Hauptseminar Medientechnik SoSe18

The Scientific Challenges for the Wide-Ranging Field of Virtual Reality

Tamay Aykut
## Topics

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Head-Mounted Display Optics</td>
<td>2</td>
</tr>
<tr>
<td>2. Head Motion Tracking</td>
<td>3</td>
</tr>
<tr>
<td>3. Direct Sparse Odometry</td>
<td>4</td>
</tr>
<tr>
<td>4. Stereoscopic 360 Degree Video</td>
<td>6</td>
</tr>
<tr>
<td>5. Feature Extraction and Object Detection in 360 Degree Videos</td>
<td>7</td>
</tr>
<tr>
<td>6. Compression of (stereoscopic) 360 Degree Videos</td>
<td>9</td>
</tr>
<tr>
<td>7. Optimal Tiling in 360 Degree Video Streaming</td>
<td>10</td>
</tr>
<tr>
<td>8. Motion Sickness - Causes and Remedies</td>
<td>12</td>
</tr>
<tr>
<td>9. Wearable Haptics for Augmented VR</td>
<td>14</td>
</tr>
<tr>
<td>10. The Visual Search Standards: From Classical Image Retrieval to Deep Learned Features</td>
<td>15</td>
</tr>
<tr>
<td>11. Drone-Augmented Human Vision for VR</td>
<td>16</td>
</tr>
<tr>
<td>12. Head-Motion Prediction</td>
<td>17</td>
</tr>
</tbody>
</table>


Topic 1

Head-Mounted Display Optics

Head-Mounted Displays (HMDs) are fully immersive VR systems and provide the user the most direct experience. Emerging technologies in microdisplay devices contributed to the breakthrough in HMD optics. The student will elaborate the theoretical background in HMD optics such as image projection, lens distortion correction, focus cues, and the vergence-accomodation conflict in human vision. Besides that, the student is asked to discuss the core display technologies like adaptive and gaze-contingent focus displays, near-eye light field displays, Maxwellian-type displays, etc.

Supervisor: Tamay Aykut

References

Gordon Wetzstein, http://stanford.edu/class/ee267/, HMD Display Optics I & II (Lecture 7, 8), 2017


Topic 2

Head Motion Tracking

Head-mounted virtual reality displays can be used to place a user inside an augmented or artificially created 3D scene. The image shown by a head-mounted display needs to be generated according to the position of the user’s head - whether the user is looking up or down, left or right in the scene. The image on the display needs to be adjusted even when the user moves his head very slightly. Therefore, accurate real-time information must be obtained about the position and motion of the head of the user and the head-mounted display.

Supervisor: Alexandra Zayets

References


Topic 3

Direct Sparse Odometry

Tracking the position of a head-mounted display is an important task in VR. Usually this is done through tracking of beacons. Another option is to estimate the ego-motion of the head-mounted display through cameras. Direct Sparse Odometry (DSO) is a state of the art visual odometry method that generates a dense model of the environment and can be used for tracking in VR tasks. The aim of this seminar topic is to present the DSO method and compare it to previous dense tracking approaches.

**Supervisor:** Adrian Garcea

**References**

Topic 4

Stereoscopic 360 Degree Video

Stereoscopic VR systems add the sensation and perception of depth and, hence, lead to a more realistic experience of the virtual or remote environment. With more and more HMD devices entering the market and the continuing high interest in VR systems, there is a strong demand for 3D 360° content, especially on platforms like Youtube or Facebook. Most 360° videos that are available nowadays are monoscopic and, thus, do not provide the perception of depth, even when watched through an HMD. This is due to the fact that capturing stereoscopic panorama videos is significantly more difficult. While streaming monoscopic 360° videos in real-time is considered state-of-the-art, this is not yet the case for the aforementioned stereoscopic 360° systems. The student is asked to elaborate the most recent state-of-the-art omnistereoscopic video technologies and compare, among other things, their performance and real-time capability. This research goes beyond typical papers and involves digging into most current patents. The student will discuss the requirements to acquire and stream stereoscopic 360° video, which includes the appropriate projection and capturing formats such as the Omni-directional Stereo (ODS) (or ODS approximation).

Supervisor: Tamay Aykut

References


Figure 4.1: (a-c) Multicamera arrangements for the acquisition of stereoscopic panoramas. (d-f) Consumer products which are based on those camera arrangements.


Feature extraction is an important part of various applications, such as object detection, homography estimation, SLAM algorithms, content-based image retrieval, etc. Challenges include similarity transformations such as rotation and translation change, as well as illumination or even viewpoint changes. Autonomous driving, for example, requires real-time processing of images, which is still a challenging task. In this seminar the differences between feature extraction and object detection in normal and 360° images are to be outlined. The student is required to explain and understand the principles of state-of-the-art techniques.

Supervisor: Martin Oelsch

References


Virtual reality (VR) creates an immersive experience of the real world in virtual environment. Due to the technological advancements in recent years, VR technology is growing very fast. Since VR is visualizing the real world experience, the image or video content which is used must represent the whole 3D world characteristics. 360° videos demonstrate such characteristics and hence are used in VR applications. However, these contents are not suitable for conventional video coding standards, which use only 2D video format content. Thus far, the focus for 360° video compression is to find a proper projection that transforms a 360° frame into a rectangular planar image that will have a high compression rate. In this seminar, we will compare the performance of state-of-the-art video codecs (e.g. AV1, HEVC and JVET) on 360 (spherical) video compression, as well as different projections and quality evaluation metrics. Finally our target is to find a good combination which is good for compression of 360° videos.

**Supervisor:** Kai Cui

**References**


Zou, Wenjie, Fuzheng Yang, and Shuai Wan. "Perceptual video quality metric for compression artefacts: from two-dimensional to omnidirectional.” IET Image Processing

For 360° spherical images and videos, current methods that directly encode them into a bit stream are not yet mature. Therefore, the state-of-the-art 360-degree video streaming systems usually seek an alternative way by exploiting the advantages of two-dimensional video coding technology (e.g., H.264/AVC, HEVC), which first projects the 360° spherical surface onto a rectangular image/frame and then applies the advanced two-dimensional video coding to encode the rectangular image/frame into a bit stream. To support the pixel resolution of the displayed viewport as 4K (3840x2160), the resolution of the rectangular projection should be at least 12K (11520x6480), which indicates that the viewport area covers roughly 14% of all pixels in the rectangular projection. This leads to recent researches on an efficient streaming technique for 360° videos, named tile based approach [1], where the two dimensional projection is divided spatially into small rectangular tiles and each temporal sequence of these small rectangular tiles in the same spatial location is treated as an individual source for video encoding. In this way, only the tiles that cover the viewport will be transmitted to the user. As shown in Fig. 7.1, the size of the tiles affects conversely the coding efficiency and the transmission efficiency. If we decrease the tile size, the transmission efficiency is improved since the non-overlapped area between the viewport and the transmitted tiles becomes smaller. On the other hand, a smaller tile size results in larger number of tiles per frame, which in turn increases the number of headers for the tiles and results in efficiency reduction in the intra- (spatial) and inter- (temporal) prediction. Therefore, we need to study the optimal determination of the tile size, by considering both the video content statistics and the viewports (and viewport prediction) of users. Another promising research direction is that, instead of fixed tiling, we may want to cover the whole rectangular frame with some sub-rectangular tiles with different sizes, which, if encoded together, could achieve the optimal tradeoff between coding and transmission efficiency.

**Supervisor:** Chenglin Li
Figure 7.1: Impact of tiling partition of the transmission and coding efficiency.

References


Topic 8

Motion Sickness - Causes and Remedies

In interactive VR systems, motion sickness is a limiting factor for the Quality of Experience and can drastically decrease the feeling of presence and lead, in the worst case, to a termination of the VR (or Remote Reality) experience. The student will first work out the fine details between various forms of motion sickness (e.g. simulator sickness, VR sickness, visually induced motion sickness (VIMS), etc.). Besides that, the student is asked to debate the fundamental causes of motion sickness and discuss state-of-the-art methods to prevent the emergence of it.

Supervisor: Christoph Bachhuber

References


Topic 9

Wearable Haptics for Augmented VR

The task of this topic consists of a summary of available wearable haptic devices for the augmentation of VR/AR systems. The student particularly should investigate, which perceptual dimensions such as stiffness, friction, or roughness, can be displayed using novel approaches in this field of research. Additionally, current limits and potential future improvement need to be identified and discussed.

Supervisor: Matti Strese

References


Topic 10

The Visual Search Standards: From Classical Image Retrieval to Deep Learned Features

This topic provides a comprehensive overview over the MPEG standardization efforts for compact descriptors for visual search and video analysis suitable for retrieval tasks in the context for mobile augmented reality. Extracting reliable local and global image cues from an image or video sequences allows for example to identify objects in a mixed virtual environment and show additional information as overlay in the current viewport. To achieve this goal, an efficient representation of the image content must be extracted, coded, transmitted and the information has to be fed back into the virtual reality system. To this end, we will analyze the concepts behind two different standards in more detail: MPEG-CDVS (Compact Descriptors for Visual Search) and the upcoming MPEG-CDVA (Compact Descriptors for Video Analysis).

**Supervisor:** Dominik Van Opdenbosch

**References**


Topic 11

Drone-Augmented Human Vision for VR

Nowadays, drones are used in many new applications and in different areas. Especially the fusion of augmented reality (AR) and drones show exciting fields of application, such as rescue missions or infrastructure inspection simulation. In this regards, the student will investigate the potential of drone-augmented human vision, e.g., exploring the environment with augmented virtual objects and controlling the drone indirectly from a self-centered viewpoint. The topic can be divided into two main fields:

- Understanding of how the user’s view can be synthesized (i.e., stereoscopic augmented reality) from a 3D reconstruction of the indoor environment using image-based rendering in line with vSLAM.
- Investigating current methods in real-time omnidirectional collision avoidance control of the drones based on the reconstructed environment by taking augmented virtual objects into account.

Supervisor: Mojtaba Karimi

References


Topic 12

Head-Motion Prediction

In the early 90’s, researchers have already been starting to work on head motion prediction to compensate the local lag between head motion and display response, the motion-to-photon latency, caused by the time needed for tracking the head and rendering the imagery onto the HMD. Even nowadays, it takes at least around 30-50ms to render imagery to an HMD. That is the reason why many proposed methods aimed to compensate latencies in the range of 10-100ms. But head motion prediction is also used for tile-based 360° video streaming algorithms. Rather than sending the whole imagery and consuming large parts of the communication capacity with unused data, only a certain portion of the imagery is sent to the user. This portion is chosen by considering the future orientation of the users head.

Supervisor: Jingyi Xu

References


