DRIVES

Development and Research on Innovative Vocational Education Skills

COMPUTER VISION FUNDAMENTALS

U2.E8. VIDEO BASICS

Computer Vision Expert

May 2021, Version 1

Co-funded by the Erasmus+ Programme of the European Union

The Development and Research on Innovative Vocational Educational Skills project (DRIVES) is co-funded by the Erasmus+ Programme of the European Union under the agreement 591988-EPP-1-2017-1-CZ-EPPKA2-SSA-B. The European Commission support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

LEARNING OBJECTIVES

The student is able to …

CVE.U2.E8.PC1	The student knows the differences between working with image and video.
CVE.U2.E8.PC2	The student understands video capturing.
CVE.U2.E8.PC3	The student knows how to perform video analysis and manipulation.

IMAGE

- An image is a 2-D array of numbers (intensity values, gray levels). Gray levels range from 0 (black) to 255 (white).

- A colored image is 3 2-D arrays of numbers: Red, Green and Blue.

- Resolution (number of rows and columns)

WORKING WITH VIDEO

A **video** is a series of **images** played in sequence at a specified frame rate, usually from software playing a **video** file (digital) or hardware playing a tape (analog). In other words, **video** has movement.

A video is a set of frames. When studying a video, we can either study **a video stream** (live image feed) or **a video sequence** (fixed-length video).

- In a video stream, we consider the current image and the previous ones (an ongoing video for on-line real-time procession, the future frames are not known).
- In a video sequence, we have access to the full video, from the first image to the last (a video of fixed length.

DIFFERENCE BETWEEN IMAGE AND VIDEO

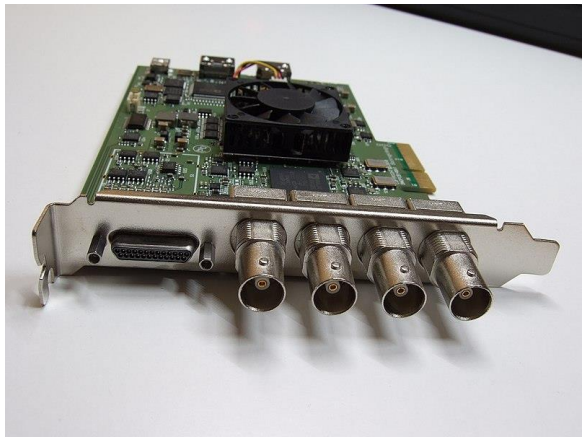
- Video is much larger object than an image.
- Frame width of consumer video is usually the line range from 3-5 image per second to 30 or 50 frames per second.
- The uncompressed data stream from Full HD video can reach 300 megabytes per second.
- Super resolution algorithms can actually reconstruct higher resolution relies on information that is removed from video during compression. Only other algorithm that hallucinate higher resolution can be applied.
- Videos need a lot of storage space and generally are not using AI. So, with video, we simply have raw image data to work with. But there is a key difference, the **motion** that is the only difference between an image and a video. Can lead to action understanding, pose estimation, or movement tracking.
- Compared to annotating individual images, video sequences provide the advantage of temporal consistency between consecutive frames.

VIDEO CAPTURING

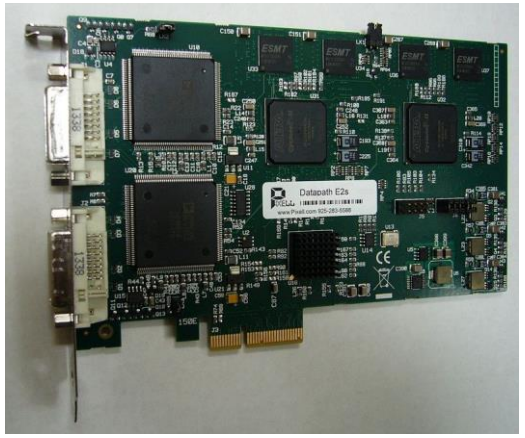
Video capture can be described as the process of converting an analog video signal like ones that are produced by a video camera, DVD player, or television tuner to digital video and sending it to local storage or to external circuitry. The resulting digital data are referred to as a digital video stream, or just a video stream.

To capture video from analog video sources, it is needed special electronic circuitry. This is performed by a **video capture device**. These devices usually use integrated circuit video decoders to convert incoming video signals to a standard digital video format, and extra circuitry to convey the resulting digital video to local storage or to circuitry outside the video capture device, or both.

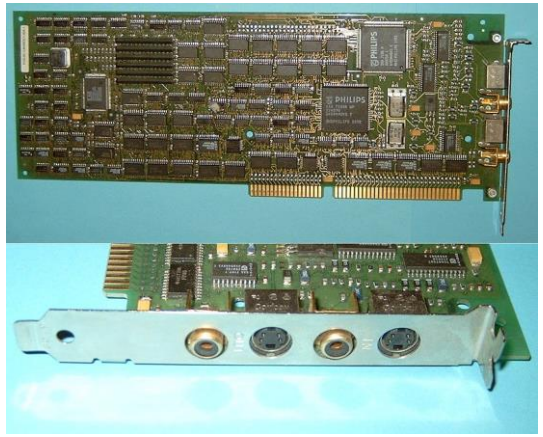
VIDEO CAPTURING



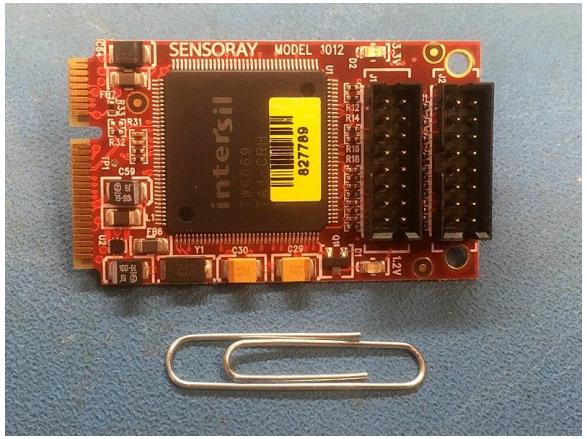
A dual link 3 Gb/s SDI video capture card for PCIe (Blackmagic Design *DeckLink HD Extreme 3D*)



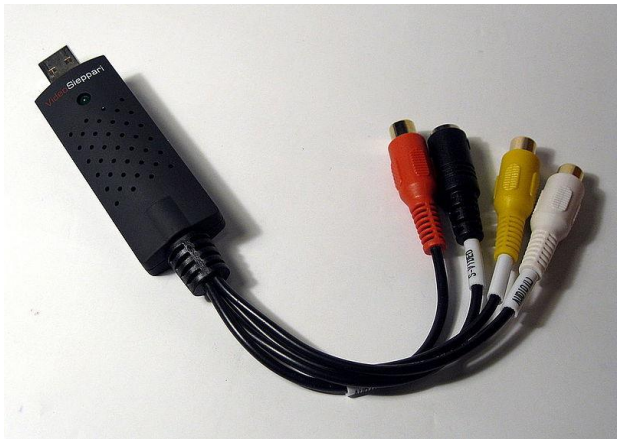
A PCIe 2-port HDMI video capture card (Datapath VisionRGB-E2s)



ISA analog video capture card



A Mini PCIe card that simultaneously captures 8 video and 8 audio signals (Sensoray 1012)



Consumer-grade USB audio/video capture device (Reddo Videosieppari)

VIDEO ANALYSIS AND MANIPULATION

Conventional surveillance systems are already installed in many areas such as traffic surveillance and security relevant scenarios.

Streaming live video contribution to visual surveillance is in the area of image sequence analysis focusing on the topics:

- Motion detection
- Object tracking
- Scene analysis (understanding)

DRIVES
Development and Research on Innovative Vocational Education Skills

Motion Detection

Motion detection algorithms are the basis for a wide range of applications in computer vision like visual surveillance, object recognition and tracking and compression of video streams.

The most common approach for motion detection in surveillance systems with static cameras are the background subtraction algorithms. In these algorithms, a (moving) foreground object is detected by comparing the current image with the static background of the scene.

Motion Detection

The main challenge of background subtraction algorithms is the acquisition of this background image, since the background image might not be static but has to adapt to several changes as:

- Illumination changes
  - sudden changes (e.g., clouds, light-switch)
  - gradual changes (e.g., position of the sun changing during the day)
- Background motion (multimodal background)
  - e.g., waving trees, waves
- Changes in the background geometry
  - e.g., parking cars, moved items

Object Tracking

Object tracking is a correspondence problem and involves finding which object in a video frame is related to another object in the next frame. Tracking methods can be classified into four major categories:

- Model based tracking
- Active contour based tracking
- Feature based tracking
- Region based tracking.

DRIVES
Development and Research on Innovative Vocational Education Skills

Scene Analysis

The goal of a “Scene Analysis” algorithm is to recognise activities in scene. Recognition algorithms are mainly based on statistical analysis of the scene. Rule based approaches are applied to identify e.g. abnormal behaviour.

What do we want from video analysis?

- We want to detect all interesting objects in video.
- We want to identify their properties including human pose estimation, attributes estimation, person identification, etc.
- We want to recognize people actions and recognize events that are happening in the video.

VIDEO ANALYSIS AND MANIPULATION

DRIVES
Development and Research on Innovative Vocational Education Skills

An object appearance varies significantly between different viewpoints. For example, some surveillance cameras are mounted on people's height, so people are large and seen these high resolution. Other cameras are mounted on top of the building to overwiew the situation, so each person is seen as a very small dot.

Current image recognition and the detection algorithm **cannot** raise sufficient accuracy and speed simultaneously for both scenarios. So, applied video analysis systems are usually obtained only when algorithm are tailored to specific video scenario.

DRIVES
Development and Research on Innovative Vocational Education Skills

Manipulation

Video manipulation is a type of media manipulation that targets digital video using video processing and video editing techniques. Current video manipulation detection approaches depend on uncovering manipulations by studying pixel domain anomalies.

Improvements in the machine learning field have allowed the creation of powerful video manipulation tools, such as Face2Face, Recycle-GAN, Deepfakes, and other face swapping techniques.

Manipulation

According to *The Washington Post*'s there are six forms of video manipulation:

- Missing context
  - *Misrepresentation:* Setting original video footage into an incorrect context to misinform the audience.
  - *Isolation:* Publishing a short segment from a video.
- Deceptive editing
  - *Omission:* Removing important segments from a video to introduce a different story.
  - *Splicing:* Matching segments from different videos to create a narrative not supported by any of the individual videos.
- Malicious transformation
  - *Doctoring:* Directly changing video frames.
  - *Fabrication:* Using technology to construct bogus videos, like deepfakes.

SUMMARY

- **Video** is a series of **images** played in sequence at a specified frame rate, usually from software playing a video file (digital) or hardware playing a tape (analog).
- **Video capture** can be described as the process of converting an analog video signal to digital video and sending it to local storage or to external circuitry.
- **Video analysis** refers to detect all interesting objects in video, identify their properties including human pose estimation, and recognize people's actions and recognize events that are happening in the video.
- **Video manipulation** is a type of media manipulation that targets digital video using video processing and video editing techniques.

REFERENCES

Computer Vision and Deep Learning: From Image to Video Analysis | by Jeremy Cohen | Heartbeat. (n.d.). Retrieved July 8, 2021, from https://heartbeat.fritz.ai/computer-vision-from-image-to-video-analysis-d1339cf23961

Convergence Culture. Where Old and New Media Collide, Buying Into American Idol, Henry Jenkins, 2006 New York University Press.

Introduction to video analysis - Object tracking and action recognition | Coursera. (n.d.). Retrieved July 8, 2021, from https://www.coursera.org/lecture/deep-learning-in-computer-vision/introduction-to-video-analysis-alApg

Janai, J., Güney, F., Behl, A., & Geiger, A. (2020). Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art. *Foundations and Trends® in Computer Graphics and Vision*, *12*(1–3), 1–308. https://doi.org/10.1561/0600000079

REFERENCES

Pea, Roy (2006). "Chapter 55: Video-as Data and Digital Video Manipulation Techniques for Transforming Learning Sciences Research, Education, and Other Cultural Practices". In Weiss, J.; et al. (eds.). *The International Handbook of Virtual Learning Environments. Springer. pp. 100–110.*

"Seeing Isn't Believing: The Fact Checker's guide to manipulated video". *The Washington Post*. June 25, 2019.

Uera, D. G. ¨, Baireddy, S., Bestagini, P., Tubaro, S., & Delp, E. J. (n.d.). *We Need No Pixels: Video Manipulation Detection Using Stream Descriptors*. Retrieved from https://github.com/dguera/fake-video-

Video Analysis | *Computer Vision Lab*. (n.d.). Retrieved from https://cvl.tuwien.ac.at/research-areas/video-analysis/

REFERENCE TO AUTHORS

Regina Sousa

- PhD student in Biomedical Engineering
- Research Collaborator of the Algoritmi Research Center

0000-0002-2988-196X

Ana Luísa Sousa

- PhD student in Information System and Tecnologies
- Research Collaborator of the Algoritmi Research Center

0000-0001-5731-3583

Diana Ferreira

- PhD student in Biomedical Engineering
- Research Collaborator of the Algoritmi Research Center

0000-0003-2326-2153

REFERENCE TO AUTHORS

António Abelha

- Assistant Professor at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0001-6457-0756

José Machado

- Associate Professor with Habilitation at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0003-4121-6169

Victor Alves

- Assistant Professor at the University of Minho
- Integrated Researcher of the Algoritmi Research Center

0000-0003-1819-7051

REFERENCE TO AUTHORS

This Training Material has been certified according to the rules of **ECQA – European Certification and Qualification Association.**

The Training Material was developed within the international job role committee **“Computer Vision Expert”:**

UMINHO – University of Minho (https://www.uminho.pt/PT)

The development of the training material was partly funded by the EU under Blueprint Project DRIVES.

DRIVES

Development and Research on Innovative Vocational Education Skills

Thank you for your attention

DRIVES project is project under **The Blueprint for Sectoral Cooperation on Skills in Automotive Sector**, as part of New Skills Agenda.

The aim of the Blueprint is **to support an overall sectoral strategy and to develop concrete actions to address short and medium term skills needs.**

Follow DRIVES project at:

More information at:

www.project-drives.eu

Co-funded by the Erasmus+ Programme of the European Union

The Development and Research on Innovative Vocational Educational Skills project (DRIVES) is co-funded by the Erasmus+ Programme of the European Union under the agreement 591988-EPP-1-2017-1-CZ-EPPKA2-SSA-B. The European Commission support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.