Forensics Investigations of Multimedia Data

A Review of the State-of-the-Art

Rainer Poisel (rpoisel)
rainer.poisel@fhstp.ac.at
Simon Tjoa (stjoa)
simon.tjoa@fhstp.ac.at

IMF Conference 2011

Agenda of this presentation

Organizational

- Organizational
  - Motivation
  - Related Work
- Fields of forensics analysis and multimedia files
- Conclusion and Discussion

Agenda of this presentation

Organizational

- Covered fields
  - Source identification
  - Environment classification
  - Content classification
  - Content forgery
  - Data recovery
  - Steganography and steganalysis
  - Standardization

Motivation, Related Work

Organizational

- General motivation
  - 80 – 90% of cases today involve digital evidence
  - Amount of data steadily increases
- Own motivation
  - Identification of future research areas

- Related Work
  - Surveys on digital images
  - Discussion of terminology
    - Multimedia forensics vs Computer forensics

Source Identification

- Comparable to gun identification
  - Bullets leave scratches
  - So do recording devices with media

Digital cameras

Source Identification

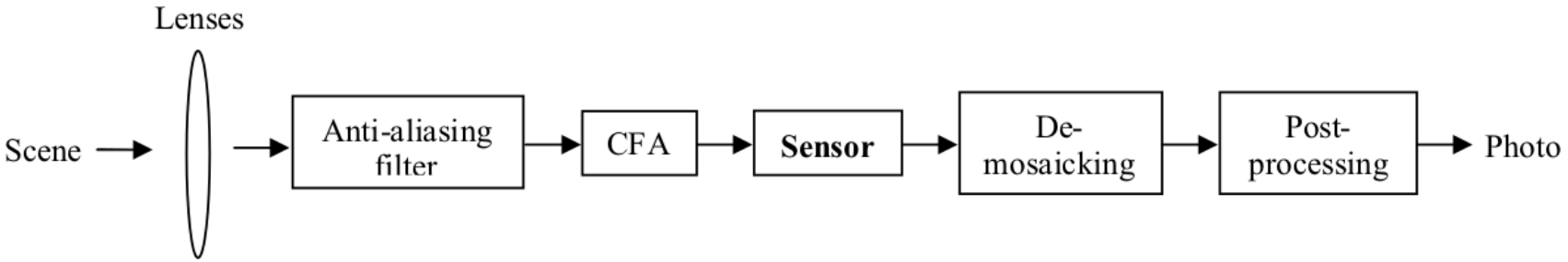
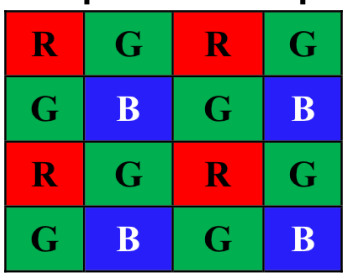


Image acquisition process[1]



Bayer Color Filter Array (CFA)[2]

[1] Source: Source Camera Linking Using Enhanced Sensor Pattern Noise, Li et al., 2009

[2] Source: Digital Camera Identification Using Colour-Decoupled Photo Response Non-Uniformity Noise Pattern, Li et al., 2010

Digital cameras

Source Identification

- Features related to digital cameras
  - Peculiarities of JPEG compression
  - Color Filter Array (CFA)
  - Sensor Pattern Noise (SPN)
- Photo Response Non-Uniformity (PRNU) Noise
  - Contamination of SPN with details from scenes
  - Details of the scene attenuated
- Colour Decoupled PRNU (CD-PRNU) considers CFA

Scanners

Source Identification

- Digital cameras reproduce natural scenes
- Scanners capture hard-copy media
- Special lighting conditions
- Moment-based features are extracted
  - Image denoising
  - Wavelet analysis
  - Neighborhood detection
- Approaches suitable for the detection of content forgery

Video cameras

Source Identification

- Application of techniques known from the picture- and audio-domain
- Better results than identification of cameras from still images
- Pretty effective

Environment Classification

- Determination recording properties
  - Location
  - Local conditions

Visual Data

Environment Classification

- Two subgroups
  - Event recognition
    - Detection of objects displayed in an image
    - Detection of environment terrain
- Place instance recognition
  - Detection of a specific place
- Place category recognition
  - Detection of the “place-type”
- Usage of the context to improve accuracy

Audio Data

Environment Classification

- Close relation between environment classification and source identification
- Approaches known from steganalysis
- Analysis of the Electrical Network Frequency (ENF)

Video Data

Environment Classification

- Techniques from the fields of visual data or audio data
- Combinations of both

Content classification

- Automated classification of data collections
  - Material from surveillance cameras
  - Evidence from financial crime
  - Pornography

Digital Images

Content classification

- Considering both file names and image analysis
- Location of skin regions (shapes) in an image
- Bag-of-visual-words (BOVW)
  - Pictures built from discrete visual words
  - Training phase
  - Histogram and filtering

Video Data

Content classification

- Analysis of keyframes and motion
  - Keyframes
    - Skin regions
    - Bag-of-visual-words
  - Motion
    - Periodicity Detection
    - Sliding Window Periodicity
    - Motion Histograms

Content Forgery

- Adding, removing or changing important features
- Digital data is easy to manipulate
- Detection of malicious manipulation

Visual Data

Content Forgery

Altered version of the Houston crisis room[3]

[3]Source: http://www.americablog.com/2010/07/bp-photoshops-fake-photo-of-command.html

Visual Data

Content Forgery

Blowup of the previous picture[4]

[4]Source: http://3.bp.blogspot.com/_1xQeOPE9ePU/TEUNdvgNqmI/AAAAAAAAFDM/i_zXzIWKpPk/s1600/bpblowup.jpg

Visual Data

Content Forgery

Image manipulation and detection[5]

[5]Source: Multimedia Forensics is not Computer Forensics, R. Böhme et al., 2009

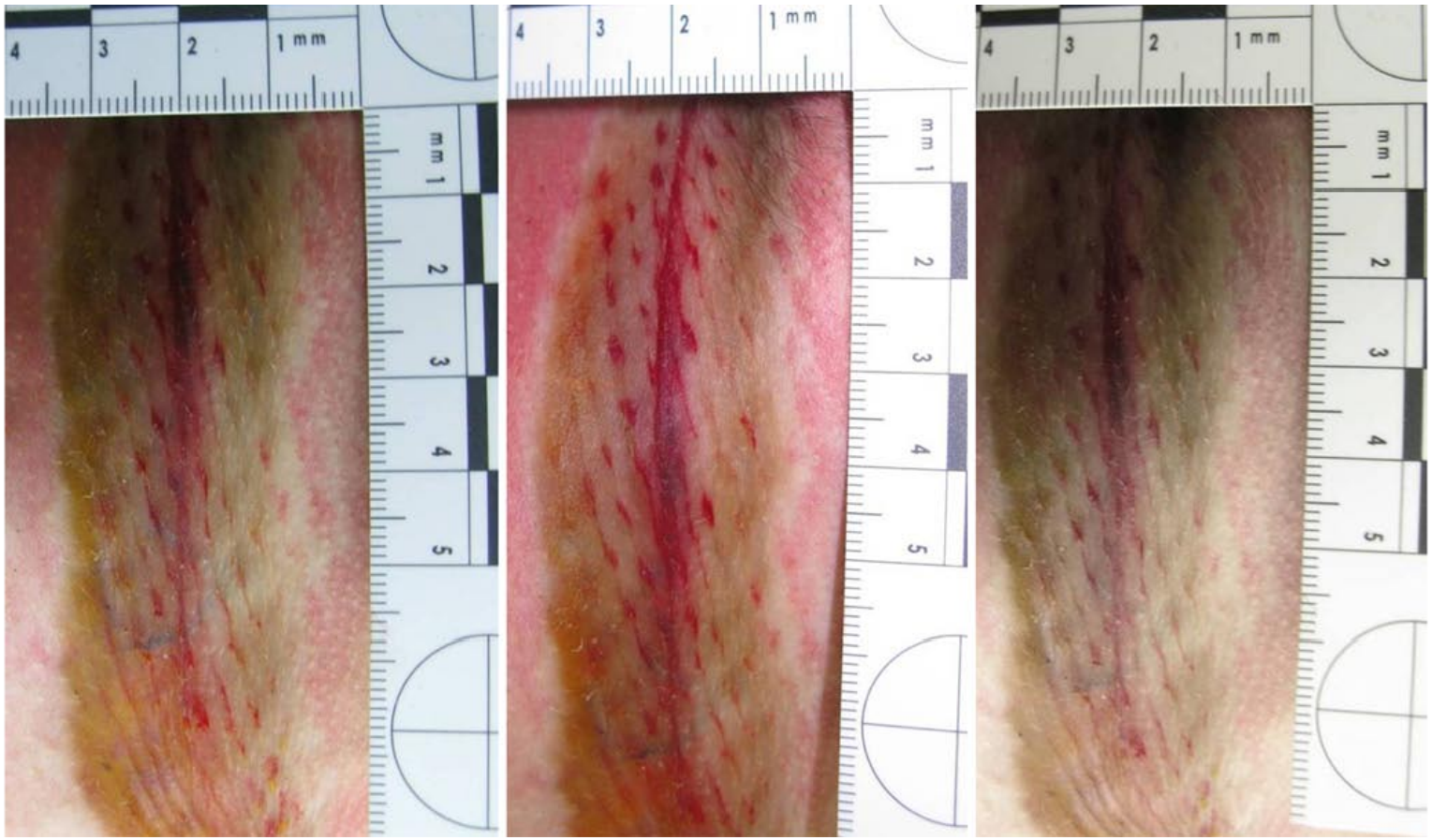
Visual Data

Content Forgery

- Copy-move forgery
- Retouching
- Filtering
- Partial deletion
- Mounting and merging
- Manipulation of geometry
- Manipulation of luminance

Visual Data

Content Forgery



Manipulation of color space[6]

[6]Source: Original oder manipuliert?, F. Ramsthaler et al., Springer, 2010

Visual Data

Content Forgery

Classification based on complexity

- Low Level
  - Statistical investigations
  - DCT coefficients, Sensor Pattern Noise (SPN)
- Middle Level
  - Simple semantic information
  - Lighting direction
  - Sharp edges and blurred areas
- High Level
  - Purely semantic tampering
  - Identification of characters and objects

Visual Data

Content Forgery

Example for “High level”[7]

[7] Source: http://www.prisonplanet.com/images/february2006/280206bushbinladen.jpg

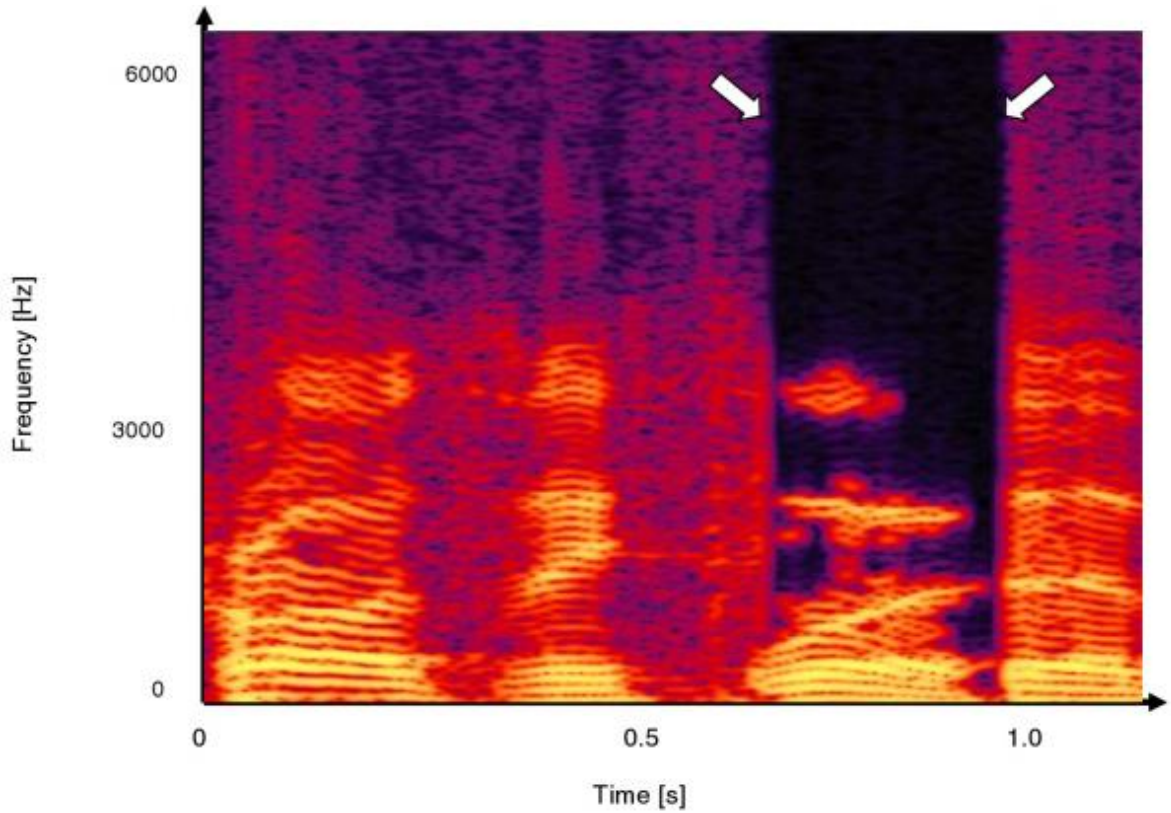
Auditive Data

Content Forgery

- Several tests
  - Visual, physical, electrical and acoustical
- Properly used recording devices
- Integrity verification of the recording medium
- Critical listening
- Checks for continuous operation
- Usage of analytical tools

Auditive Data

Content Forgery



Likely alteration in a spectogram of a speech recording[8]

[8] Source: Overview of Audio Forensics, Robert C. Maher, Springer, 2010

Video Data

Content Forgery

- Similar to the detection of image data manipulations
- Methods of different complexity
  - Low Level: Detection of artifacts and noise characteristics
  - Middle Level: Detection of duplicated frames and regions
  - High Level: Detection of persons and objects
- Methods known from auditive domain

Data Recovery

- File recovery from various storage media
  - Tradition approaches using filesystem information
  - Content-based recovery of data
- Field closely related to Computer forensics

File Carving

Data Recovery

- What does the term "'‘File Carving"' mean?
  - *“File carving is a forensics technique that recovers files based merely on file structure and content and without any matching file system meta-data.”* – Anandabrata Pal
- What is it good for?
  - Recovery of files
    - based on their structure only
    - with unknown file-system structures
    - with manipulated/deleted file-system metadata
- What is the typical usage context?
  - Can be applied to any storage medium
  - Digital forensics and general data recovery

Potential areas

Data Recovery

- Potential areas containing fragments of files
  - Deleted Files
  - Clusters that have been marked as corrupt
  - Host Protected Area (HPA)
  - Device Configuration Overlay (DCO)
  - Unallocated Areas
  - Partition Slack

File Carving Approaches

Data Recovery

- Techniques for unfragmented files
  - Header/footer carving
  - Header/maximum size carving
  - Header/embedded length carving
  - File trimming

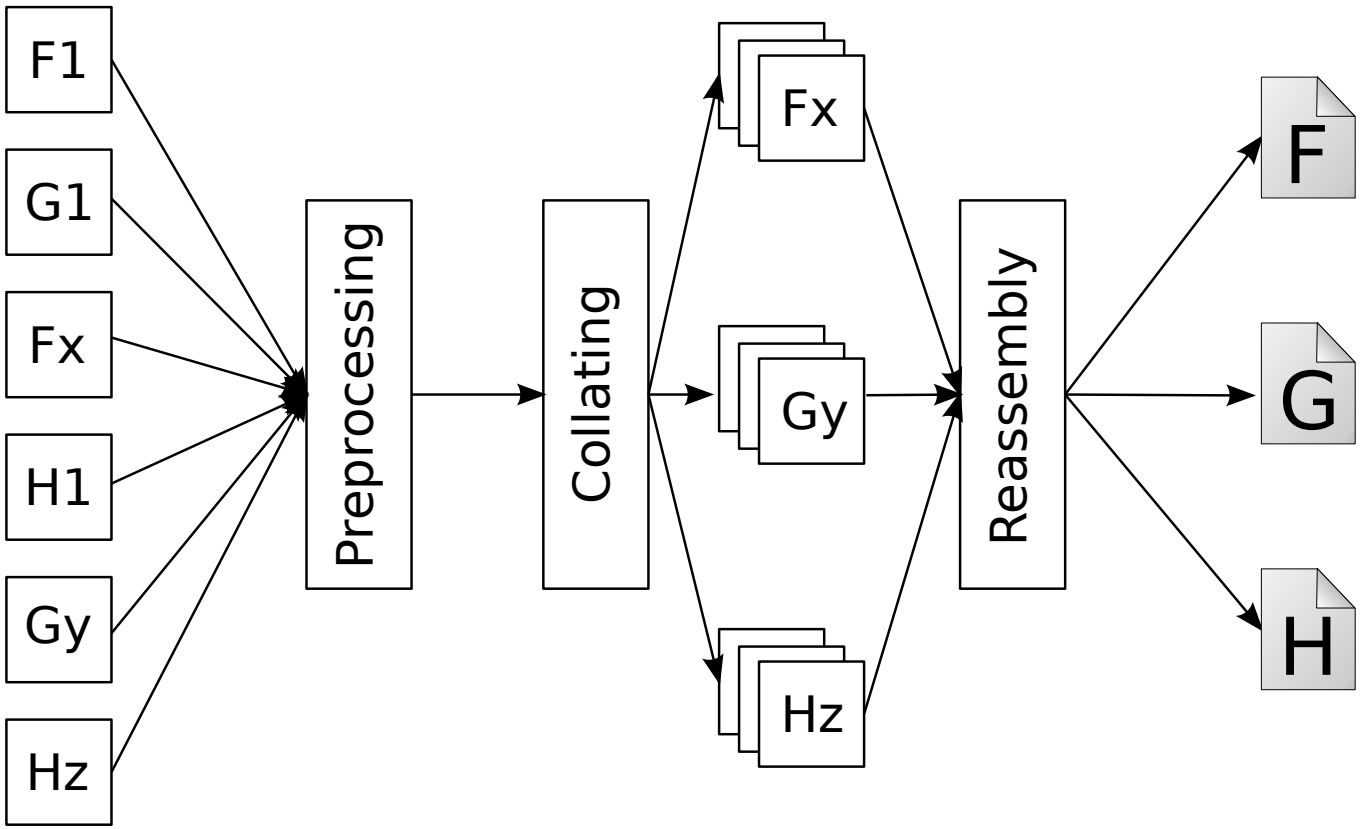
File Carving Approaches

Data Recovery

- Techniques for fragmented files
  - Bifragment gap carving
  - Graph-theoretic carvers (Parallel Unique Path, PUP)
  - Sequential Hypothesis Test (SHT-PUP)
  - Approaches that regard the recovered file format
  - Smartcarving

Smartcarving

Data Recovery



Smartcarving Architecture[9]

[9] Source: The evolution of file carving, Pal et al., 2009

Fragment Classification

Data Recovery

- Essential step when finding parts of a whole file
- Techniques
  - Magic numbers
  - Statistical approaches
  - Normalized Compression Distance (NCD)
  - Specific approaches

Steganography and Steganalysis

- *Steganography is an ancient art of embedding private messages in seemingly innocuous messages in such a way that prevents the detection of the secret messages by a third party.* – Richard Popa
- *Steganalysis deals with the detection of embedded information.* – Huaiqing Wang and Shuozhong Wang

Steganography: Auditive and Visual Data

Steganography and Steganalysis

- Digital image steganography
  - Spatial domain
  - Frequency domain
- Transfer to auditive domain is possible

Steganography: Auditive and Visual Data

Steganography and Steganalysis

- Techniques
  - Low-bit coding or Least Significant Bit Hiding
  - Echo Hiding
  - Phase coding
  - Spread spectrum
  - Combinations

Steganalysis: Auditive and Visual Data

Steganography and Steganalysis

- Techniques
  - Signature based
  - Statistical

Research considerations

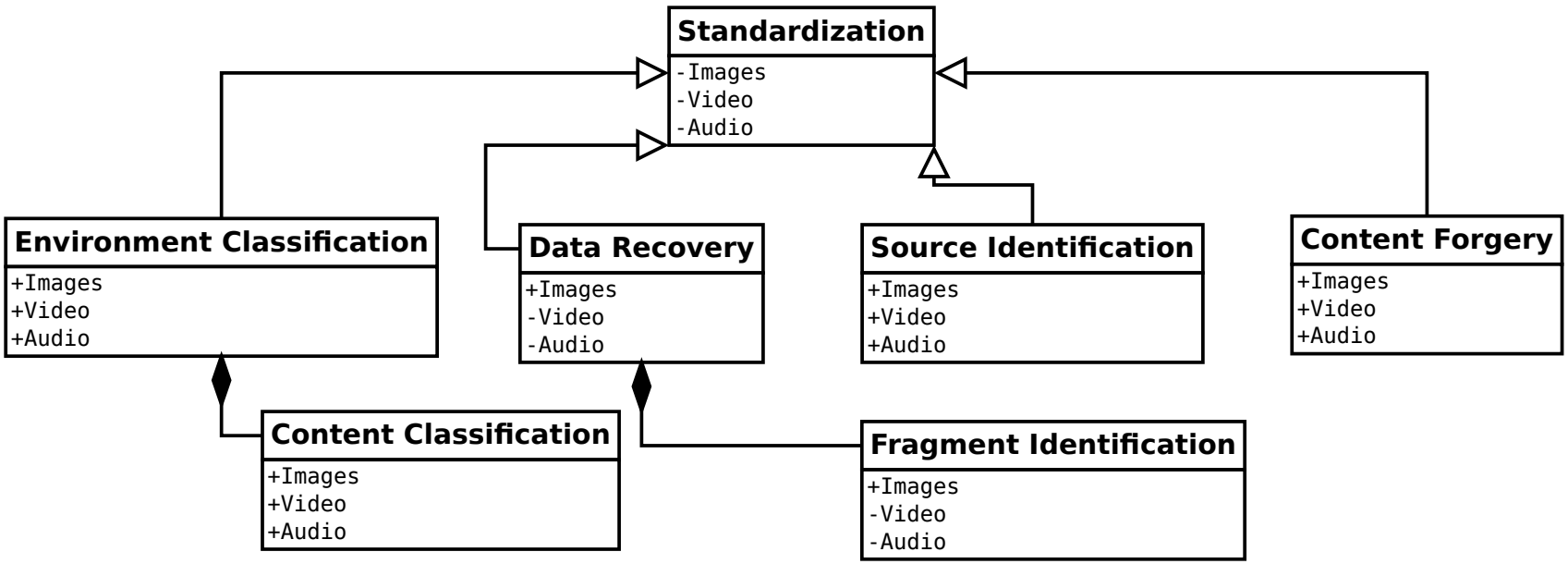
Standardization

- Standardization: addressing of proven methods, procedures or algorithms
  - Paper standards
  - Material standards
- Research: exploration of new approaches
- Nevertheless: standardized procedures needed
- Advances for common file formats, schemas and ontologies
- “Hyper-formalized” processes and approaches
- But no single universal standard

Research considerations

- Multimedia/Computer forensics
  - Video file carving
- FREDI: FRamEwork for Digital Investigations
- Localization of WLAN Access-Points

Conclusion and Discussion



Conclusion and Discussion

Q? & A!

Thank you for your attention!