NTNU | Norwegian University of Science and Technology

AN INTRODUCTION TO MULTIMEDIA FORENSICS

Audio, Image and Video

Jens-Petter Sandvik

2021–11–12

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

- Outline

Introduction

Metadata

Audio forensics

Image forensics

Video forensics

Introduction - What is this about?

- Multimedia content is found everywhere
- “Everyone” has recording equipment
- People tend to trust something we see more than something we read
- Multimedia:
  - Audio
  - Image
  - Video
- Data related to the multimedia content
  - Metadata
- ENFSI has general and best practices guides on their webpage: `https://enfsi.eu/`

Introduction - Learning outcomes

- Understand some of the possibilities and challenges in multimedia forensics
- Understand metadata
- Know the process that craetes multimedia and the artifacts that is created
- Understand a few methods that are used in multimedia forensics
- What makes deepfakes “deep”, and how to detect it

Metadata - What is *metadata*?

- “Data about data”
- A description of the content
- Parameters for playing the content
- Description of equipment used for creating content
- Metadata can be found many places
  - In media file
  - Text files
  - In databases
  - Other archives?

Metadata - File system data

- Data about the file itself
- File name
- MAC times
  - Modified, Last accessed, Created/Metadata changed
  - But contemporary operating systems don't update Last accessed times
  - Created is mostly updated to the time the file is created in the file system
  - Modified often survives when unpacked from a zip archive
- Username of owner
- Access rights to file

Metadata - Container file data

- Multimedia file typically contains:
  - Content streams: Video and audio content
  - Information about the content
- EXIF, MP3tags, etc.
- Written by creator and editing processes
- . . . but can also be modified by others

Metadata - EXIF

- Exchangeable image file format
- Set by the camera or image creation program
- Can be updated by other programs
- Includes information about the equipment
- Sometimes also GPS coordinates
- Many programs can print the EXIF data
  - exiv2, exiftool, etc.

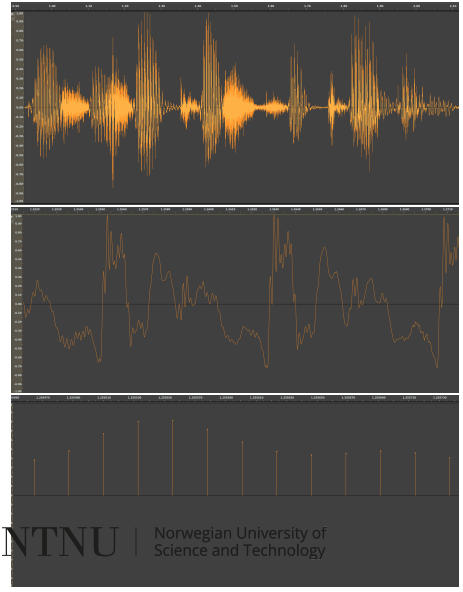
Metadata - EXIF example

```
$ exiftool 20211011/20211011-1806-S10-5851.jpg
ExifTool Version Number         : 12.16
File Name                       : 20211011-1806-S10-5851.jpg
...
File Size                       : 2.4 MiB
File Modification Date/Time     : 2021:10:11 18:06:48+02:00
File Access Date/Time           : 2021:11:04 18:14:17+01:00
File Inode Change Date/Time     : 2021:10:17 22:31:22+02:00
...
Make                            : samsung
Camera Model Name               : SM-G973F
...
GPS Latitude                    : 59 deg 54’ 31.76" N
GPS Longitude                   : 10 deg 48’ 44.12" E
...
```

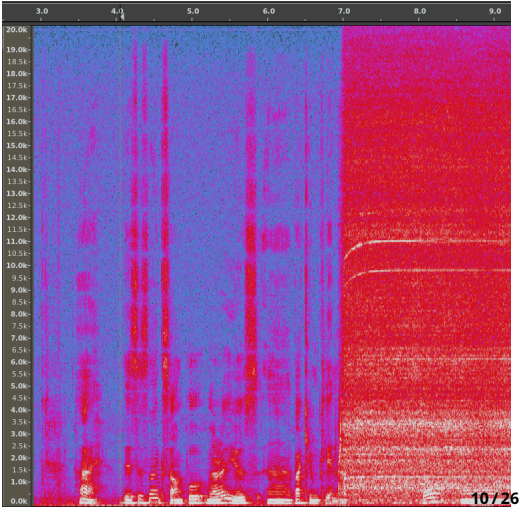
Audio forensics - Audio fundamentals

- Sound waves are pressure waves in a medium (air, solid materials)
- The pressure differences over time is the sound pressure
  - Measured with regard to a reference pressure: dB
- Frequency is the number of pressure tops/bottoms per second
  - Measured in Hz
  - A complex wave can consist of several waves, each with different frequencies
- A microphone convert the sound waves to electrical waves
  - Has a *frequency response* — different sensitivity for different frequencies
- Analogue to Digital Conversion (ADC) introduces noise to the process
- Lossy compression of digital signal also introduces artifacts
  - Lossy: mp3, aac; Lossless: wav, flac

Audio forensics - Visualizing sound



← Waveform ↓ Spectrogram



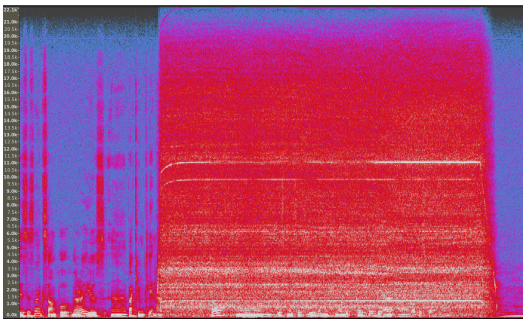
NTNU | Norwegian University of Science and Technology

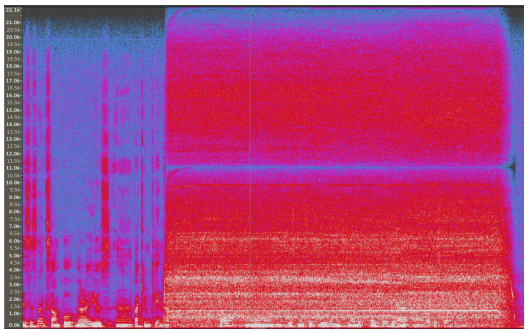
Audio forensics - Cleaning

- Remove noise or other sounds to enhance the sound of interest
- Want to better understand what is happening or being said in the redording
- For speech, a **high risk of bias** when interpreting result
  - We tend to hear what we expect to hear
- Mosty a *subtractive* action
- Remove frequencies that contain noise components
  - But keep most of the speech components
  - Works fairly well for a static noise component
- Be aware that removing parts of the spectrum can make words sound differently
  - e.g. sh $\rightarrow$ s, sharp sounds becoming more “muffled”, etc.

Audio forensics - Notch filter example

- Notch filter will remove only a small range of frequencies
- Other main type of filters are band-pass and -stop filters, high- and low-pass/ -stop filters
- Below is speech interrupted by a vacuum cleaner, to the left using a notch filter for one of the major noise frequencies:



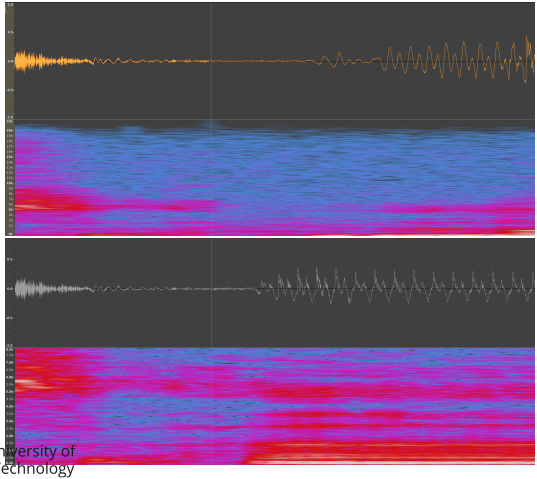


Audio forensics - Authenticity

- Authenticity of a recording is to determine whether:
  - The recording is in its original state
  - The recording is of the described event
- Finding edited areas
  - Cuts, splices
  - Find abrupt changes in the content
  - Noise suddenly changing characteristics
- Does metadata match the content?
- Lack of evidence of manipulation does not guarantee authenticity

Audio forensics - Splicing example

- The upper image is spliced, the bottom is the original
- Note the change in noise at the splice



Audio forensics - Doppler effect

- Sound waves generated while moving will be compressed in the direction of movement and stretched in the opposite direction
- The same is true for a static audio source and a moving recorder
- Compressed waves means a higher frequency, or pitch
- Stretched waves means a lower frequency, or pitch
- Example of the Doppler effect: Sirens passing by, train horn as it passes.

Image forensics - Capture process

- Light enters through lenses — focuses image on sensor
- Color Filter Array (CFA) — each pixel only see one color component
- Sensor — Transform photons to electric current and digitizes the current
- Camera processor:
  - Demosaic — four color pixels from the CFA to one pixel with three colors
  - In-cam processing — Color/ white balance, contrast, saturation adjustments
  - Image encoding — JPEG compression
- Post-processing of image
- Editing

Image forensics - Capture process

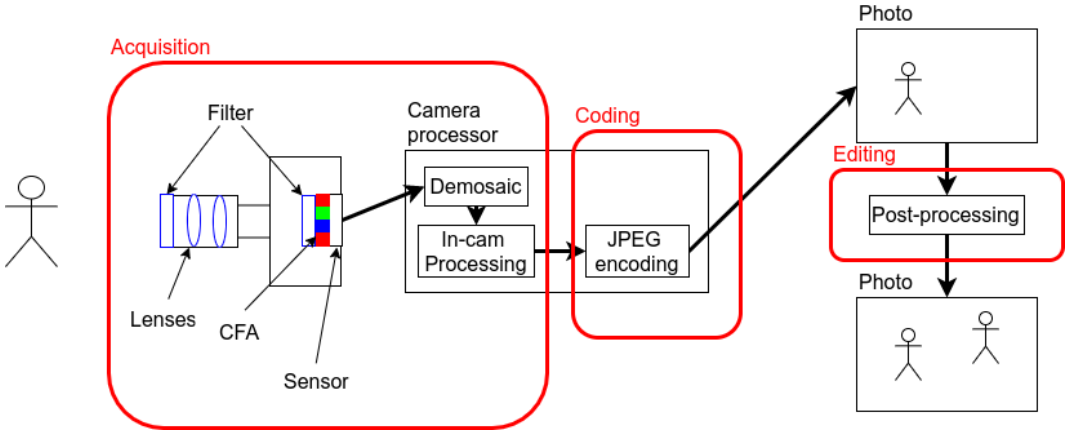


Image forensics - Photogrammetry

- Measure angles, distances, sizes in photos
- Mapping from 2D to 3D space
  - Trigonometry
  - Compare to objects with known sizes
- Need to know the effects lenses have on the photo: Optical distortion
  - Straight lines curving in the photo
  - Barrel distortion — lines curving away from the center
  - Pin-cushion distortion — lines curving toward the center
  - Moustache distortion, a combination of barrel and pin-cushion distortion
  - Photo editing programs often have filters to adjust optical distortions
- Perspective distortion
  - Wide-angle distortion — Objects closer to the camera appear bigger
  - Compression distortion — Objects further away appear bigger, closer

Image forensics - Distortion from wide angle lens

- A grid notebook page, wide angle lens from a phone
- The example shows a pin-cushion distortion
  - Can be from in-camera lens correction

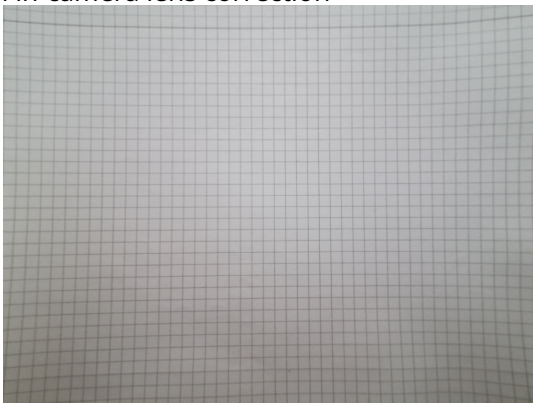


Image forensics - Editing detection

- Malicious editing operations change the perceived meaning of the image
  - Copy part of an image to another location in the same image
  - Copy part of another image into the image
  - Remove part of an image, change perspective, etc.
- Analysis of the encoded data
  - Anomalies in blocking of JPEG images
  - Error Level Analysis
  - Anomalies in histogram of JPEG DCT coefficients
- Analysis of the scene
  - Lighting/shadow anomalies
  - Detection of similar areas in the photo
  - Perspective anomalies

Image forensics - Equipment identification

- Each sensor consist of millions of pixels, each have slight variations due to production imprecision
- Photoresponse Non-uniformity (PRNU)
  - Unique for each photosensor
  - The PRNU can be suppressed by strong compression
  - Experiments show that this is stable over the lifetime of the chip
  - Exist python libraries for extracting and comparing the PRNU
- Color Filter Array / demosaicing artifacts
  - Don't uniquely identify device
  - Identify type of device, camera model

Image forensics - Deepfakes

- Term from AI — Deep learning
- Most generators today use Generative Adversial Networks (GAN)
  - One deep learning module generate images
  - The other tries to detect which is generated
  - Result fed back to generator that tries to improve the generated image
  - Many iterations
- Sometimes the generated image have details that don't make sense for a human eye
- Often lack PRNU, but this can be synthetically created (if implemented in generator)
- Machine learning detection
  - By adjusting GAN or compressing image: detection rate drops
  - Don't trust AI/ML detection methods more than at an advisory level

Video forensics - Different than images?

- One image per frame plus audio
- Videos are typically more compressed than images
  - Removes PRNU, demosaicing artifacts
- Less standardized, more configurable encoding steps
  - Also a temporal component to the encoding and compression
  - Parameters for encoding can be used for identifying models of equipment
- Harder to hide evidence of editing operations, as every frame need to be undetectable
- Some operations
  - Remove noise, encoding and compression artifacts
  - Find editing operation such as greenscreen
  - Detect deepfakes

Video forensics - Audio/ video correlation

- Speed of light is different from the speed of sound
  - 299 792 458 m/s vs. 343 m/s (at 20 °C, dry air)
- One second difference between visual source of sound and audible sound means that the event was 343 meters from the camera
- Can be hard to know exactly when a visible event generates the sound
- Have to find the offset between video and audio from close events
  - Check that the audio/video offset is stable throughout the video

Video forensics - Deepfakes

- Videos can be generated from scratch, but this is resource demanding
- Add a face to the body of someone else
  - Face swapping apps
- Make a person say something different
  - Change the audio to something else
  - Change the face to give new expression, mouth movements to match the audio
  - Can also generate deepfake audio
- Video deepfakes are often easier to detect, as the generation is harder and leaves more anomalies
  - E.g. blinking, eye movements, but many deepfake generators implement this now
- Anomalies can be hidden by harder compression
- A search on Youtube on “deepfake” shows many examples of face swaps

Video forensics - Other sources

- Also use external sources:
  - OSINT
  - Interpretation of the recorded scene (audio, photo, video)
- Does the content fit the broader picture, or is it inconsistencies between the content and the broader context?
- As deepfake technology gets better and more accessible, this will be used for all types of information
  - Elections
  - Polarized topics
  - +++

Thank you for your attention