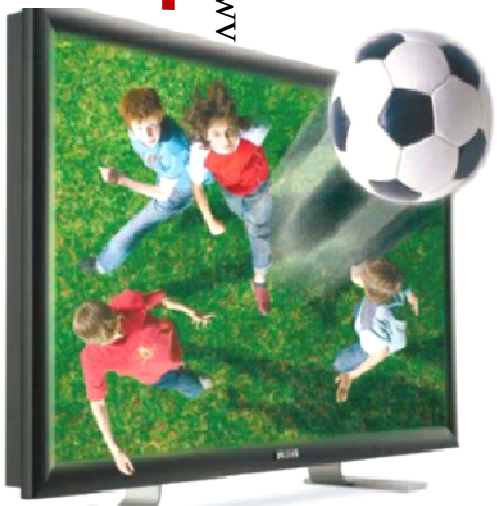
Portability and mobility

Multi-Viewv



Multi-Viewv

2D to 3D
Temporal and spatial



2xFull HD (2x2M pixels)
60 to 240 Hz
24 Giga bits/sec

Full HD: 1920x1080 (2M pixels)
60 to 120 frames per sec (Hz)
6 Giga bits/sec

Spatial resolution

HD: 1280x720 (922K pixels)
60 to 120 frames per sec (Hz)
3 Giga bits/sec

SD: 720x480 or 720x576 (346K to 415K pixels)
60 to 120 frames per sec (Hz); 1.2 Giga bits/sec

3D Video Acquisition

- Left –Right stereo pair (LR)
- Multi-View (MV) – allows “look around”
- 2D Video plus Depth (VD)

Efficient solutions needed

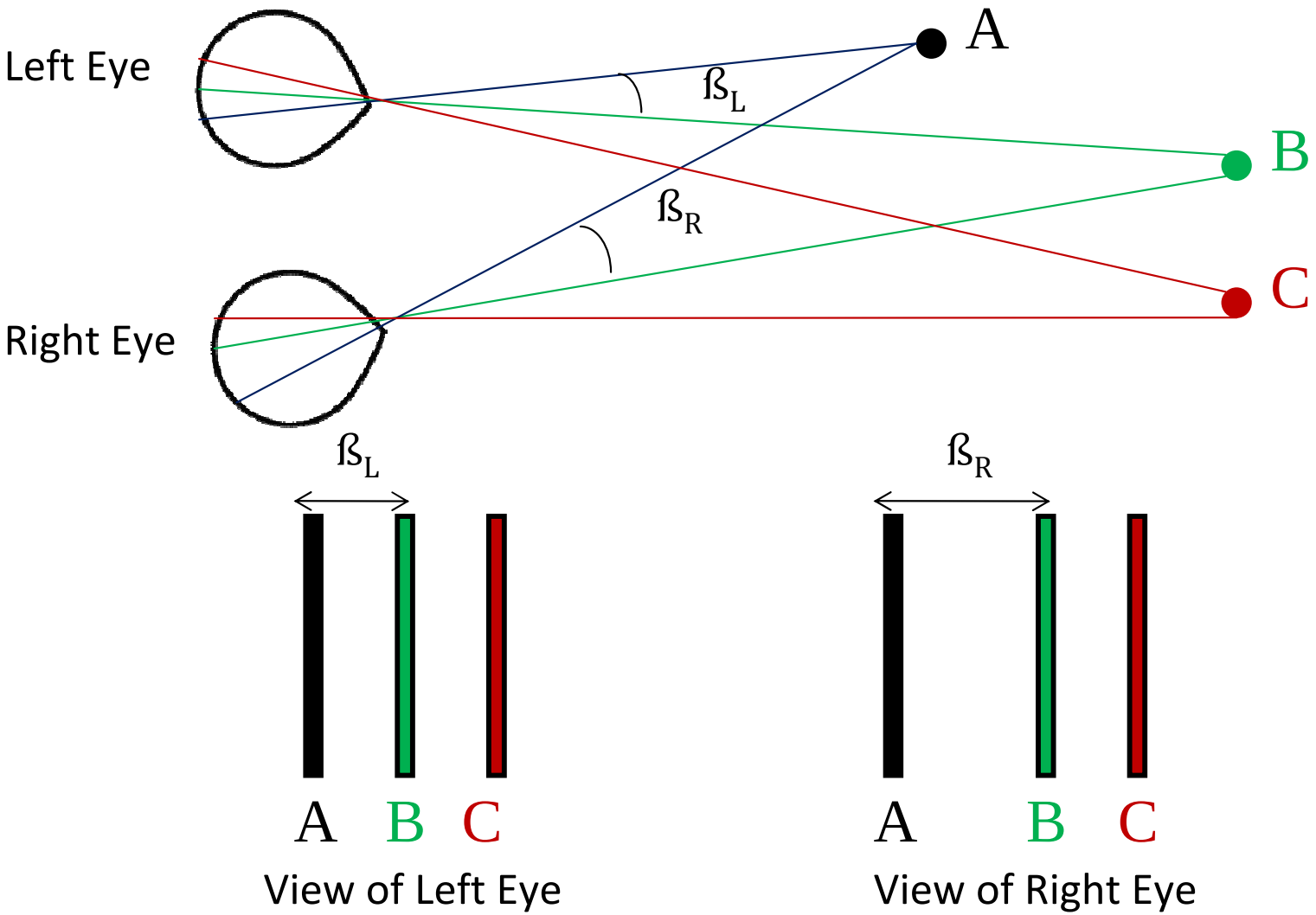
3D Video Delivery

3D Video Content (Post-)Production

- Improper 3D content is annoying! (binocular rivalry; discomfort; headaches)
- Need to ensure “good” 3D video content
  - Careful planning and capturing
  - Assessing quality of content
  - Editing and fixing captured content
  - Generating 3D content

3D Video Delivery

Binocular Vision

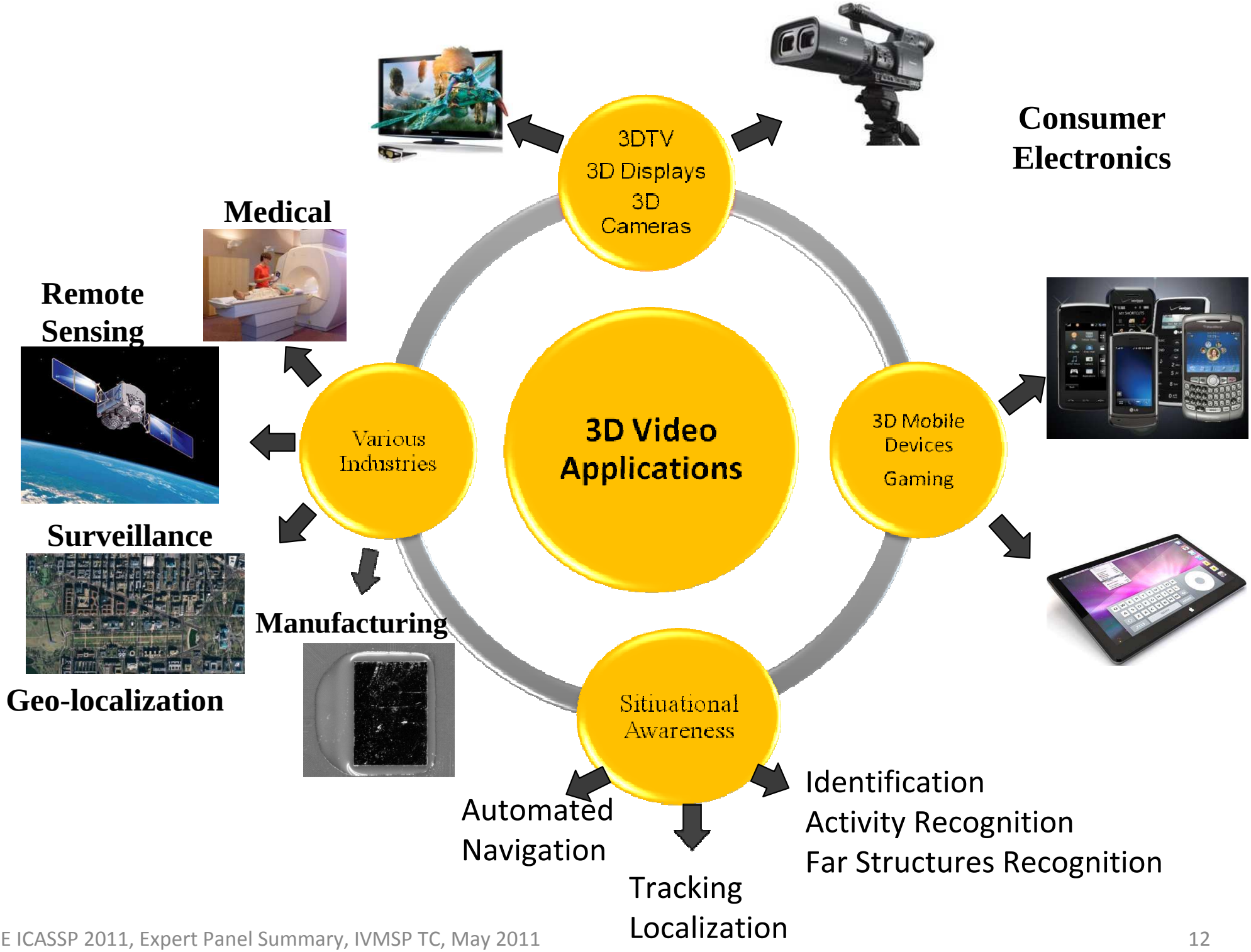


Binocular Vision

Binocular stereo is not the only means for depth perception

Spatial depth perception possible through monocular cues:

- Head movement parallax
- Linear perspective
- Shading and texture gradients
- Occlusion of more distant objects by near ones
- Accommodation (muscular tension to focus objects)


3DTV
3D Displays
3D
Cameras
Consumer
Electronics
Medical
Remote
Sensing
Surveillance
Geo-localization
Manufacturing
Various
Industries
3D Video
Applications
3D Mobile
Devices
Gaming
Sitiuational
Awareness
Automated
Navigation
Tracking
Localization
Identification
Activity Recognition
Far Structures Recognition

Enabling Solutions

- 2D to 3D Conversion
- Disparity Estimation; depth information generation
- Virtual View Synthesis
- 3D Video Compression and Transmission
- 3D Video Representation and Reformatting
- 3D Video Content Quality Assessment
- 3D Video Post-Processing (e.g., distortion correction, color matching, enhancement, …)
- Perceptual-based 3D Video Processing and Coding

2D to 3D Conversion

To compute the 3D structure (depth information) of points of a 3D scene from 2D images or 2D video:

- For any 2D video sequence, captured from a static camera and using the relative motion between an independently moving object and the camera.
- Captured by a single camera moving around a static scene
  - Structure from Motion methods (Jehara et al., IEEE SP Magazine, 1999; Pollefeys et al., Int. Journal Comp. Vis., 2004; N. Snavely et al. SIGGRAPH, 2006; K. Ni et al, "3D Image Geo-Localization using Vision-Based Modelling," ICASSP 2011)
- Captured simultaneously by multiple cameras (multi-view)
  - Estimation of camera intrinsic and extrinsic parameters
  - Structure from Silhouette methods (Liang & Wong, Image Vis. Comput., 2010)
  - Disparity estimation techniques (Structure from Stereo)
- Captured by a static camera in a static scene: monocular depth estimation
  - Exploit monocular cues: Shape from Focus/Defocus, Shape from Shading (Durou et al., Comp. Vis. Imag. Underst., 2008), Shape from Texture, Shape from Geometrical Properties (e.g., vanishing points and lines) , Occlusion Cues (Palou et al., Occlusion Based Depth Ordering on Monocular Images with Binary Partition Trees," ICASSP 2011)
  - Machine learning or rule-based methods (Saxena et al., T-PAMI, 2008)
  - User-guided and Segmentation based (Phan et al., "Semi-Automatic 2D to 3D Depth Conversion using a Hybrid Ransom Walk and Graph Cuts based Approach," ICASSP 2011)

Disparity/Depth Estimation

Estimate disparity or depth from a stereo image pair or from multiple views of the same scene (sparse or dense disparity map)

Two main categories:

- Local disparity estimation: depends only on local region surrounding current processed pixel (Cech & Horaud, "Joint Disparity and Optical Flow by Correspondence Growing," ICASSP 2011; faster than global methods and more accurate than pure local methods as neighboring pixel relations are not ignored completely)
- Global disparity estimation: finds disparity field by minimizing a global cost function over the entire image : can be more accurate than local methods but more computationally intensive (El Gheche et al., "Proximal Splitting Methods for Depth Estimation," ICASSP 2011; Gamal-Eldin et al., "A Novel Algorithm for Occlusion and Perspective Effects using a 3D Object Process," ICASSP 2011; Thirumala & Frossard, "Dense Disparity Estimation from Linear Measurements," ICASSP 2011; Khoshabeh et al., "Spatio-Temporal Consistency in Video Disparity Estimation," ICASSP 2011)
- Need to ensure: smoothness, temporal consistency (to avoid flicker), and need reliable handling of occlusions

Virtual View Synthesis

Key for 3D post-production and processing (e.g., depth adaptation depending on application, screen size and resolution, 3D display, to ensure a pleasant viewing experience; content creation for autostereoscopic multi-view displays, and Free Viewpoint Video functionality)

Typically performed from dense depth maps:

- Depth Image-Based Rendering (DIBR): estimates the depth maps from existing views, and these are then used to synthesize virtual views. Issues considered include handling depth discontinuities to reduce artifacts along object borders, checking consistency of disparity maps of views, handling occlusions, reducing blur due to errors in camera parameters (Surn et al., "Error Compensation and Reliability Based View Synthesis," ICASSP 2011; Tran et al., "Spatially Consistent View Synthesis with Coordinate Alignment," ICASSP 2011; Pearson et al., "Accurate Non-Iterative Algorithm for Image Based Rendering," ICASSP 2011; Jain et al., "Efficient Stereo-to-Multiview Synthesis," ICASSP 2011)
- Layered Depth Images (LDI): several color and depth values are stored for each pixel to compensate for possible occlusions
- Intermediate View Reconstruction (IVR): intermediate views between neighboring views with dense depth information are rendered; no occlusion problem (Zitnick et al., SIGGRAPH, 2004)
- Image-based warping: does not require reprojection to 3D space and operates directly in image space by defining feature correspondences between source and target images, which are used to generate virtual view through interpolatation and morphing (D. Mahajan et al., ACM Trans. Graphics, 2009)

3D Video Compression and Transmission

- Reduction of inter-view redundancies through inter-view prediction (disparity estimation) in addition to temporal redundancies (Davidoiu et al., "Rate-Distortion Analysis in a Disparity Compensated Scheme," ICASSP 2011; Konieczny et al., "Extended Inter-View Direct Mode for Multiview Video Coding," ICASSP 2011)
- Stereo and MVC Extensions of H.264/MPEG-4 AVC (Vetro et al., "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard," Proceedings of the IEEE, 2011)
- Stereo coding using existing standards (e.g., MPEG-2, H.264/MPEG-4 AVC; SVC): requires frame-sequential format (left, right, left, right, …) or frame-compatible formats (e.g., side-by-side, top-bottom, row-interleaved, column-interleaved, checkerboard)
- Asymmetric 3D Video coding (Saygili et al., ICIP 2010; Aflaki et al., ICIP 2010)
- 2D plus Depth Video Coding
- 2D plus Delta Video Coding
- 3D Video Transmission (Gurler et al., Proc. of the IEEE, 2011; Schierl et al., Proc. of the IEEE, 2011; Xiang et al., "Auto-Regressive Model Based Error Concealment Scheme for Stereoscopic Video Coding," ICASSP 2011)

3D Video Content Quality Assessment

- Important for evaluating produced content, assessing effects of degradations due to acquisition, compression, or transmision, and for assessing the performance of 3D processing schemes, and 3D displays.
- Subjective quality of experience assessment for 3D Video: in addition to typical overall quality, factors include depth perception, naturalness, immersiveness, comfort (*ITU-R BT.1438 Subjective assessment of stereoscopic television pictures*)
- Objective quality metrics for 3D video: still work needed (Maalouf et al., "CYCLOP: A Stereo Color Image Quality Assessment Metric," ICASSP 2011; Sazzad et al., QoMEX 2010; Malpica et al., VPQM 2009; upcoming IEEE Signal Processing Magazine, Special Issue on Multimedia Quality Assessment to be published in 2011; upcoming QoMEX 2011)
- Shortage of good 3D Video databases for evaluating quality metrics

3D Video Content Post-Processing

- Improve the similarity between the left and right views to reduce binocular rivalry
- 3D Restoration (Bal et al., "Detection and Removal of Binocular Luster in Compressed 3D Images," ICASSP 2011)
- 3D Enhancement (Lu et al., "A Revisit to MRF-Based Depth MAP Super-Resolution and Enhancement," ICASSP 2011)
- 3D Reformatting and Conditioning

Perceptual-based 3D Video Processing and Coding

- Need to exploit more HVS characteristics:

  -improve the performance and/or efficiency of 3D Video Processing

  - perceptually-optimized 3D Video Compression

  - perceptual-based enhancement and restoration

Further Reading

- Proceedings of the IEEE, Special Issue on 3D Media & Displays, vol. 99, No. 4, April 2011

Thank You