Digital and Interactive Multimedia

Notes

Alberto Lazari

I Semester A.Y. 2023-2024

Index

1. Presentation

1.1. Outline

1. 3d vision and acquisition
2. Lab + 3d processing
3. Seminar?
4. Lidar and automotive
5. VR
6. AR
7. SLAM
8. Seminars

1.2. Exam

Two parts:

1. 2 open questions (10 pts each)
2. 10 multiple choice (1 pt each)

Dates:

- Jan 22 12:30
- Feb 07 12:00

Multimedia Multiple types of media combined (audio, video, text, …)

2. Media types

2.1. Image

Images can be formed by combining:

- Illuminance: $i(x, y)$
- Reflectance: $r(x, y)$

2.2. Video

- Fast sequence of single images
- At least 25 fps to see motion, because of retina's persistence phenomenon

3. 3D Perception

3 different ways to perceive:

- Oculomotor (binocular vision)
- Static visual
- Motion

3.1. Oculomotor cues

- Accomodation: changes in focal aperture in the crystalline
- Vergences: movements of the eyes to merge the two images

3.2. Visual cues

Features of images that allow to create a 3D (static) perception:
- Occlusions
- Relative dimensions: far → small, close → big
- Textures
- Linear perspective: straight lines in perspective
- Aerial perspective: fog in the far distance
- Shadows

3.3. Motion

Also motion can create 3D perception of a (2D) video:
- Motion parallax
- Relative angular velocity: far objects appear slower
- Radial expansion
- Shadow movement

4. Binocular vision

Requires:
1. Simultaneous perception: two images in both eyes
2. Fusion:
   - Motor: accomodation + vergence
   - Sensory: create single image
3. Stereopsis: interpret two images add 3D perception to the fused image

Horopter area where points don't produce duplicate images (Panum area)

Lecture 3 – 10/10

5. 3D media

Technologies are divided in:
- Passive
- Active

w.r.t. the viewer

5.1. Passive 3D rendering

- Lens arrays (3D cards)
- Parallax occlusion: bands with holes, similar to lens (3DS?)
- Anaglyph: red/blue glasses (colors are off, though)
- Dolby 3D: slightly different colors per eye (requires wheel on projector and double the frame-rate)
- Polarized light: need to stay still and not rotate head
- Circular polarization: it needs:
  - Two projectors with polarizers
  - Special silver reflective screen
  - Glasses

5.2. Active 3D

Example: Nvidia glasses, alternate shutters

6. Stereo images

- Disparity map shows intensity of parallax effect between two images (two eyes)
- Stereo images have to be rectificated → point the object in focus

Lecture 4 – 31/10

7. Camera parameters

- Intrinsic: depend on the camera itself
- Extrinsic: camera location/orientation

8. Features

Feature recognition can be useful for:

- Camera calibration
- Stereo image creation
- Tracking
- Image mosaicing

They have to be invariant to:

- Illumination
- Scale
- Rotation
- Affine (similarity, slight changes)
- Perspective projection

8.1. SIFT algorithm

Used for feature recognition (ex, for image matching)

Lecture 5

8.2. Direct Linear Transform (DLT)

Infer the 11 parameters (5 intrinsic + 3 rotations + 3 translations) from image. At least 6 points are needed

Lecture 9

9. 3D reconstruction

9.1. Rectification

Make the image rows match with epipolar lines

9.2. Point clouds

Set of points in the space. Provides information for each point about:

- Geometry: position
- Color, reflectance, … (optional)

10. Camera arrays

Possible applications:

- HDR
- Higher resolution
- Tiled panoramas
- Synthetic aperture photography: show subjects partially hidden behind occluders
- Hybrid aperture photography: mix various apertures in the same image (ex light fields)

10.1. Light fields

Use microlens arrays to merge various point of views, apertures and focus in a same image, allowing for post-processing access of those informations

Lecture 10

11. VR

VR experience the user feels **immersed** in a **responsive** virtual world → dynamic control of view point

11.1. Immersion

VR is immersive because of:

1. Stereovision, provided by **headset**
2. Dynamic control of viewpoint
3. Surrounding experience

Can also provide:

- Various Degrees Of Freedom (DOF)
- Interaction with controllers
- Aptic feedback

11.2. Navigation

- Controller/keyboard/joystick: more nausea-prone
- Teleporting (movement has to be not too quick)
- Threadmills

Lecture 11

12. 360 images

Acquisition with:

- Multiple cameras
- Catadioptric: reflection on curved mirrors
- Fish-eye lens

Sphere construction needs:

- Multiple cameras (can't acquire the whole sphere)
- Stitching

12.1. Sphere representations

How to represent a sphere on a flat topology?

- Equirectangular projection: geographical maps' method

Great distortions and low algorithms performance
- Cube map: good performances + natural images, but artifacts
- Pyramid projection: lots of discontinuities, but clear center (pyramid basis). Useful for streaming

Lecture 13

13. Quality of Experience

How to objectively measure it?

13.1. Saliency maps

Interesting regions, that catch user's attention and focus

Can be generated with:
- Bottom-up approaches: ex Gabor filters, based on feature detection
- Top-down ones

13.2. 360 content

Rendering can be done either:
- Client-side: requires full video streaming (90% of the FOV is disregarded) and processing
- Server-side: render and stream only necessary parts → reduce bandwidth. Can be done with:
  - Two-tier streaming: parallel stream of base, low-res video + HD viewport area. Bad performance, because two streams compete for resources
  - Viewport-adaptive streaming: more versions for different possible viewports. Requires server-side storage
  - Tile-based streaming: sphere divided in tiles, to be streamed
    - at different resolutions (full delivery)
    - possibly not streamed at all (partial delivery, bad QoE)

Predict head movements with saliencies

Lecture 14 – Seminar

14. Immersive media compression

Point clouds are difficult to compress: sparse, irregular… → quantize (voxelize)

Then just use ***AI*** to reconstruct:
- Uses 3D convolutional neural network
- Works perfectly for dense point clouds, not so much on sparse ones
- Works on static point clouds (models, not animations/videos)

Alternatively use graph-based solutions:
- No voxelization
- Results are too smooth
- Point properties difficult to compress (color)

Lecture 17

15. Objective evaluation (QoS)

Image quality assessment: compare and provide evidence of improvement

Subjective tests are too complicated, expensive, difficult…

15.1. Full reference

Requires a reference of the original picture (?)
- PSNR/MSE: not consistent with human perception (blur looks not destructive)
- SSIM $\in [0, 1]$: improvement, measures similarity between two images. It compares luminance
- VMAF $\in [0, 100]$: for video

15.2. Reduced reference

Uses feature extraction

15.3. No reference

Brisque and NIQE (lower is better)

16. QoE

Depends on many factors:
- Technological
- Multi-sensory
- Emotions (frustration, surprise)

Lecture 18

17. Subjective assessment

Most reliable way of measuring multimedia quality

In order to be reliable needs:
- Large number of users (at least 15, screened for visual acuity)
- Description of:
  1. Laboratory equipment: screen, distance, illumination, …
  2. Data set: contents used
  3. Methodology: rating target (quality, comparison, impairment) and scale, stimuli (single/double)
  4. Score processing: mean, outlier detection, …
- Introduction to method, training sequence. Consider a break after that (to answer questions)
- No more than 30 mins sessions

Spatial Information (SI) complexity in image (spatial detail present)
Temporal Information (TI) frequency of changes in video

17.1. Learning effect

Calibrate time to balance:
- Training: user becomes more sensitive
- Tiredness: user becomes less sensitive

Control it by:
- Showing full range of stimuli (SI/TI)
- Short sessions
- Pay participant

- Randomize stimuli

17.2. Methods

- Single-Stimulus/Absolute Category Rating (SS/ACR): single image at a time, index of presentation
- ACR with Hidden Reference (ACR-HR): a picture is secretly a reference. Differential MOS between scores (against the reference)
- Double-Stimulus Impairment Scale (DSIS): rate degradation of image, given a non-impaired reference (first, then the other is showed)
- Double-Stimulus Continuous Quality-Scale (DSCQS): two images, one is reference (don't know which). Vote on whole presentation, on vertical scale.

  Results are to be considered as differences from reference
- Single-Stimulus Continuous Quality Evaluation (SSCQE): continuously rate video quality, with slider
- Simultaneous Double-Stimulus for Continuous Evaluation (SDSCE): continuously rate side-by-side video, knowing which is reference
- Subjective Assessment of Multimedia Video Quality (SAMVIQ): various different sequences, with explicit and hidden references. User can go backward etc…
- Pair-wise Comparison (PC): two videos, one after other. Select the best
- Simulator Sickness Questionnaire (SSQ): 360° video, 0-3 rating. At least 28 subjects, no more than 25 continuous mins, no more 50 rating mins. < 1.5 h participation

17.2.1. Comparison

- Methods that use explicit references measure fidelty (DSIS)
- ACR is easier to implement
- ACR-HR is even better, because it only considers the difference between the reference (no bias towards specific pictures)
- PC can be used as a last resort for the items that have the same rating (direct comparison, 1v1)

17.2.2. Designs

Need to show all pairs to compare:
- Full design: $O(n^2)$
- Reduced design: assume transitivity + make sorting algorithm: test becomes human merge sort $O(n \cdot \log n)$

17.3. MOS process

Mean Opinion Score (MOS) average observer rate

Standard deviation $s = \sqrt{\frac{1}{n} \sum_{i=0}^{N} (x_i - m)^2}$, where N: sample size, m: mean

Standard error $\mathrm{SE} = \frac{s}{\sqrt{N}}$

Confidence Interval $\mathrm{ci} = m \pm 0.95 \cdot \mathrm{SE}$

95% probability that user's average is within confidence interval

17.4. Crowdsourcing

Alternative method, ask people from internet, under compensation:

- No controlled environment

Lecture 22

18. Augmentation/Mediation

Augmentation amount of virtual content on top of real world

Examples:
- Information overlay
- Spatial anchor of virtual objects

Mediation change surroundings

Examples:
- Beautification
- Diminished reality

18.1. AR

- Strong AR: full surroundings knowledge (precise tracking, semantic understanding)
- Weak AR: little tracking/interaction

Technological solutions:
- Marker-based AR: very precise tracking, if light conditions are good. When no marking the experience disappears
- Marker-less AR: more flexible, might not be suitable for the experience (not enough space/ does not make sense)
- Location-based AR: Google Maps, not always accurate, because of technologies/sensors

18.1.1. SLAM algorithm

Combine visual + inertial sensors to:
- Create map of environment
- Continuously position device

System:
1. Sensors
2. Front-end: feature extraction of real environment
3. Back-end: localize POV, reconstruct model, analyze frames
4. Estimate: reconstruction of environment, with locations of features + POV

3D maps are usually meshes