Residual-based forensic comparison of video sequences

[1]**Patrick Mullan**,
[2]Davide Cozzolino,
[2]Luisa Verdoliva,
[1]Christian Riess

[1] FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

[2] UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

Increased prevalence of video content

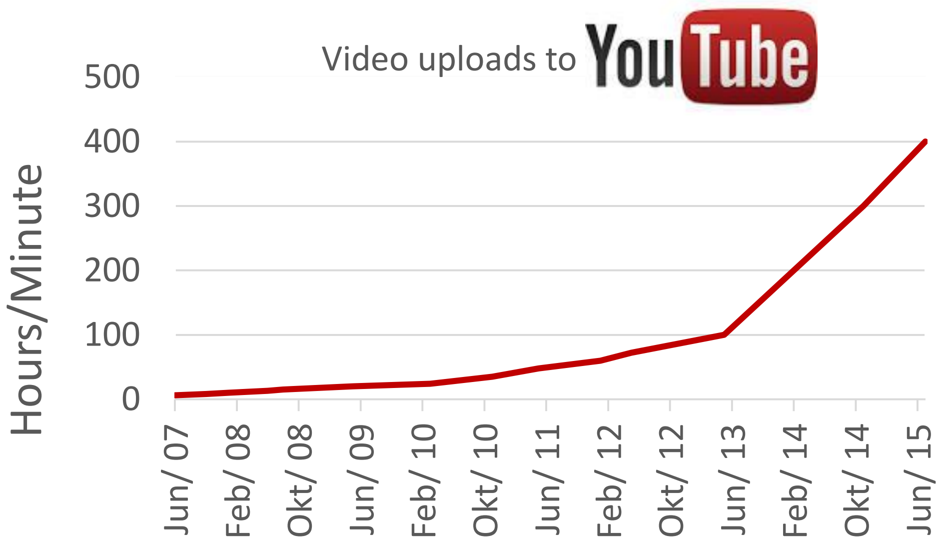
Creator’s side

- Visual content simpler to create and share than ever before
- Easy-to-use tools for editing videos are already widely present

Increased prevalence of video content

Creator's side

- Visual content simpler to create and share than ever before
- Easy-to-use tools for editing videos are already widely present



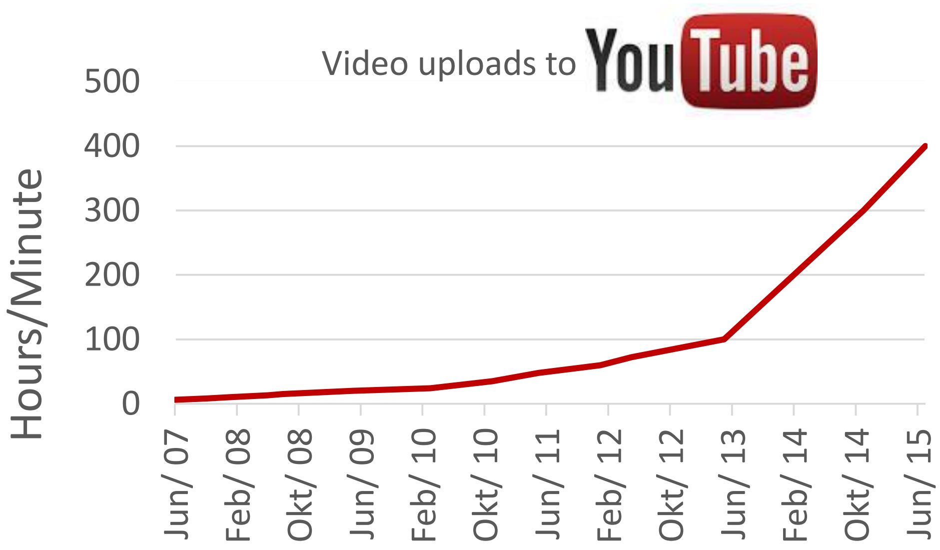
Increased prevalence of video content

Creator's side

- Visual content simpler to create and share than ever before
- Easy-to-use tools for editing videos are already widely present

Analyst's side

- Some content is altered with malicious intents
- Few tools exist to automatically assess authenticity of video data



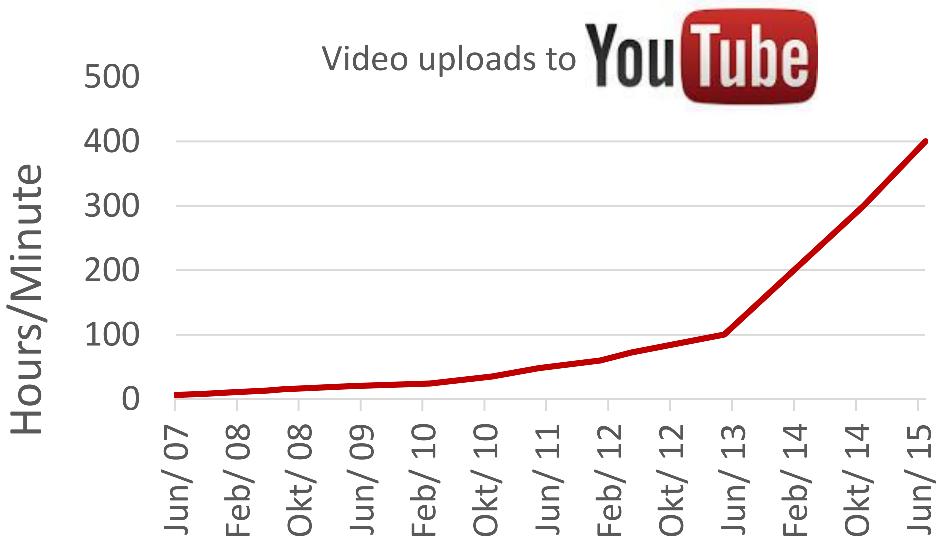
Increased prevalence of video content

Creator's side

- Visual content simpler to create and share than ever before
- Easy-to-use tools for editing videos are already widely present

Analyst's side

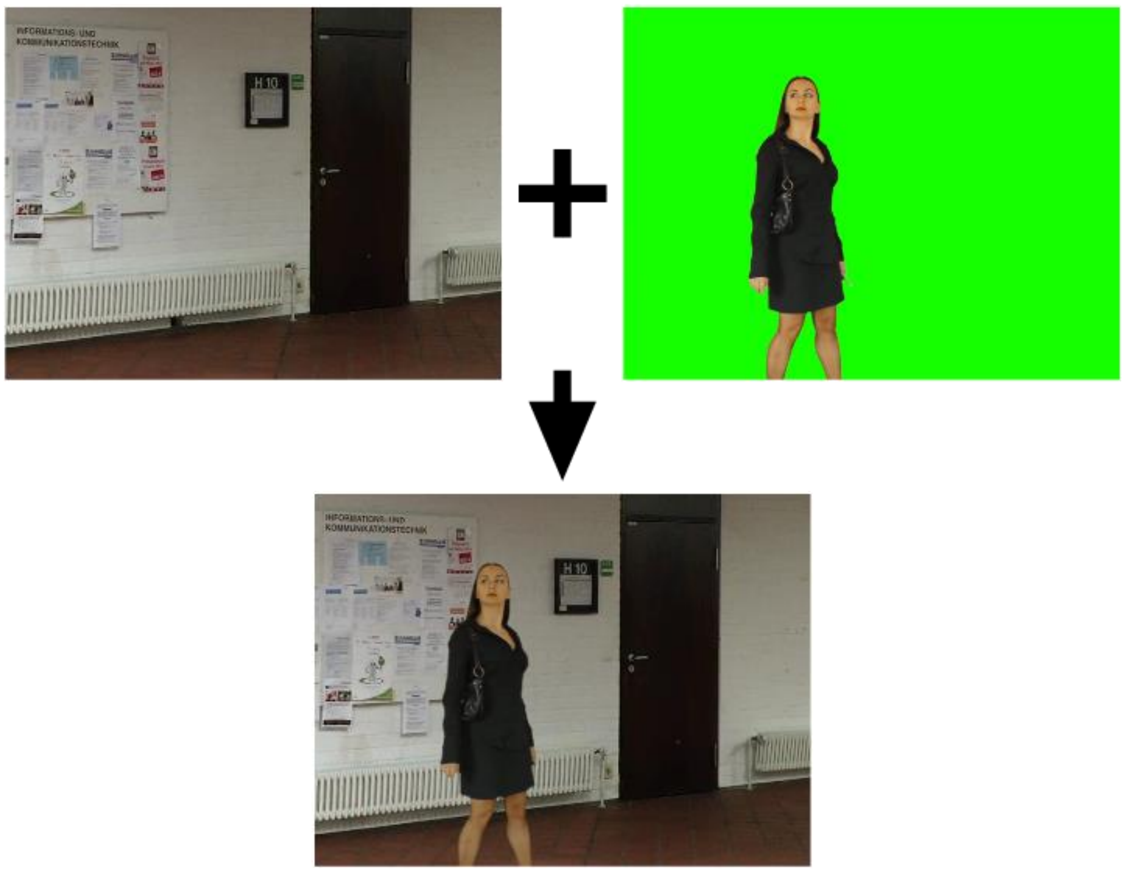
- Some content is altered with malicious intents
- Few tools exist to automatically assess authenticity of video data



source: mediathek.zdf.de

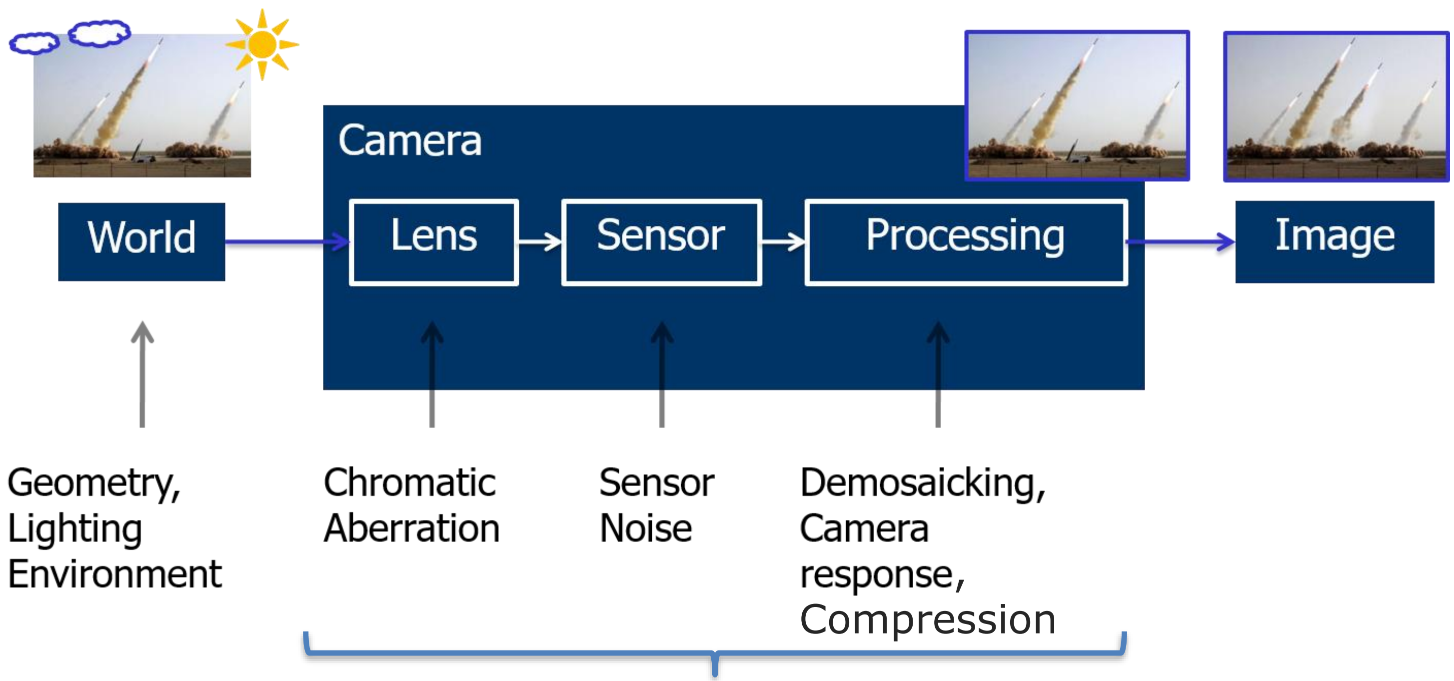
Chroma keying

- One manipulation attack is chroma keying (e.g. greenscreening)
- If done well, forged video offers no visual clues on manipulation



Assumption

- Each camera has its own, unique, processing pipeline
- They introduce characteristic, high frequent noise, in each frame and over frames
- Often not visually perceivable



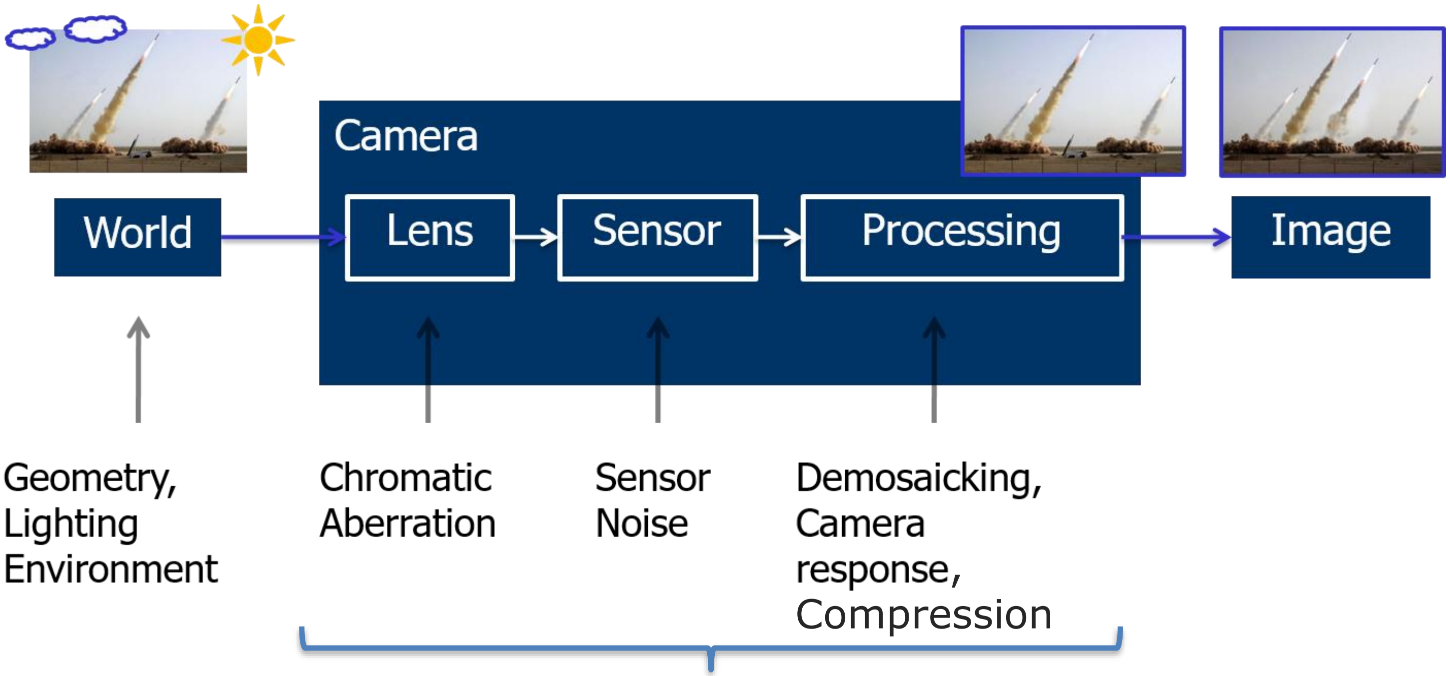
Camera unique components and settings/parameters

➢ Introducing unique high frequent noise

Assumption

- Each camera has its own, unique, processing pipeline
- They introduce characteristic, high frequent noise, in each frame and over frames
- Often not visually perceivable

➢ Manipulations break those statistics or make them inconsistent



Camera unique components and settings/parameters

➢ Introducing unique high frequent noise

Feature extraction from noise

Inconsistencies in noise patterns well exploited in different fields:
For example, in “steganography” [1] or “forgery detection in images” [2]

[1] J. Fridrich, J. Kodovský “Rich Models for Steganalysis of Digital Images”, in *IEEE Transactions on Information Forensics and Security*, June 2012
[2] D. Cozzolino, G. Poggi, L. Verdoliva, “Splicebuster: A new blind image splicing detector,” in *IEEE International Workshop on Information Forensics and Security*, Nov. 2015

Feature extraction from noise

Inconsistencies in noise patterns well exploited in different fields:
For example, in “steganography” [1] or “forgery detection in images” [2]

Common algorithm:

1. High-pass filtering input image I, returning residual image R, where image I has pixels at $I_{xy} \in [0|255]$
   → retrieves noise domain

[1] J. Fridrich, J. Kodovský “Rich Models for Steganalysis of Digital Images”, in *IEEE Transactions on Information Forensics and Security*, June 2012
[2] D. Cozzolino, G. Poggi, L. Verdoliva, “Splicebuster: A new blind image splicing detector,” in *IEEE International Workshop on Information Forensics and Security*, Nov. 2015

Feature extraction from noise

Inconsistencies in noise patterns well exploited in different fields:
For example, in “steganography” [1] or “forgery detection in images” [2]

Common algorithm:

1. High-pass filtering input image I, returning residual image R, where image I has pixels at $I_{xy} \in [0|255]$
   $\rightarrow$ retrieves noise domain

2. Quantize and truncate: $R^*_{xy} = \min\{t, \max\{-t, round(\frac{R_{xy}}{q})\}$
   $\rightarrow$ large residuals (like edges) are all mapped to t or $-t$
   $\rightarrow$ the “interesting” coefficents lie between $[-t+1 \mid t-1]$

[1] J. Fridrich, J. Kodovský “Rich Models for Steganalysis of Digital Images”, in *IEEE Transactions on Information Forensics and Security*, June 2012
[2] D. Cozzolino, G. Poggi, L. Verdoliva, “Splicebuster: A new blind image splicing detector,” in *IEEE International Workshop on Information Forensics and Security*, Nov. 2015

Feature extraction from noise

Inconsistencies in noise patterns well exploited in different fields:
For example, in "steganography" [1] or "forgery detection in images" [2]

Common algorithm:

1. High-pass filtering input image I, returning residual image R, where image I has pixels at $I_{xy} \in [0|255]$
   → retrieves noise domain

2. Quantize and truncate: $R^*_{xy} = \min\{t, \max\{-t, round(\frac{R_{xy}}{q})\}$
   → large residuals (like edges) are all mapped to t or $-t$
   → the "interesting" coefficents lie between $[-t+1 \mid t-1]$

3. Build co-occurences of length d: $C_{nm} = \{R^*_{xy}, R^*_{xy+1}, \dots, R^*_{xy+d}\}$
   → incorporates neighborhood relationships

[1] J. Fridrich, J. Kodovský "Rich Models for Steganalysis of Digital Images", in *IEEE Transactions on Information Forensics and Security*, June 2012

[2] D. Cozzolino, G. Poggi, L. Verdoliva, "Splicebuster: A new blind image splicing detector," in *IEEE International Workshop on Information Forensics and Security*, Nov. 2015

Descriptors applied to image forensics

Grayscale input frame

Descriptors applied to image forensics

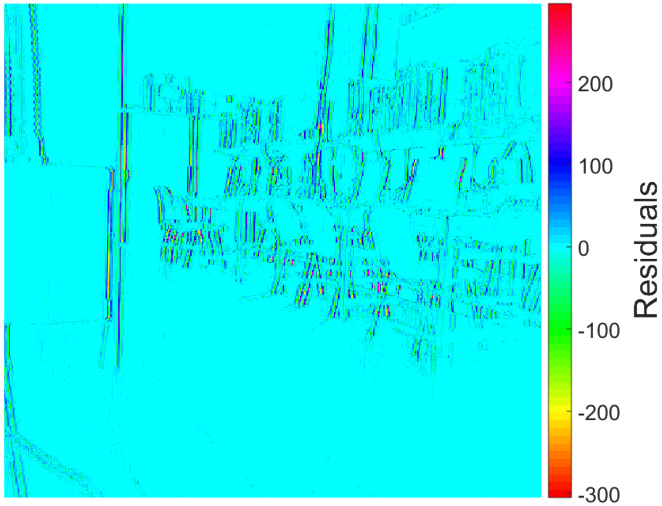
Grayscale input frame

High-pass
[1,-3,3,-1]

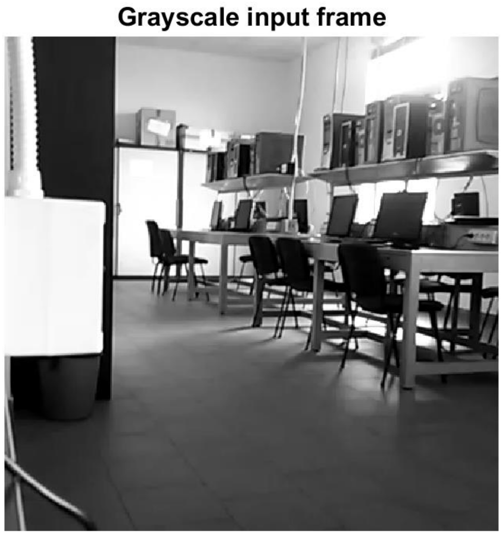
Descriptors applied to image forensics

Grayscale input frame

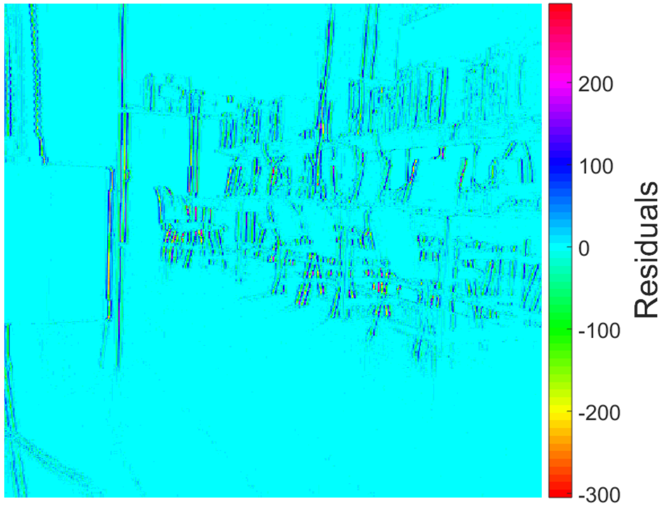
High-pass
[1,-3,3,-1]



Descriptors applied to image forensics

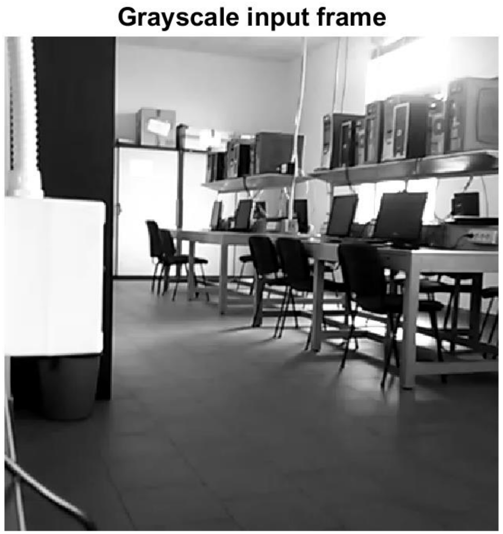


High-pass
[1,-3,3,-1]

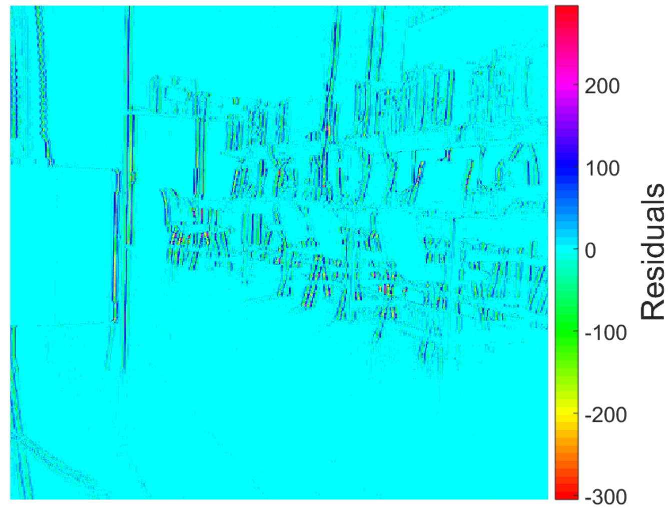


Quantize
& truncate

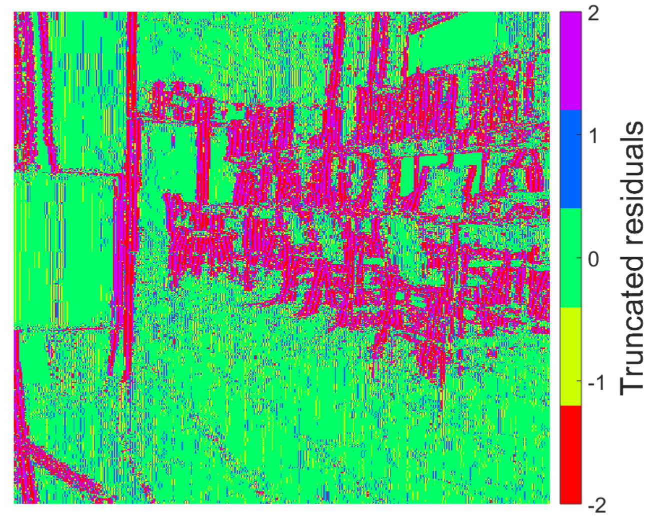
Descriptors applied to image forensics



High-pass
[1,-3,3,-1]



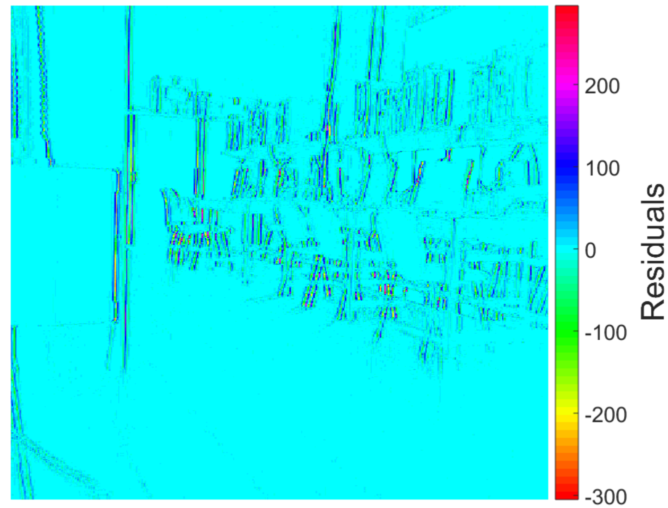
Quantize
& truncate



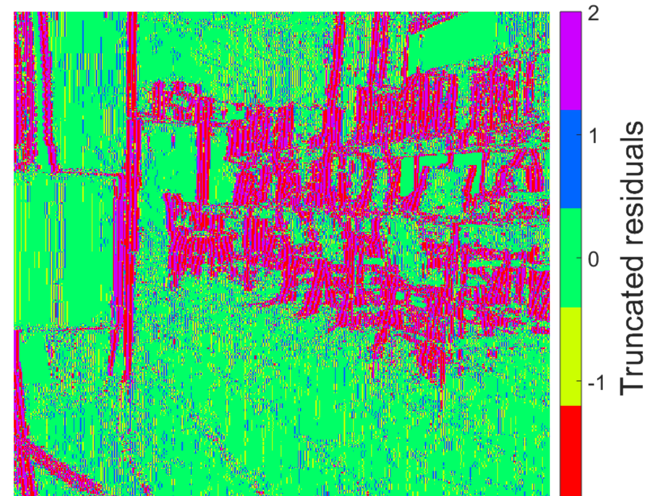
Descriptors applied to image forensics

Grayscale input frame

High-pass
[1,-3,3,-1]



Quantize
& truncate

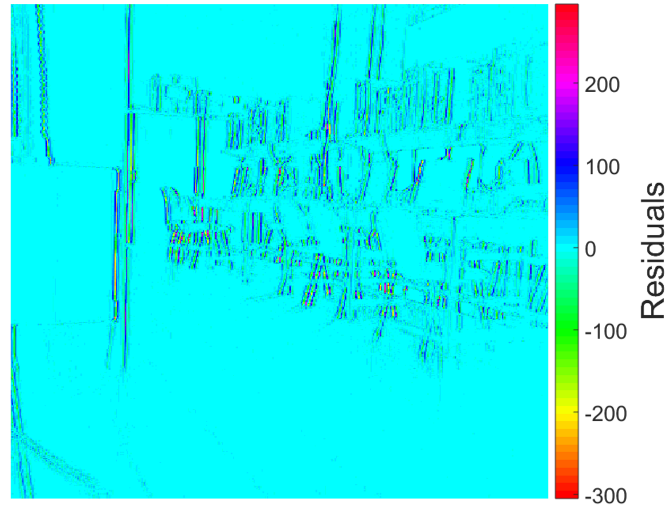


Co-occurrence
matrix

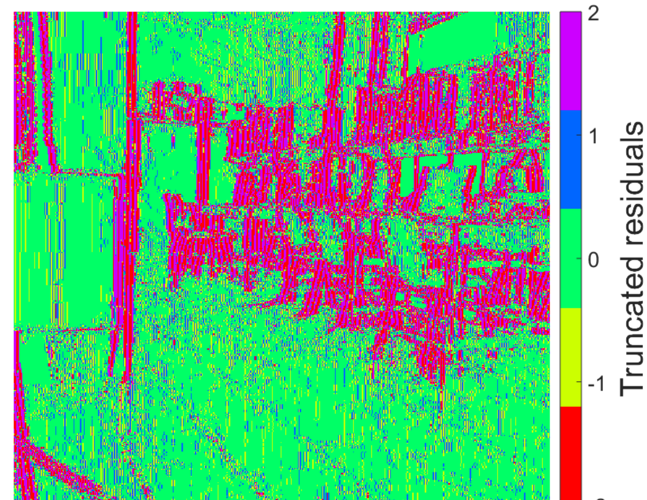
Descriptors applied to image forensics

Grayscale input frame

High-pass [1,-3,3,-1]



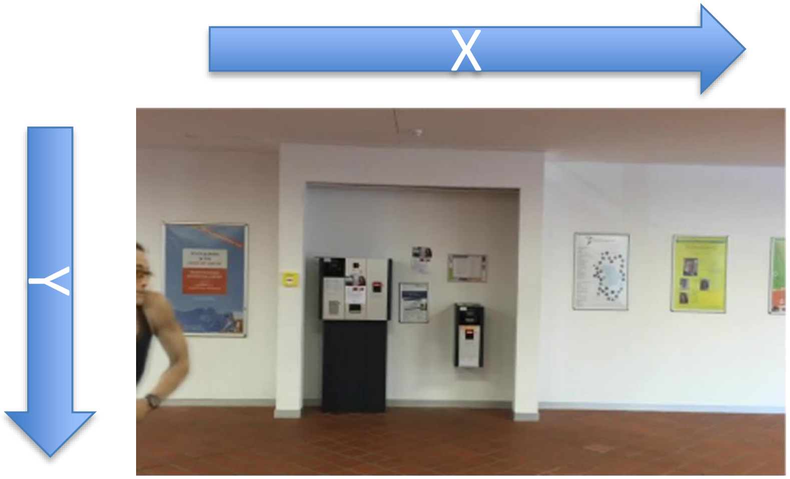
Quantize & truncate



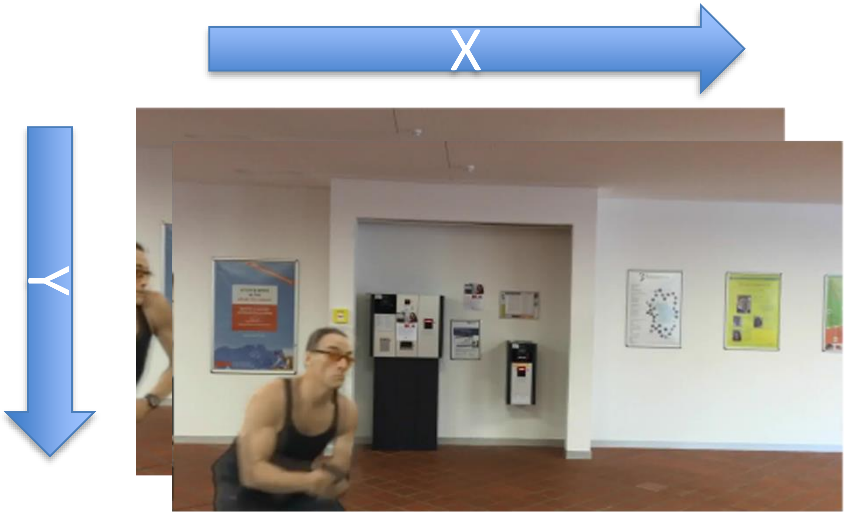
Co-occurrence matrix

	-2	-1	0	1	2
-2	8087	1256	2317	2713	15095
-1	1163	947	12097	11592	2600
0	2147	11892	84896	10277	2475
1	2732	11587	10317	854	1255
2	15340	2755	2182	1316	8208

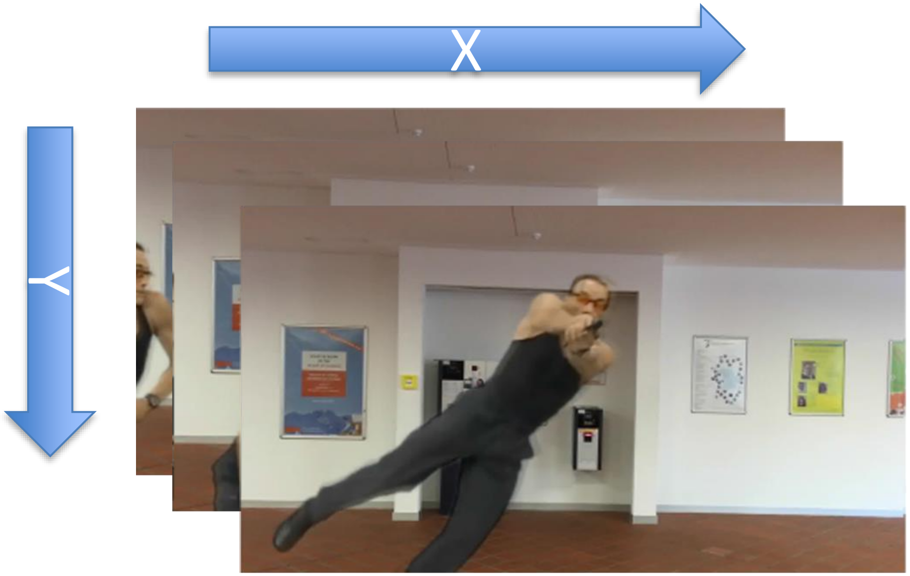
Directions



Directions



Directions



Directions

Directions

Directions

Video:

- Enlarges feature space
  → time offers new, third dimension
- Can be used to track motion by optical flow
  → to align slided windows of features

Classification pipeline

Feature Extraction

- Histogram of co-occurrence residuals
- In different directions
- On sliding windows
- Optional: align features by "optical flow"

Classification

- Calculate mahalanobis distance
- Can be thresholded

Decision

- Frame authentic?
- Frames from same camera?

Training

Train on known pristine frames

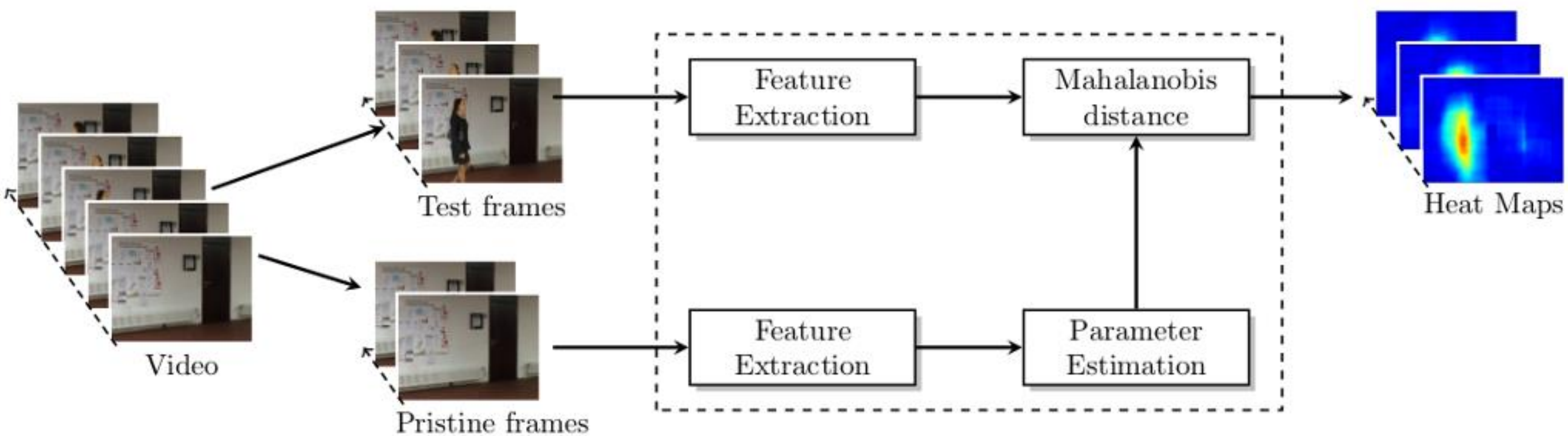
Classification pipeline

Feature Extraction
- Histogram of co-occurrence residuals
- In different directions
- On sliding windows
- Optional: align features by "optical flow"

Classification
- Calculate mahalanobis distance
- Can be thresholded

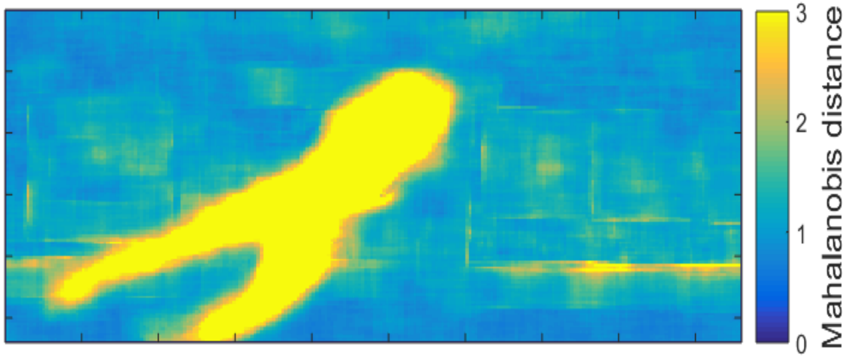
Decision
- Frame authentic?
- Frames from same camera?

Training
Train on known pristine frames

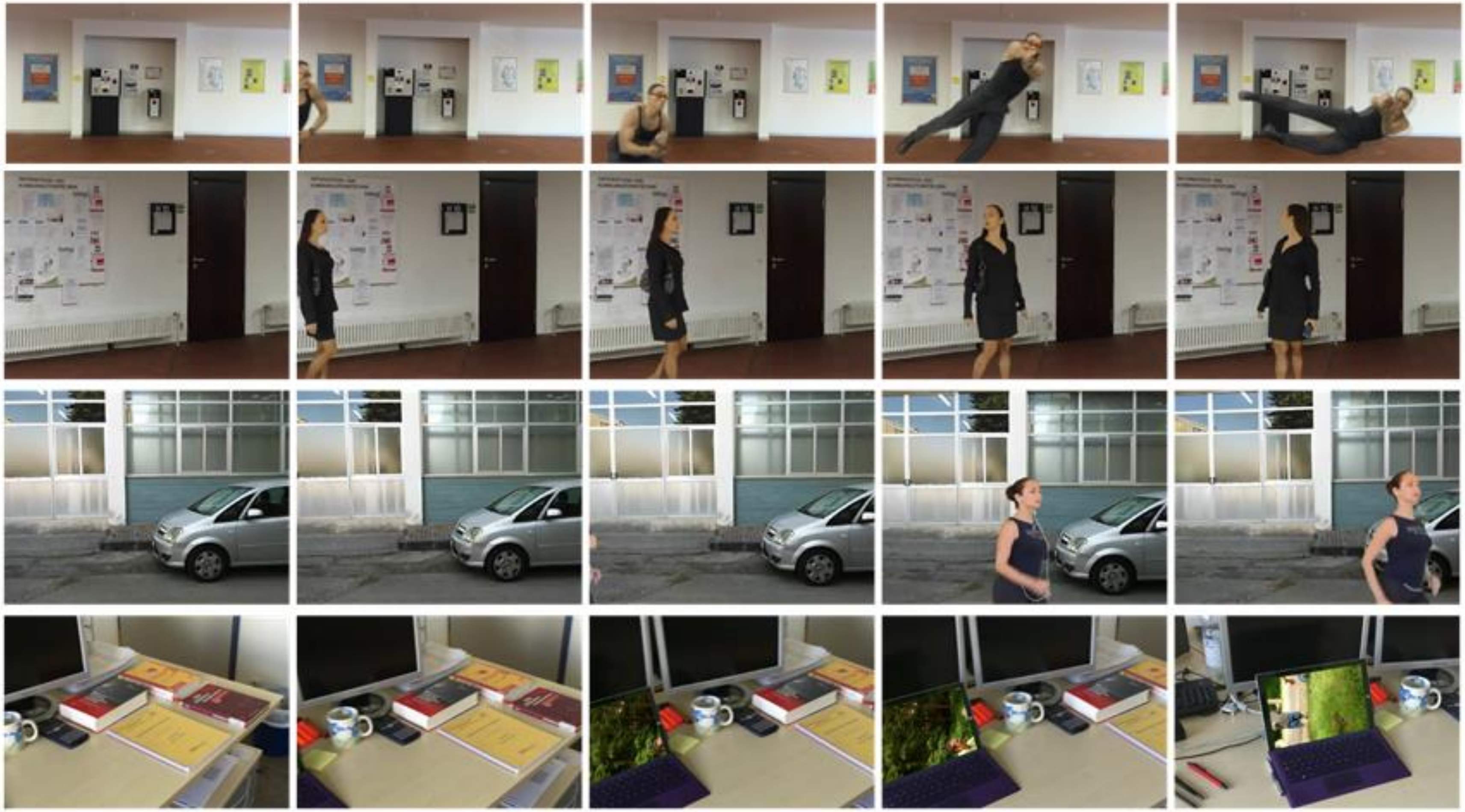


Mahalanobis distance as heatmap

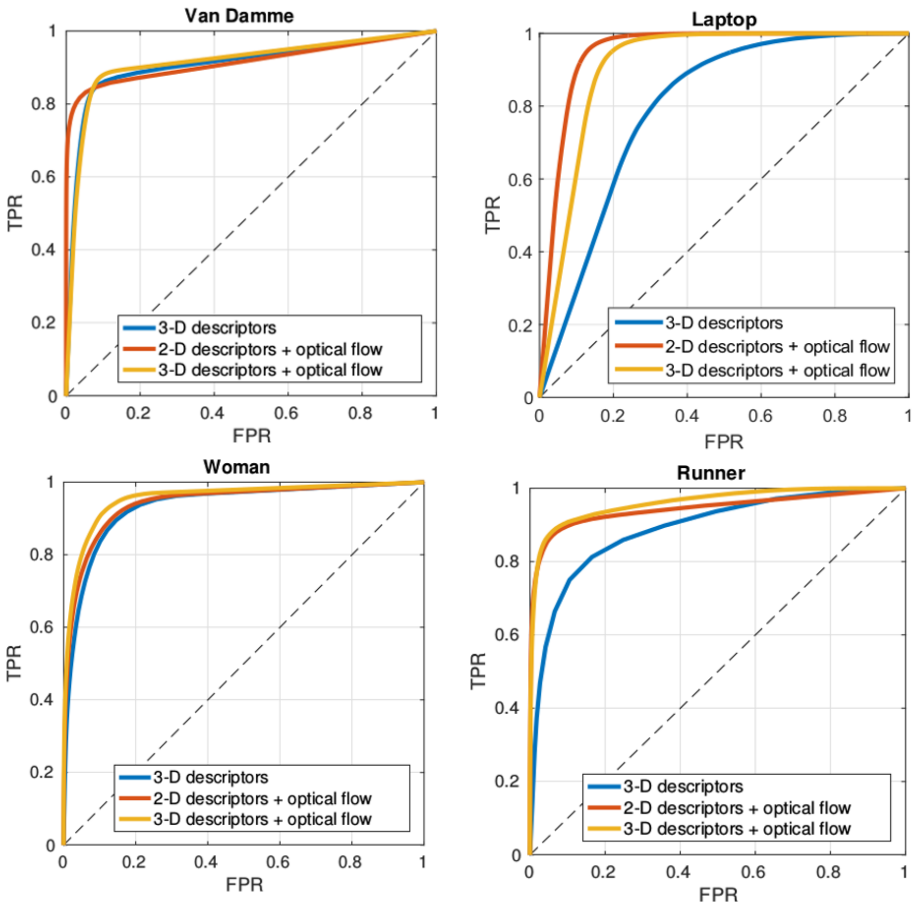
- Mahalanobis distances can be illustrated in heatmaps
- Objects spliced onto the background are revealed visually



Dataset

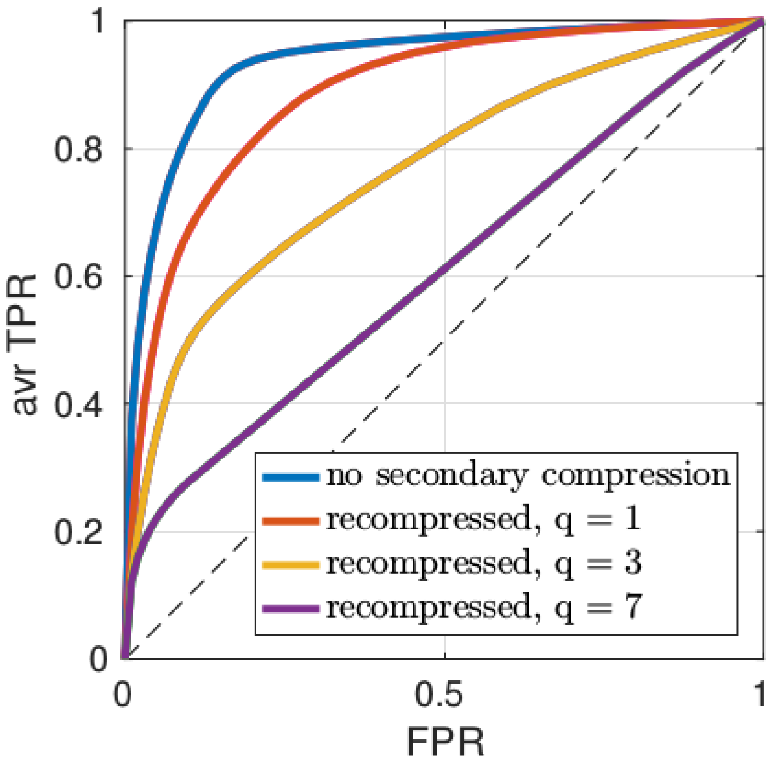


Evaluation



- Suggested method detects splicing reliable
- Incorporating optical flow to can improve results

Evaluation under compression

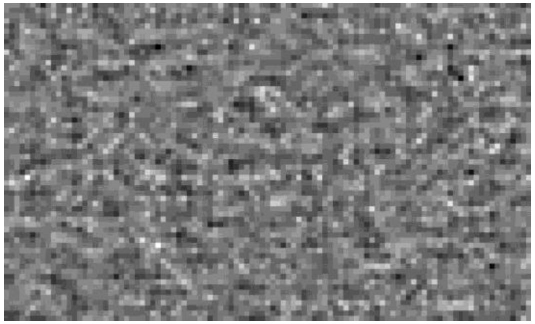


Secondary recompression of spliced material:

- Weakens its localization
- Detection results correlates (negatively) with compression factor

Related work

- Photo-response nonuniformity (PRNU) based:
  - PRUN is a profoundly unique pattern inherently present in any imaging device [1]
  - Also applied to localize video manipulations [2]

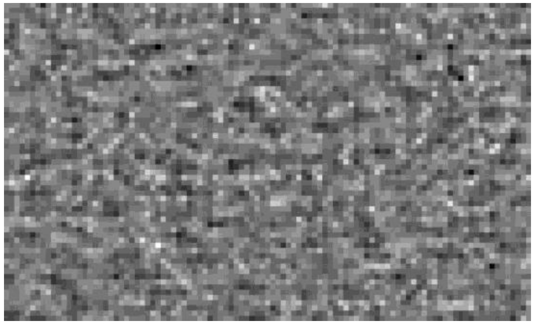


Example PRNU, amplified

[1] J. Lukás, J. Fridrich, M. Goljan, "Detecting digital image forgeries using sensor pattern noise," in *Proceedings of the SPIE*, vol. 6072, 2006
[2] W. Van Houten, Z. Geradts, "Source video camera identification for multiply compressed videos using sensor photo response non-uniformity" in *Proc. Of SPIE Security, Steganography, and Watermarking of Multimedia Contents IX*, Feb. 2007
[3] L. D'Amiano, D. Cozzolino, G. Poggi, L. Verdoliva: "Autoencoder with Recurrent Neural Networks for Video forgery detection", in *IS&T Electronic Imaging: Media Watermarking, Security and Forensics*, Feb. 2017

Related work

- Photo-response nonuniformity (PRNU) based:
  - PRUN is a profoundly unique pattern inherently present in any imaging device [1]
  - Also applied to localize video manipulations [2]



Example PRNU, amplified

- Autoencoder (AE) based [3]:
  - AEs are a special neural network architecture
  - Training subject to reconstruct input from compressed state z with little error as possible: $\min\{\mathcal{L}(f,\hat{f})\} \rightarrow$ If new input differs, $\mathcal{L}$ becomes large

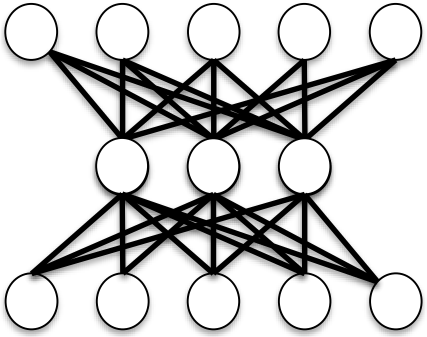
Output vector $\hat{f}$

↑

Compressed state z

↑

Input vector f



Decode

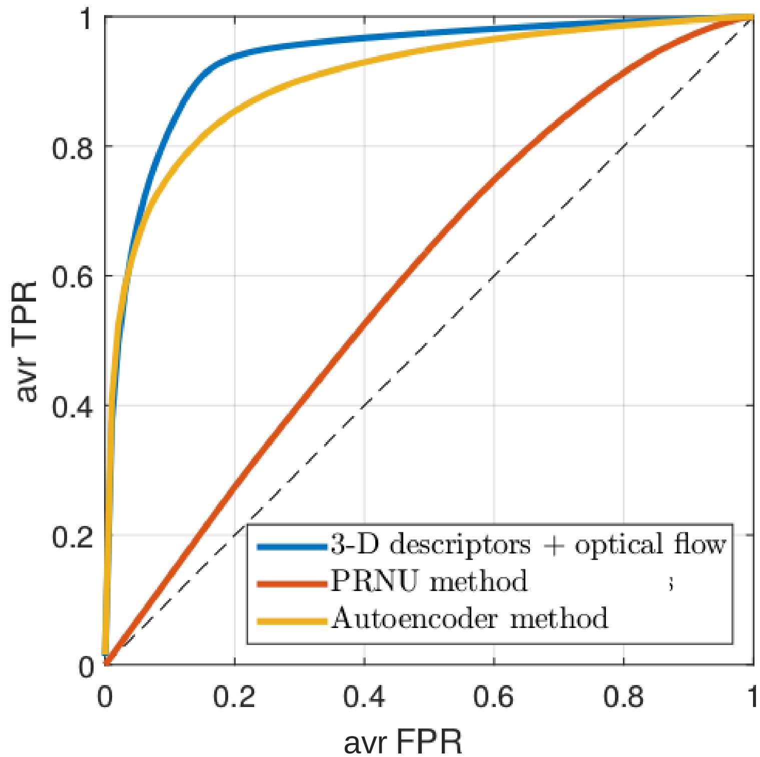
Encode

[1] J. Lukás, J. Fridrich, M. Goljan, “Detecting digital image forgeries using sensor pattern noise,” in *Proceedings of the SPIE*, vol. 6072, 2006

[2] W. Van Houten, Z. Geradts, “Source video camera identification for multiply compressed videos using sensor photo response non-uniformity” in *Proc. Of SPIE Security, Steganography, and Watermarking of Multimedia Contents IX*, Feb. 2007

[3] L. D’Amiano, D. Cozzolino, G. Poggi, L. Verdoliva: “Autoencoder with Recurrent Neural Networks for Video forgery detection”, in *IS&T Electronic Imaging: Media Watermarking, Security and Forensics*, Feb. 2017

Comparison with other methods



- Suggested framework can produce better results then other works
- AE does not utilize information about movement in videos, like incorporating optical flow in the suggested framework
- PRNU might have difficulties to build a meaningful model from correlated frames

Summary and Outlook

Presented Algorithm:

- Distinguishes different noise distributions, present in a spliced video
- Tested successfully on green screen splicing
- Additional secondary compression influences performance

Future Work:

- Build up bigger database
- Apply algorithms to different kinds of forgeries
- Also apply to video source identification (e.g. on non-forged videos)

Thanks for your attention!
Questions?

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT