POLITECNICO DI MILANO

REVERSE ENGINEERING OF VIDEO CONTENT FOR FORENSIC ANALYSIS

Paolo Bestagini
Ph.D.

paolo.bestagini@polimi.it

Motivations

Motivations

Diffusion of multimedia **sharing platforms**

Motivations

- Diffusion of multimedia **sharing platforms**

Motivations

- Diffusion of multimedia **sharing platforms**
  - Huge amount of user-generated content

Motivations

- Diff
  - H

Availability of user-friendly editing software

Motivations

- Diffusion of multimedia **sharing platforms**
  - Huge amount of user-generated content

- Availability of user-friendly **video-editing software**

Motivations

- Diffusion of multimedia **sharing platforms**
  - Huge amount of user-generated content
- Availability of user-friendly **video-editing software**
  - Easy to tamper with videos

Motivations

- Diffusion of r
  - Huge amo

We **cannot trust** what we see!

- ftware

Motivations

- Diffusion of r
  - Huge amo

We cannot trust what we see!

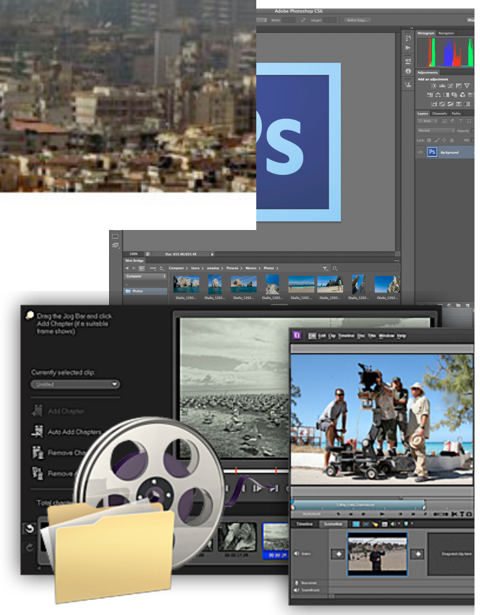
ftware

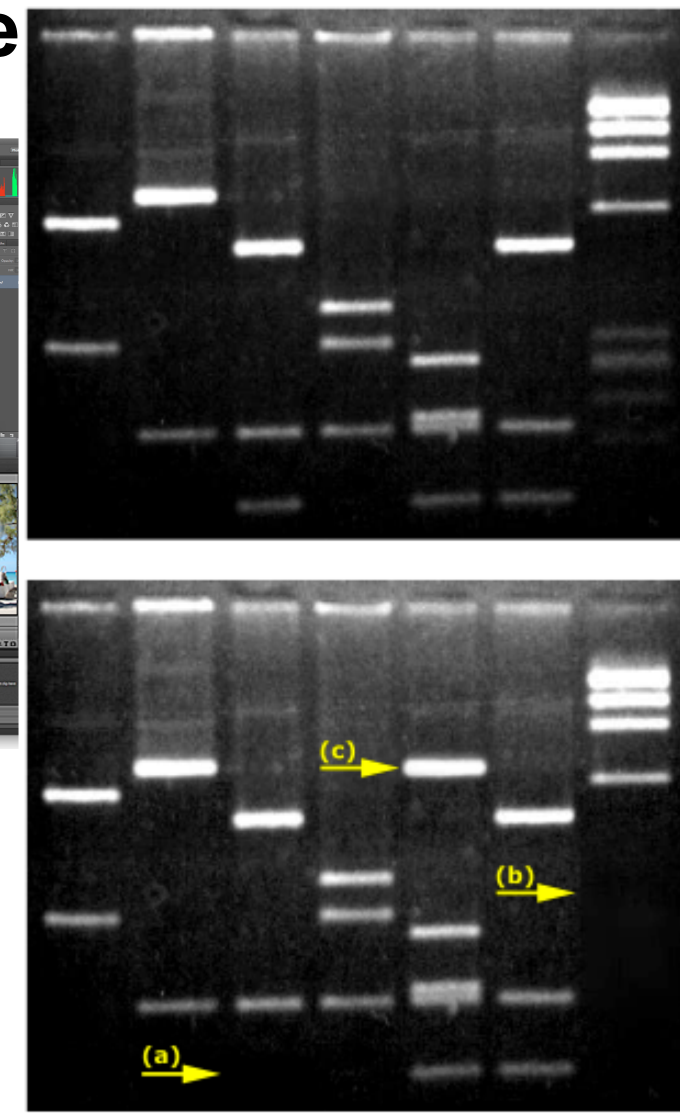
Motivations

- Diffusion of r
  - Huge amo

We cannot trust what we see!

ftware



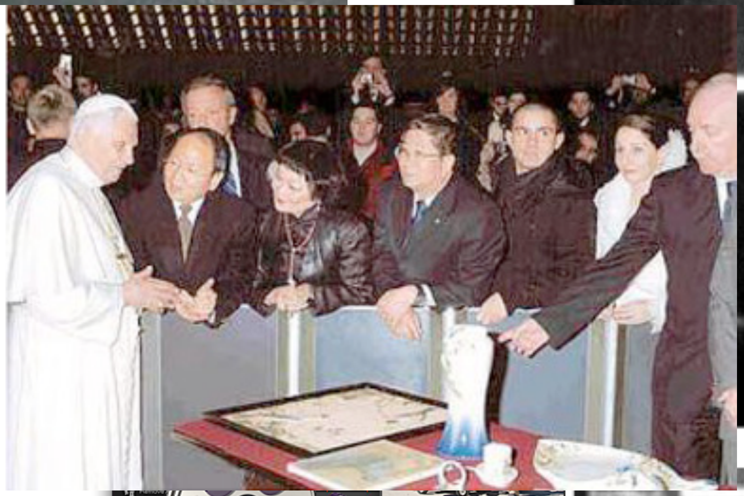


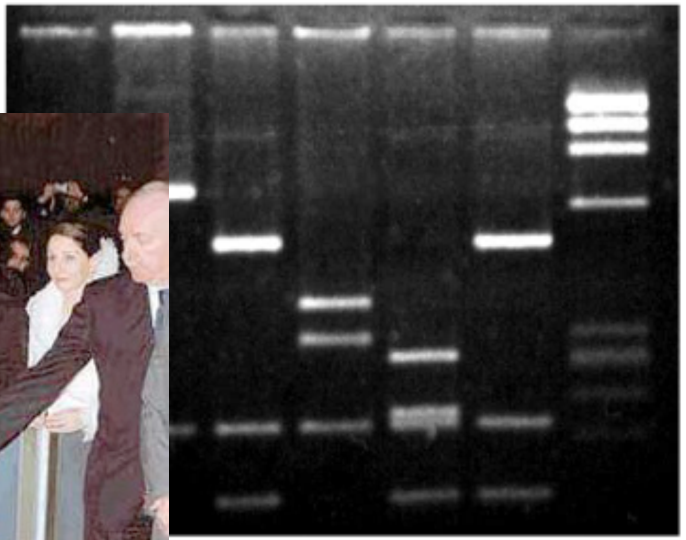
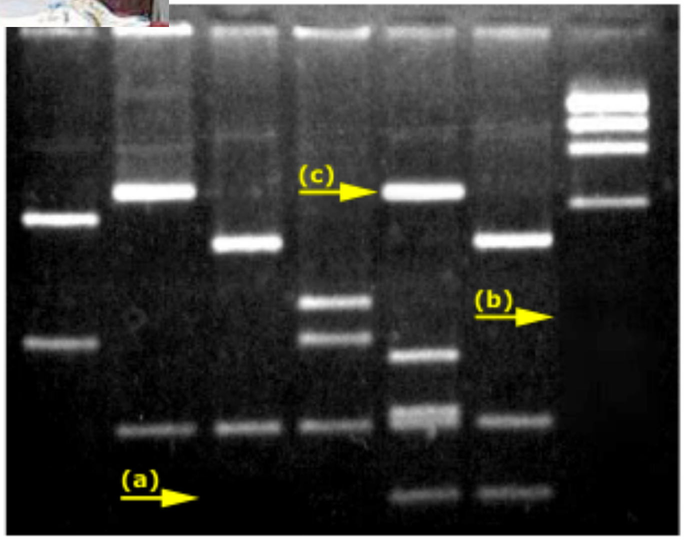
Motivations

- Diffusion of r…
  - Huge amo…

We cannot trust what we see!


…ftware

Motivations

- Diffusion of r
  - Huge amo

We cannot trust what we see!

Motivations

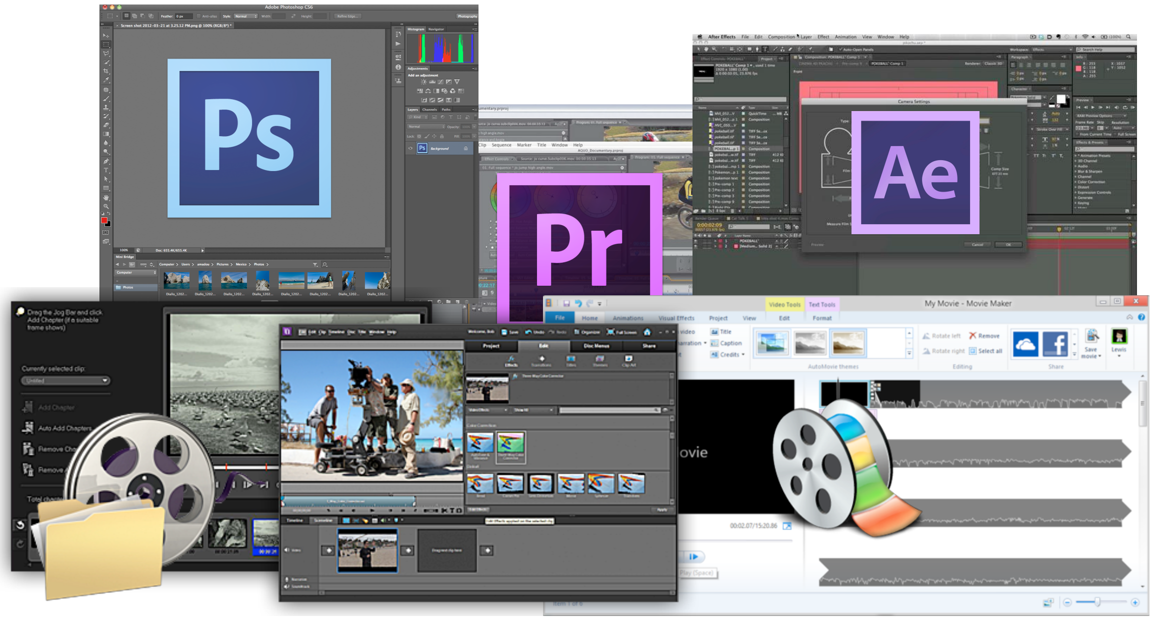
- Diffusion of r
  - Huge amo

We cannot trust what we see!

ftware

Motivations

- Diffusion of multimedia **sharing platforms**
  - Huge amount of user-generated content
- Availability of user-friendly **video-editing software**
  - Easy to tamper with videos



- We **cannot trust** what we see!

Motivations

- Diffusion of multimedia **sharing platforms**
  - Huge amount of user-generated content
- Availability of user-friendly **video-editing software**
  - Easy to tamper with videos
- We **cannot trust** what we see!

TRACE BACK FILE HISTORY

Objectives

Objectives

- **What can we do?**
  - To develop a series of **blind** algorithms and tools for **video forensic analyses** working in a real world scenario.

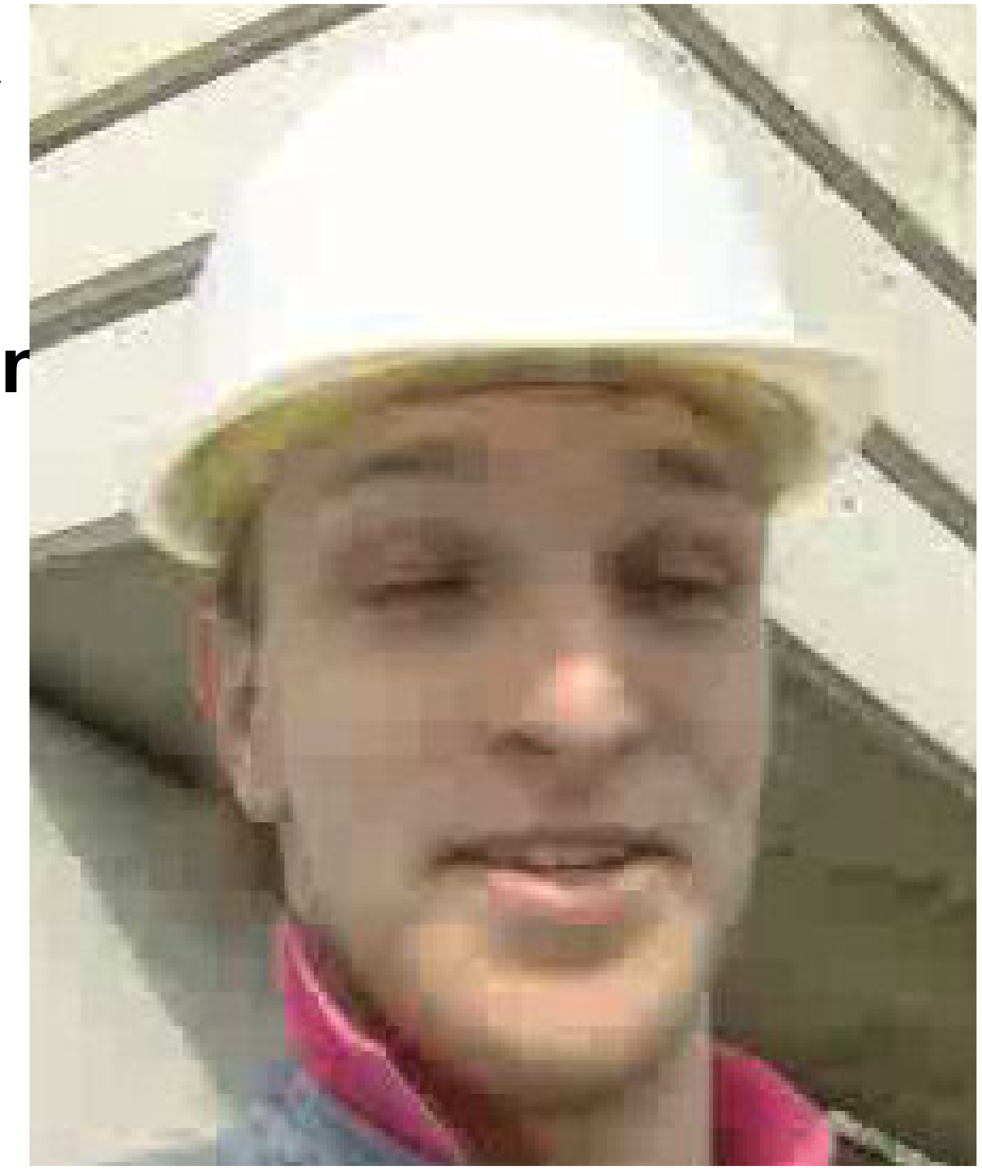
Objectives

- **What can we do?**
  - To develop a series of **blind** algorithms and tools for **video forensic analyses** working in a real world scenario.

- **How?**
  - Every **non-reversible operation** leaves peculiar footprints.
  - **Footprints** as an asset.

Objectives

- **What can we do?**
  - To develop a series of **blind** algorithms and tools for **video**
    for a
- **How?**
  - Eve or
  - **Foc**



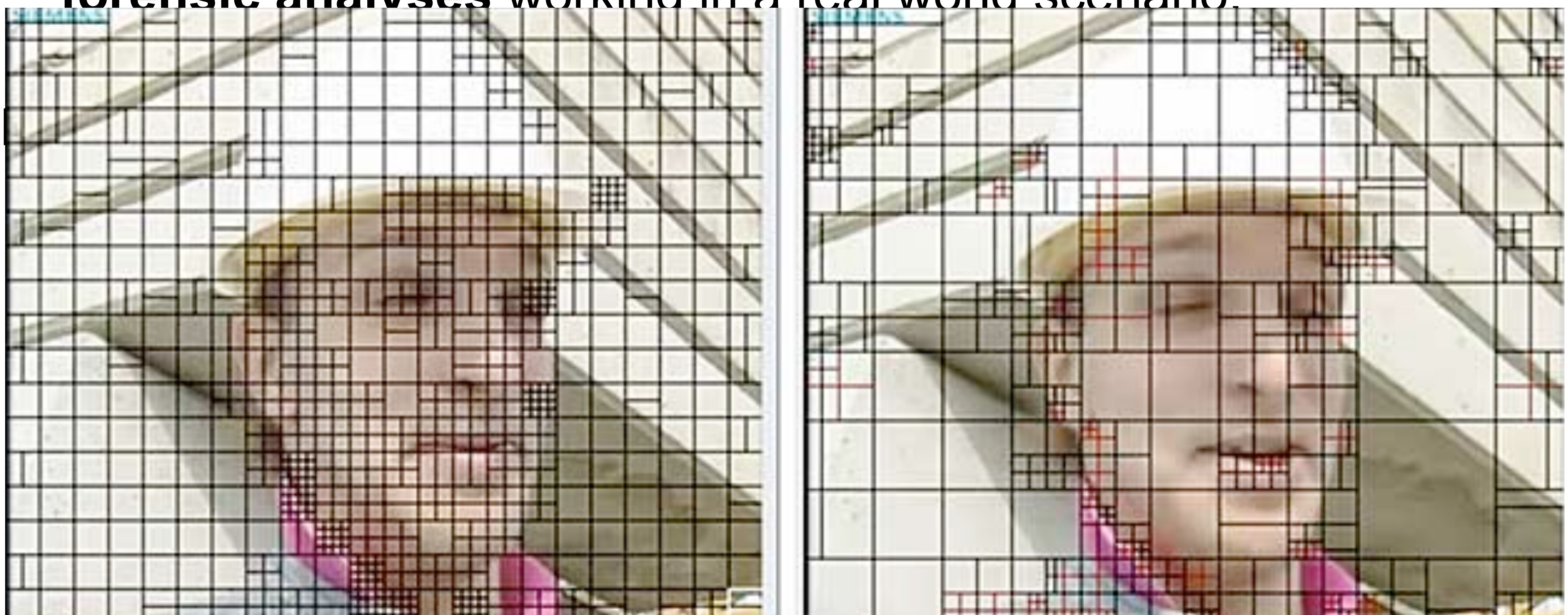
Objectives

- **What can we do?**
  - To develop a series of **blind** algorithms and tools for **video forensic analyses** working in a real world scenario.

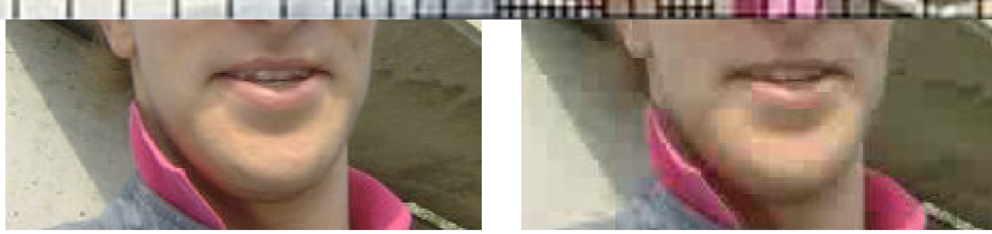
- **How?**
  - Every **non-reversible operation** leaves peculiar footprints.
  - **Footprints** as an asset.

Objectives

- **What can we do?**
  - To develop a series of **blind** algorithms and tools for **video forensic analyses** working in a real world scenario.



-



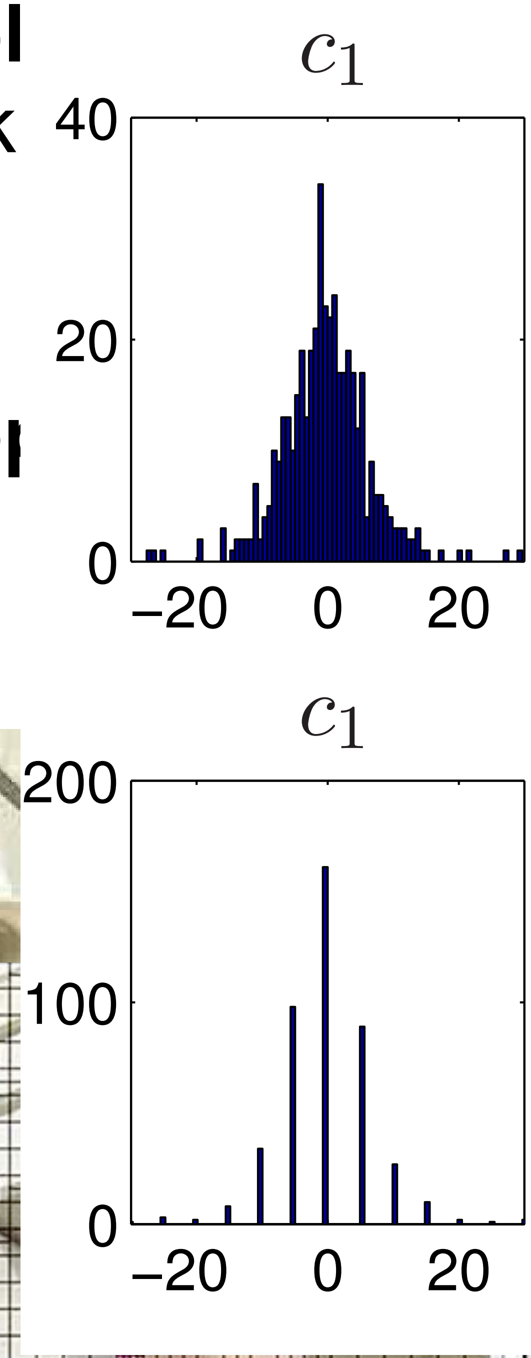
Objectives

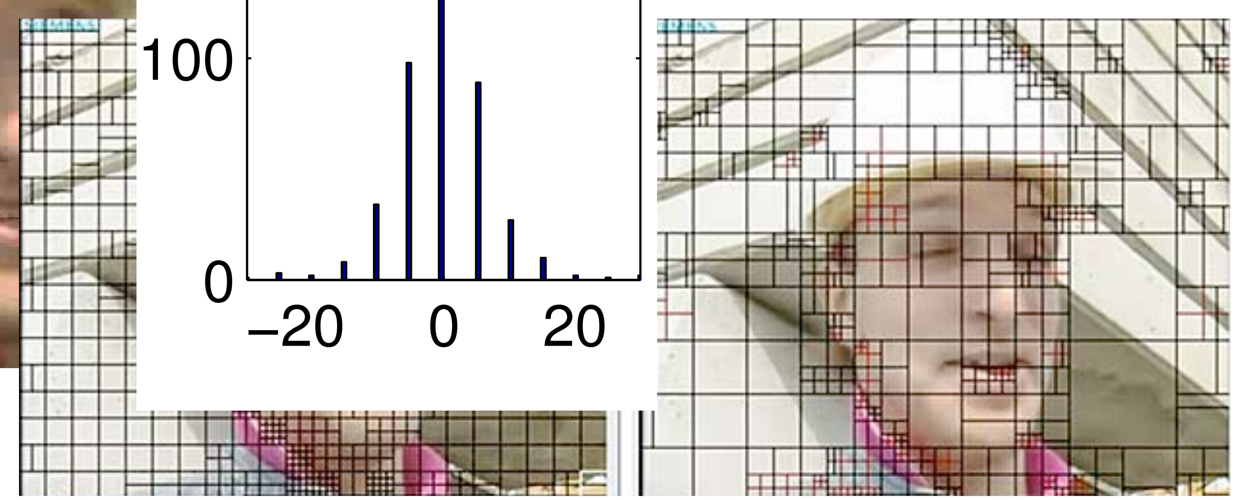
- **What can we do?**
  - To develop a series of **blind** algorithms and tools for **video forensic analyses** working in a real world scenario.

- **How?**
  - Every **non-reversible operation** leaves peculiar footprints.
  - **Footprints** as an asset.

Objectives

- **What can we do?**
  - To develop a series of **bl** and tools for **video forensic analyses** work rld scenario.
- **How?**
  - Every **non-reversible o** s peculiar footprints.
  - **Footprints** as an asset.

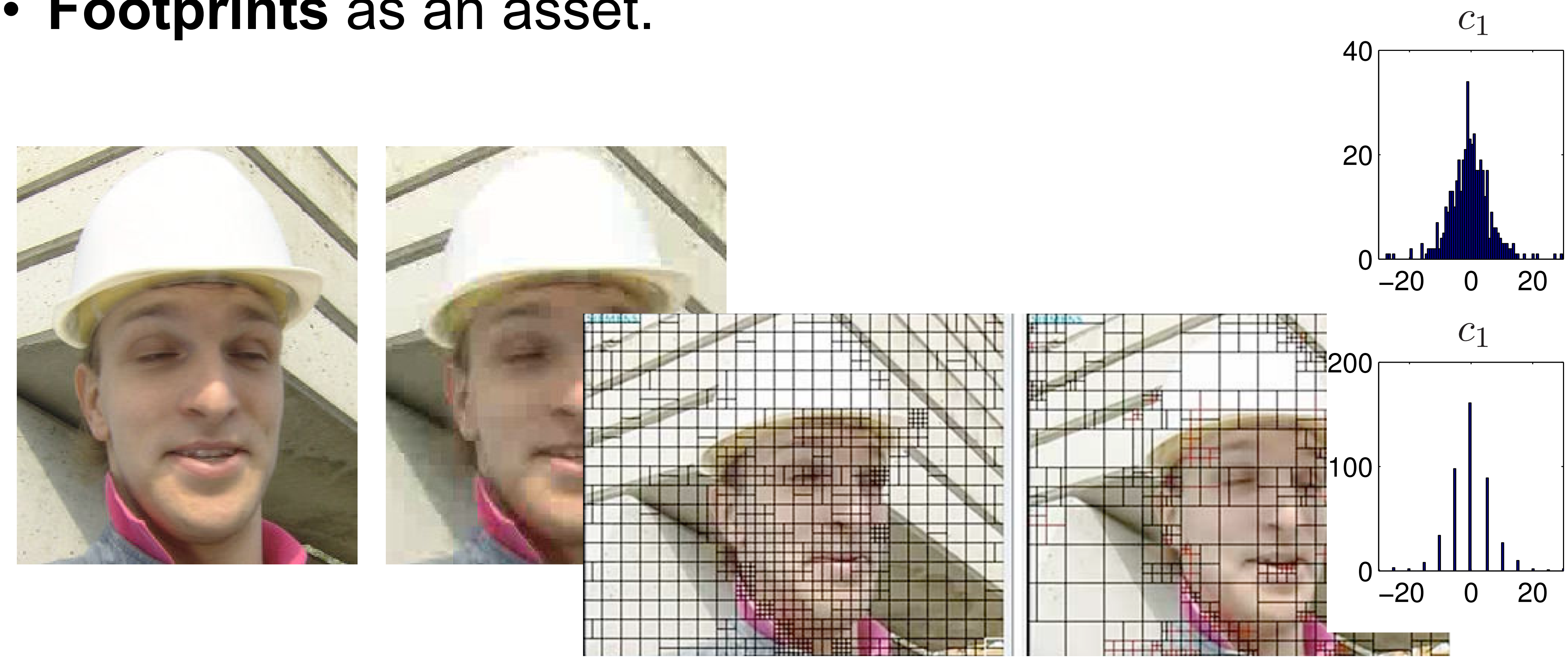




Objectives

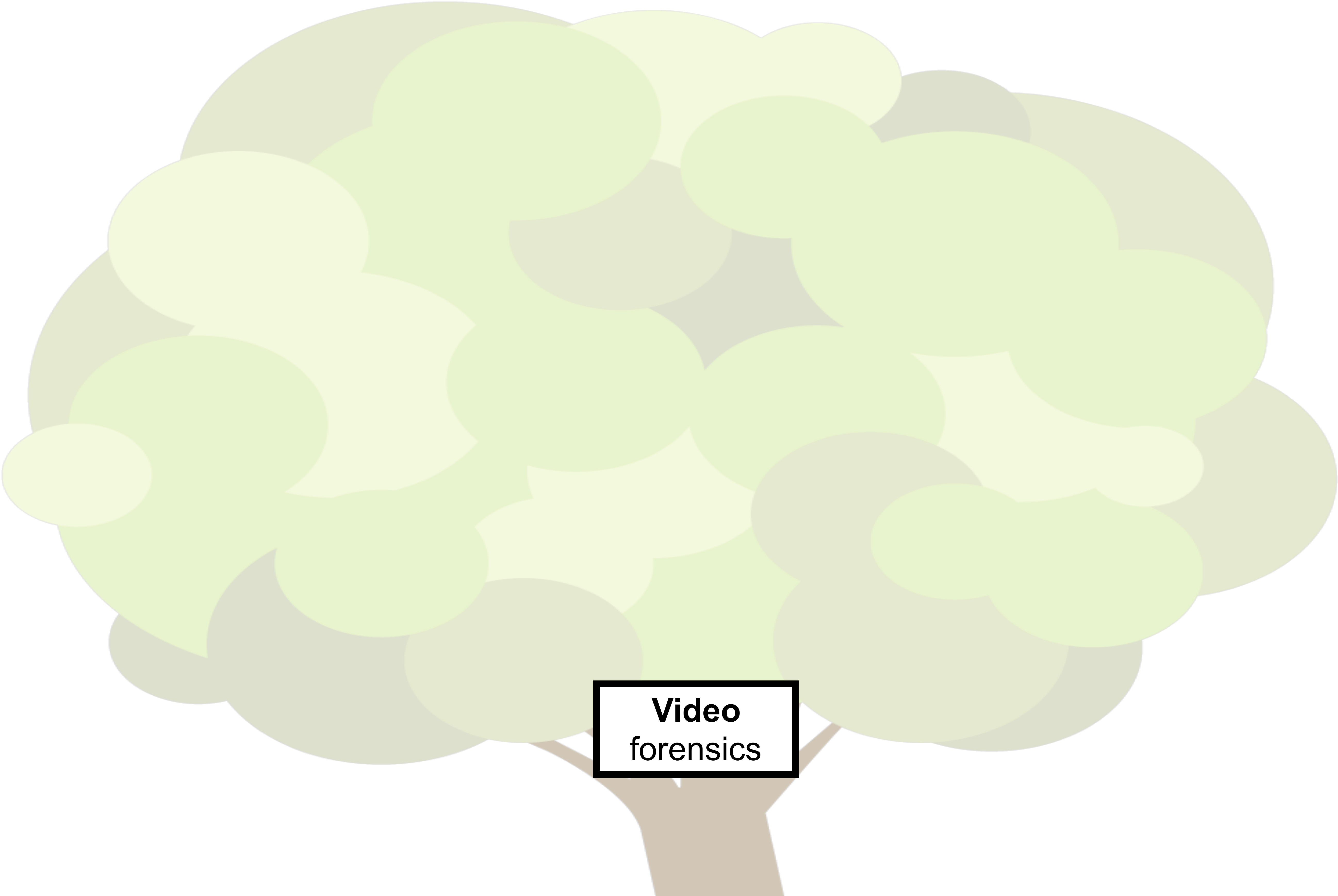
- **What can we do?**
  - To develop a series of **blind** algorithms and tools for **video forensic analyses** working in a real world scenario.

- **How?**
  - Every **non-reversible operation** leaves peculiar footprints.
  - **Footprints** as an asset.

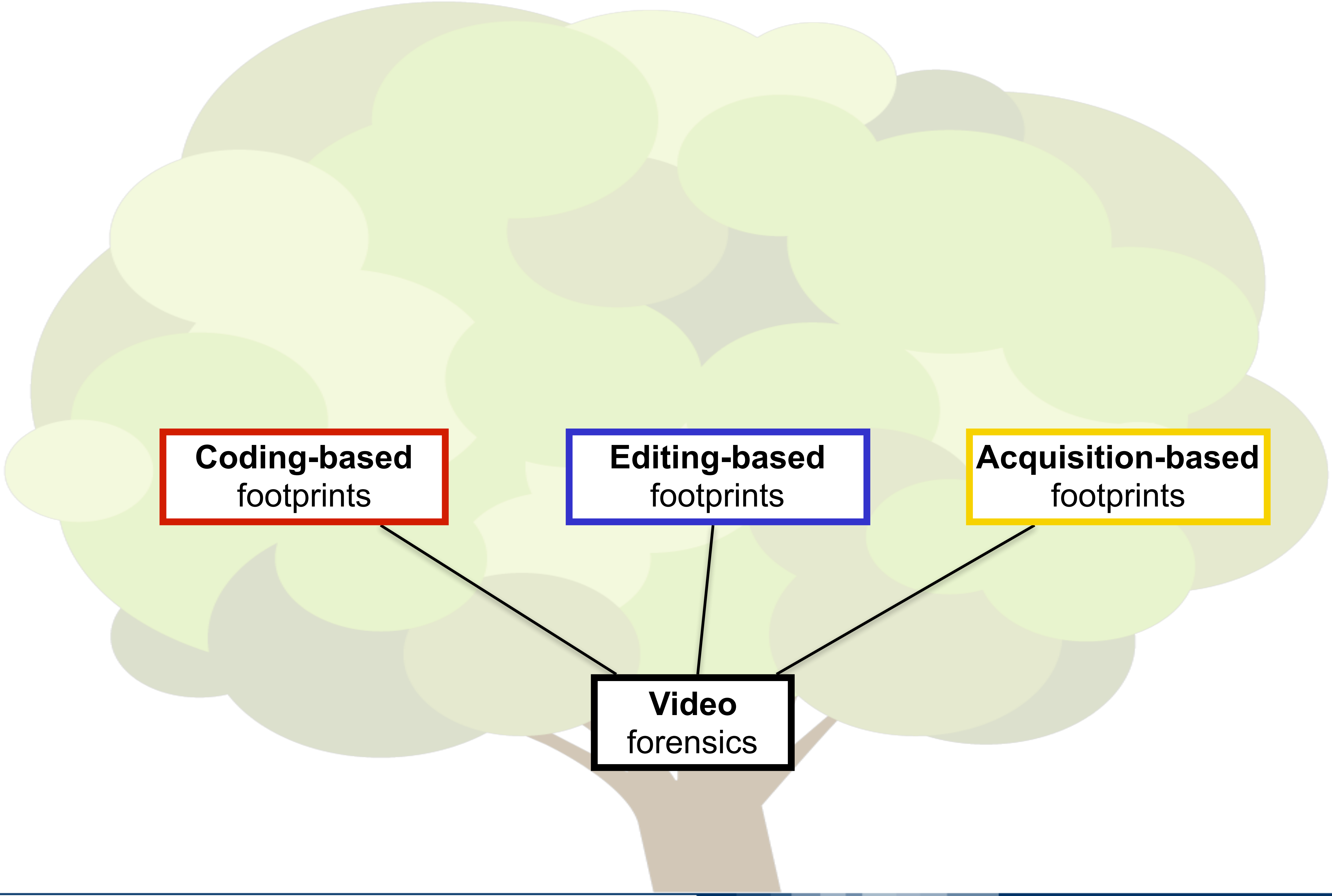


Analysis on a single video

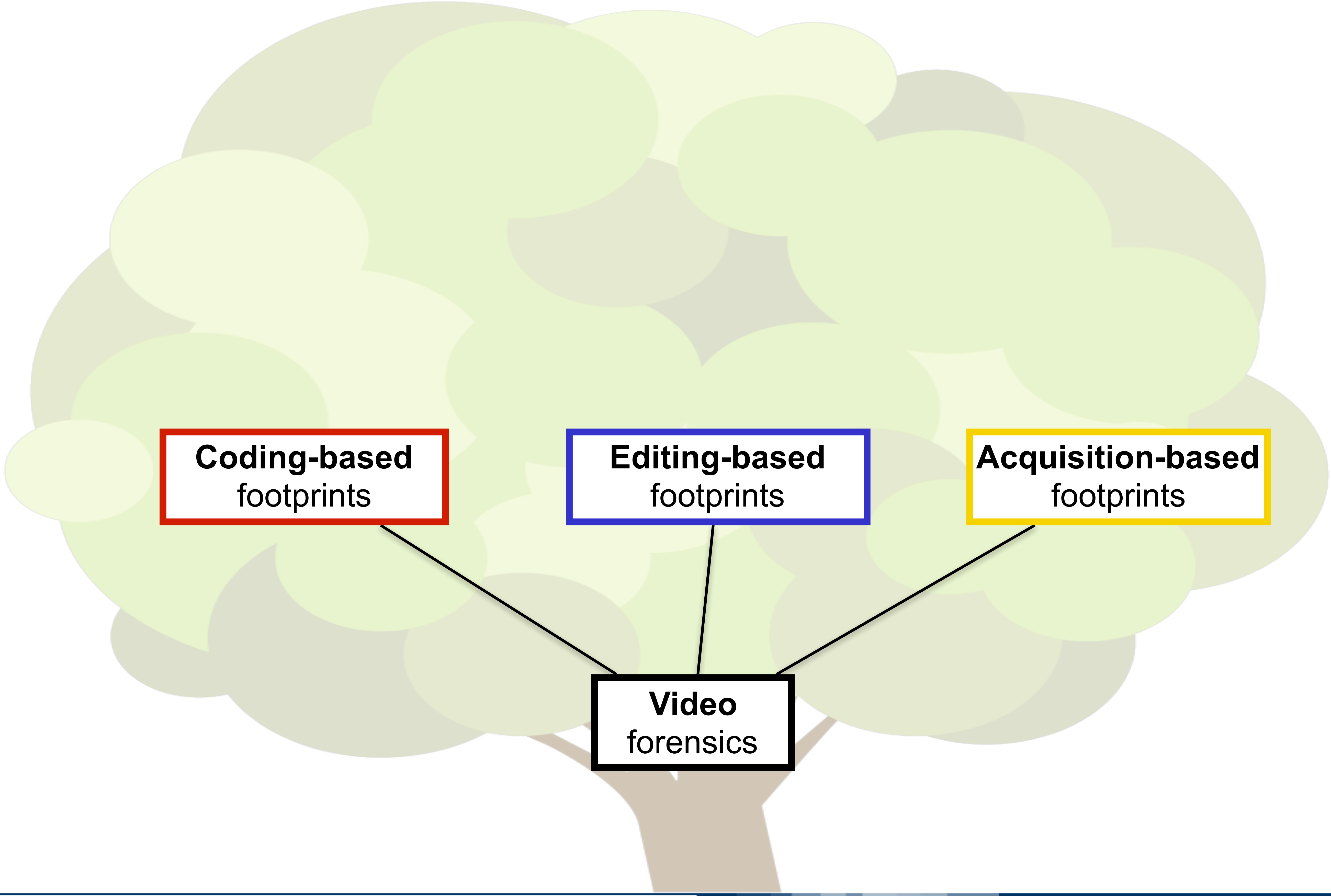
Analysis on a single video



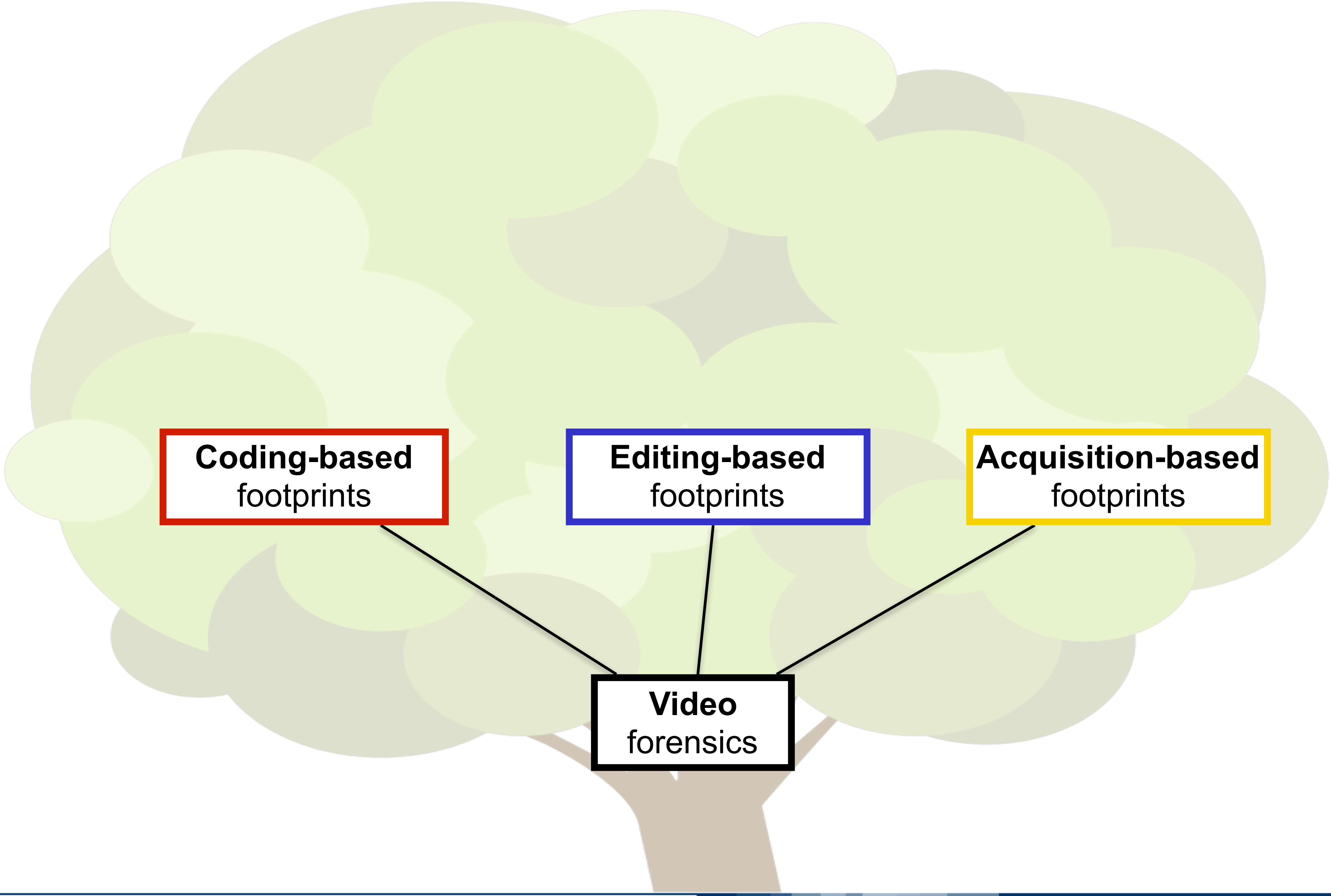
Analysis on a single video



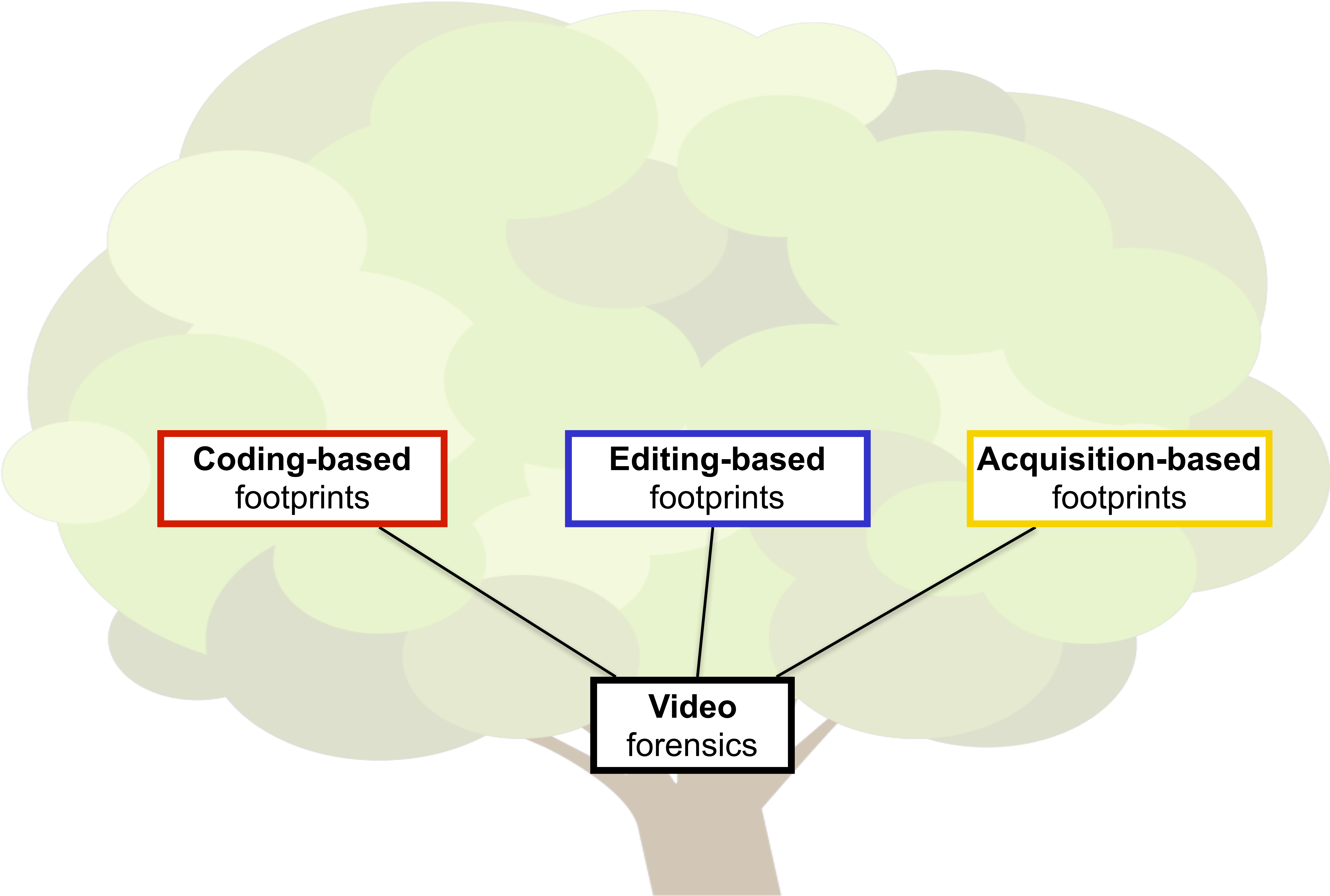
Analysis on a single video



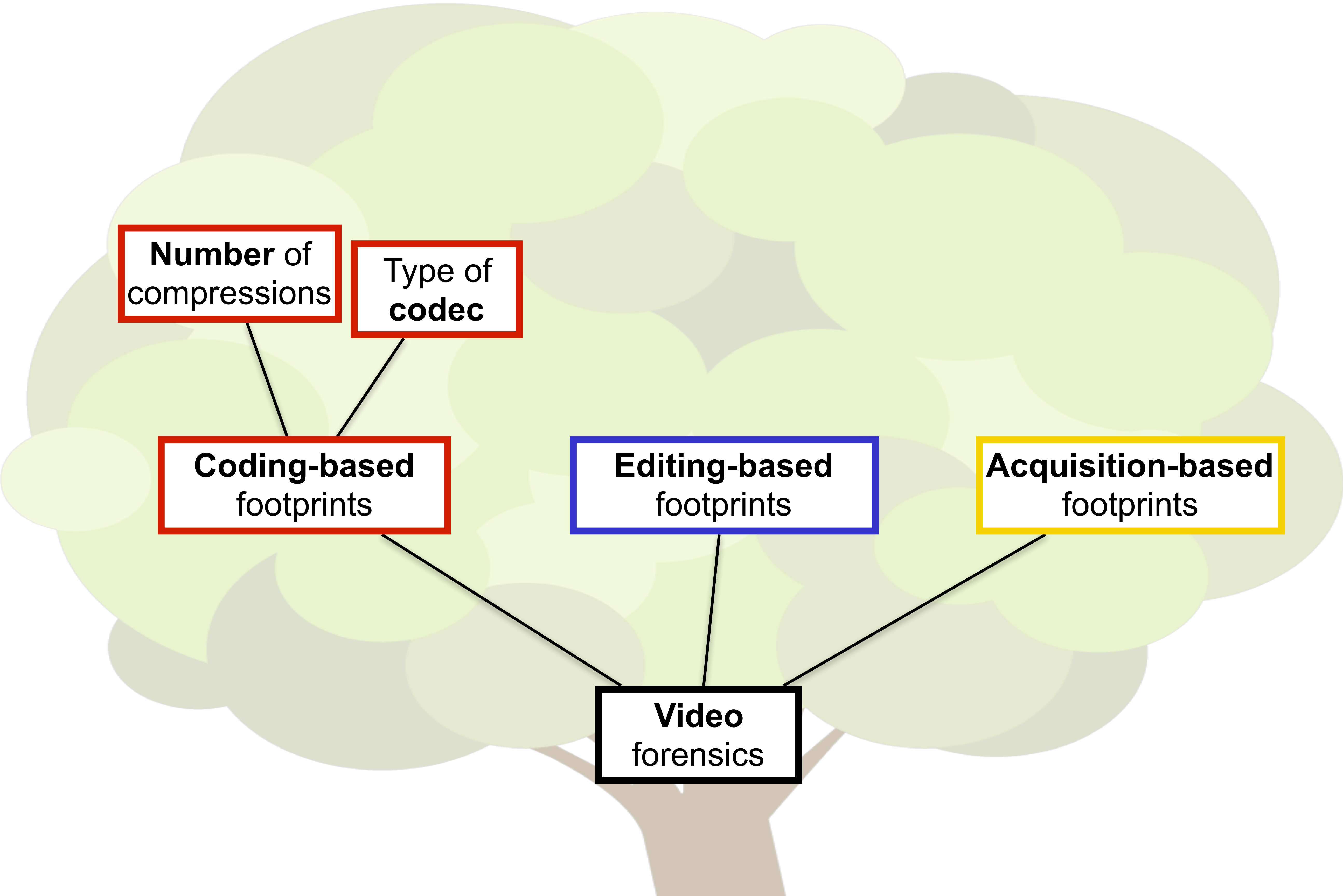
Analysis on a single video



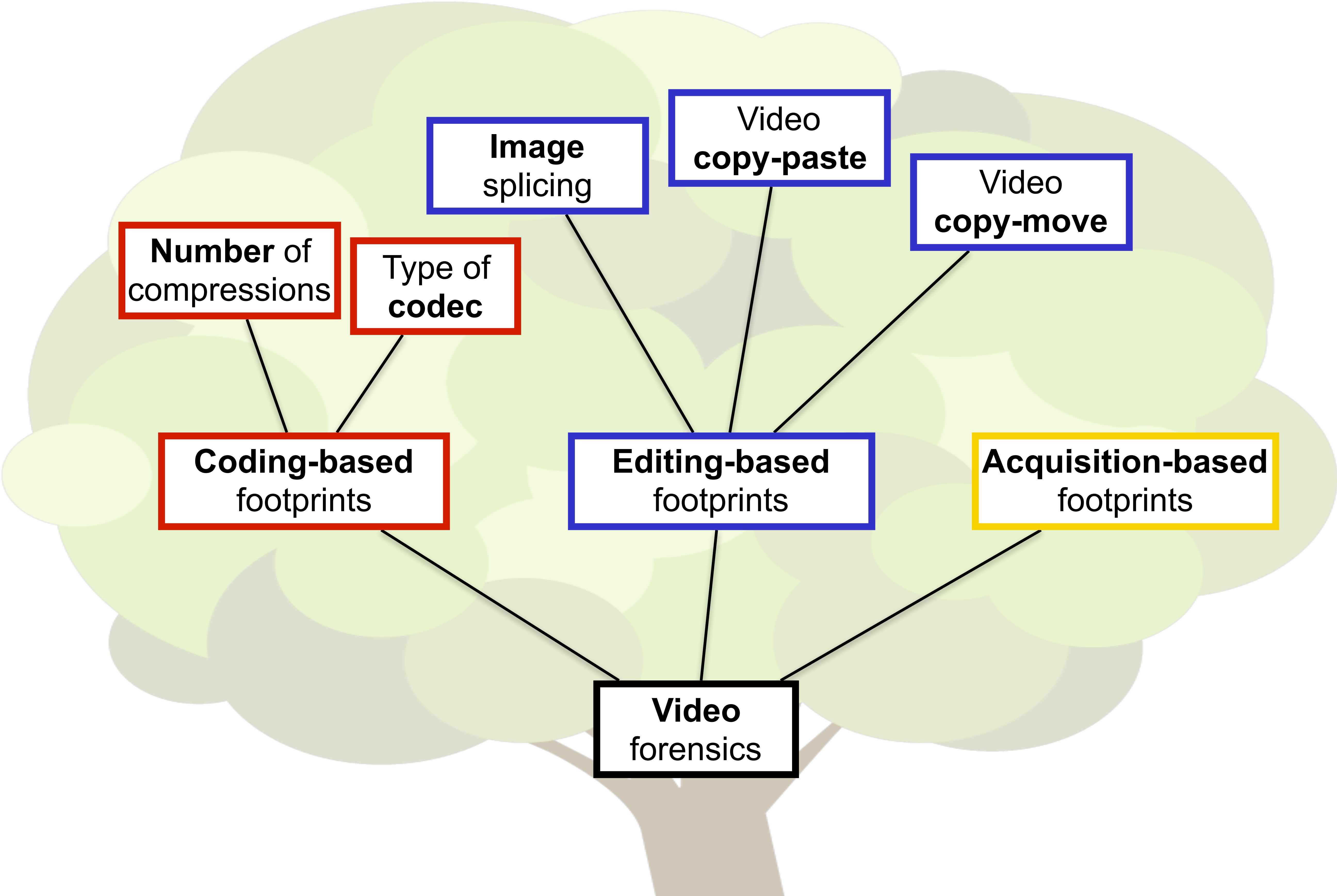
Analysis on a single video



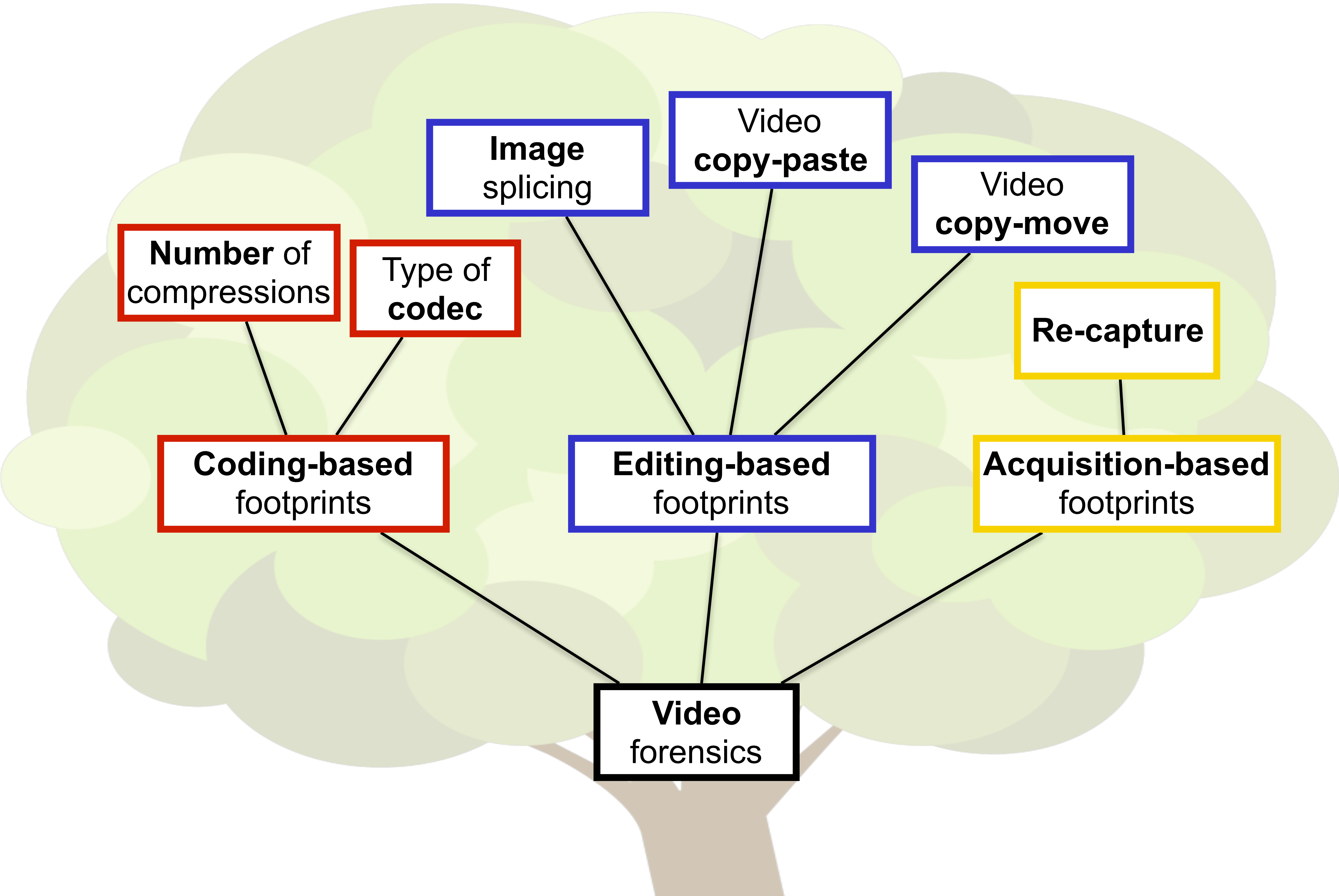
Analysis on a single video



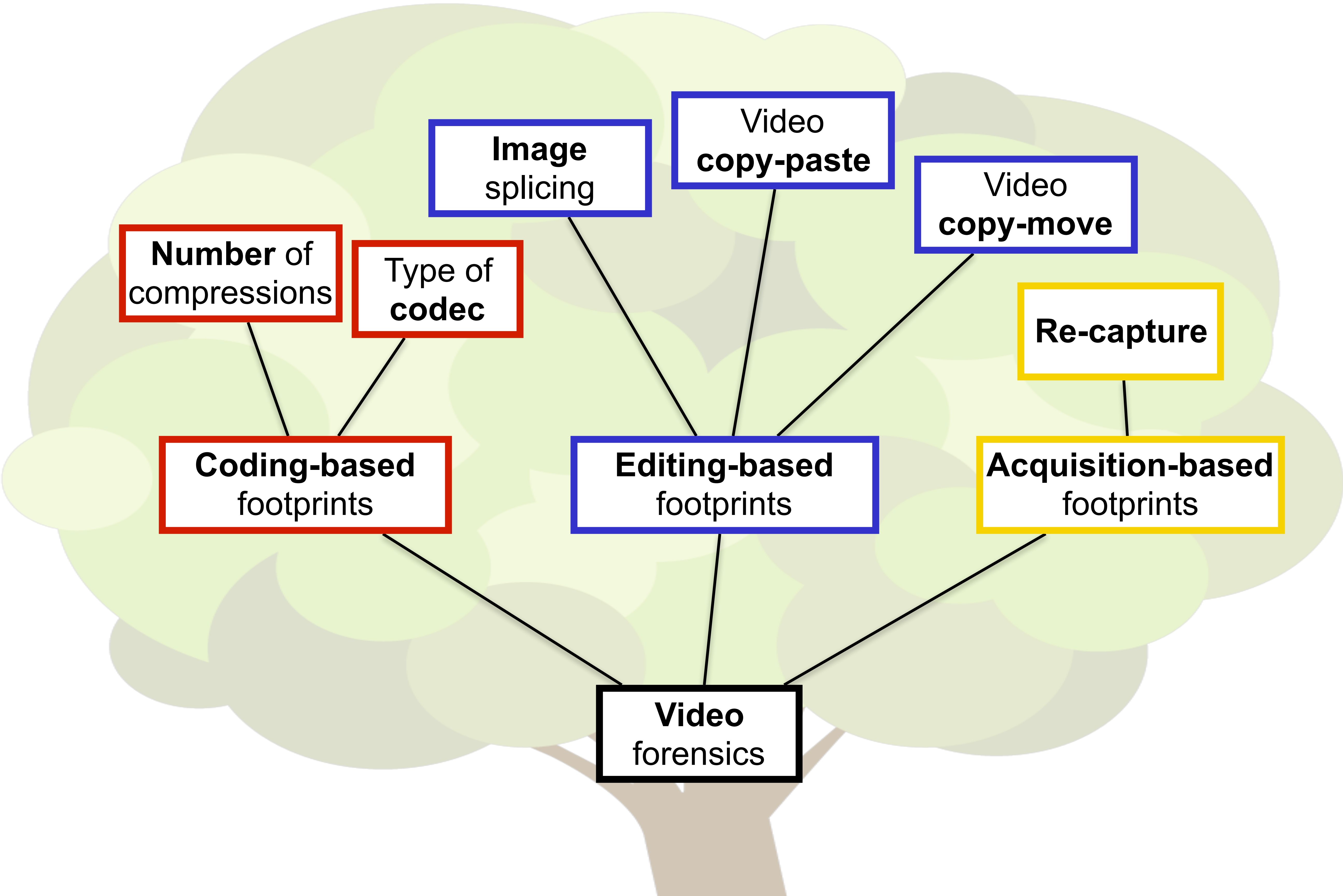
Analysis on a single video

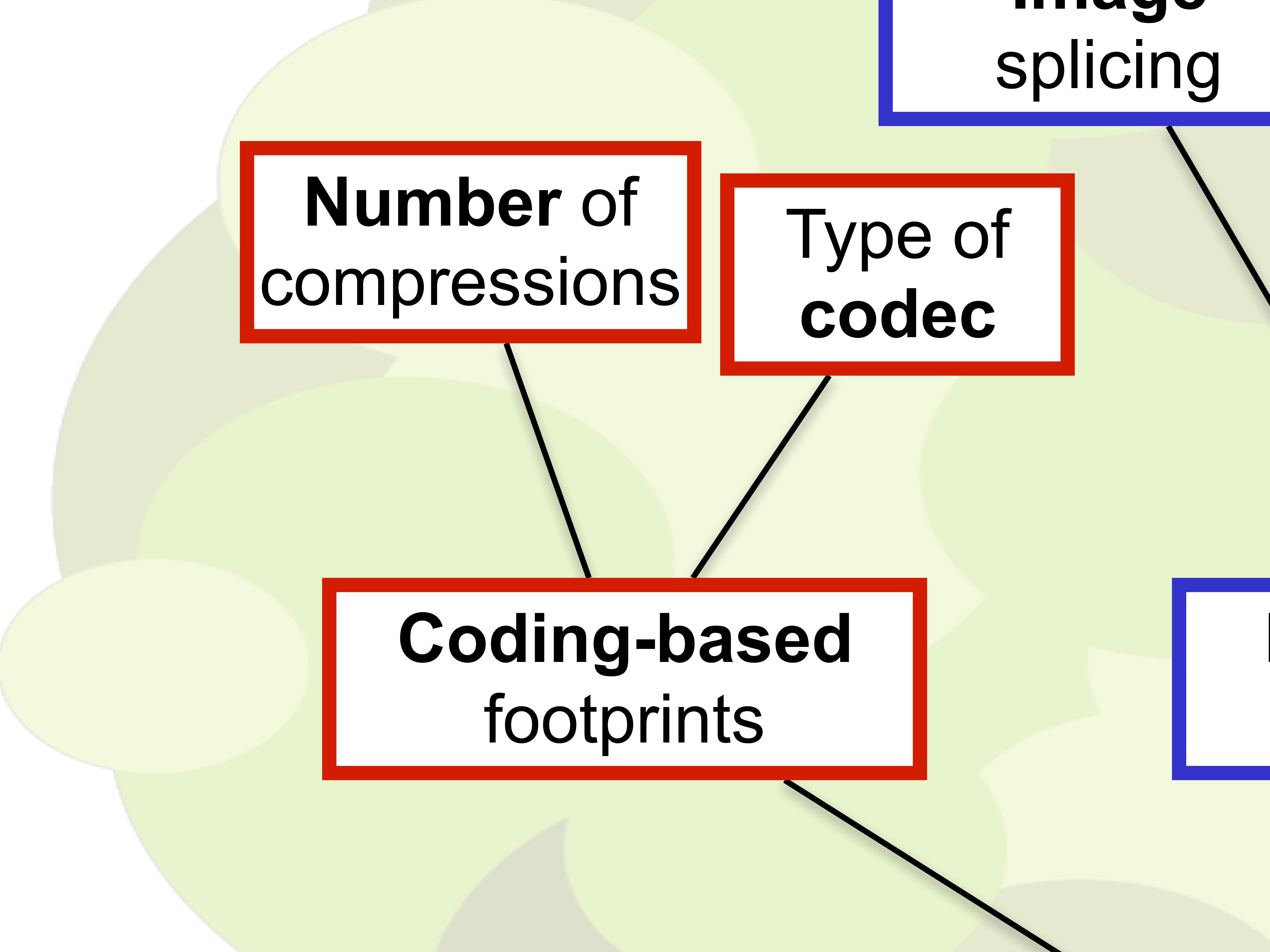


Analysis on a single video



Work organisation

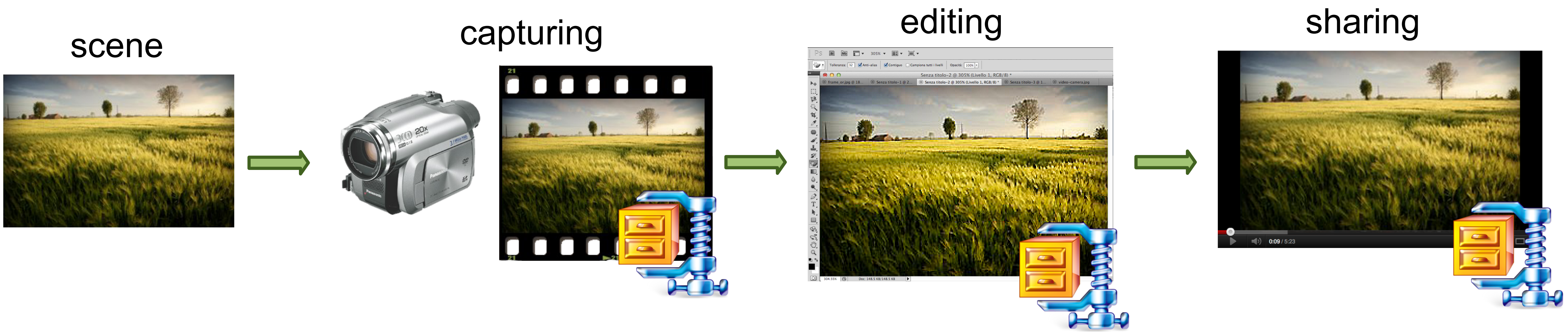




splicing
Number of compressions
Type of codec
Coding-based footprints

Coding-based footprints

- Videos are often **encoded multiple times** during their life-time
  - Information about acquisition device
  - The number of compression steps is an indicator of a video reliability

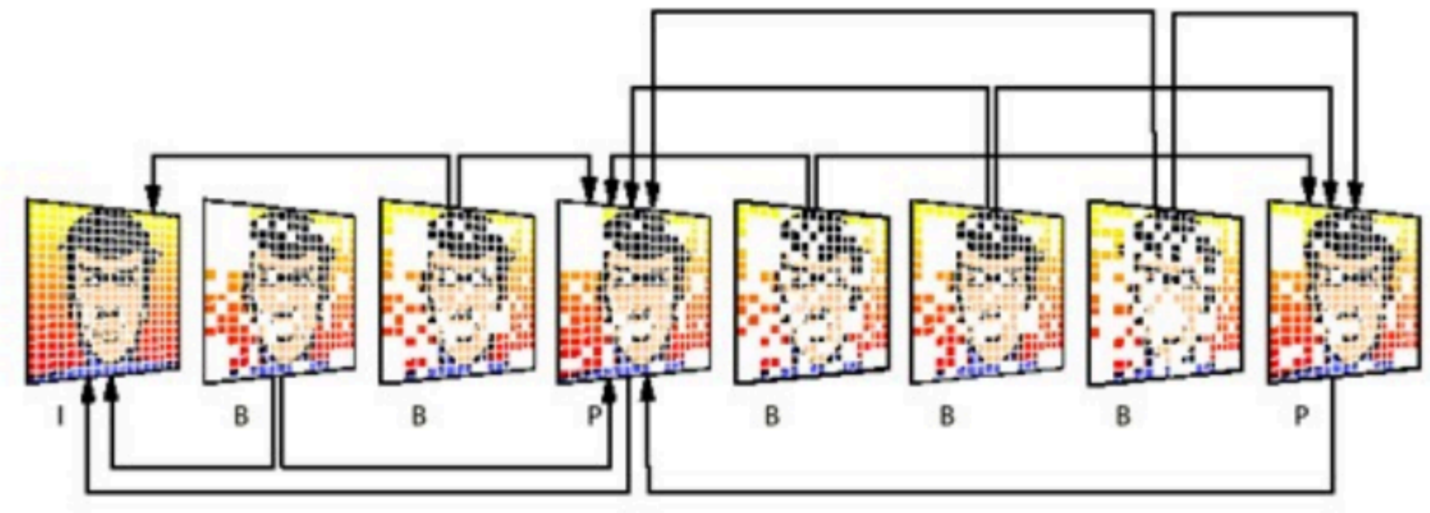


How many times has the video been compressed?

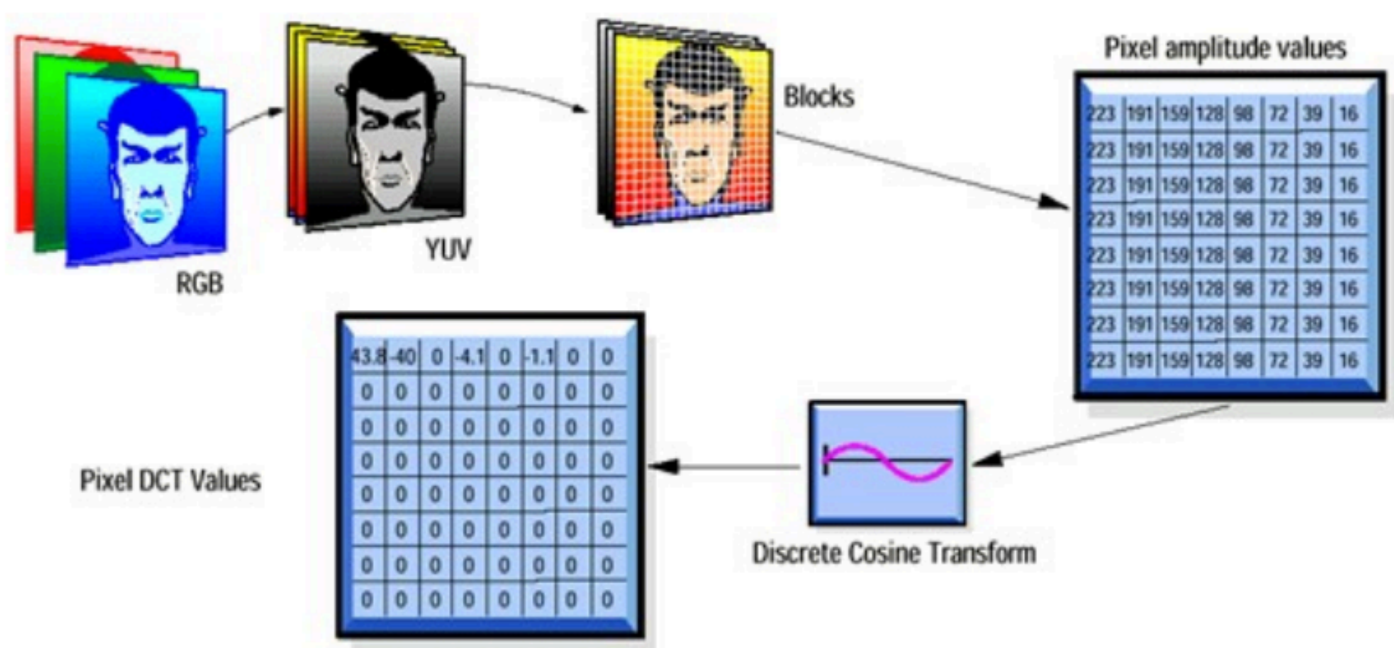
[Bestagini et al. **MMSP 2012**]

Coding-based footprints: multiple compression

- **Video coding:**
  - Temporal redundancy



  - Spatial redundancy



Coding-based footprints: multiple compression

- **Benford's law:**

  - The distribution of the first digit (FD) of a **single quantized** DCT coefficient approximatively follows Benford's law:

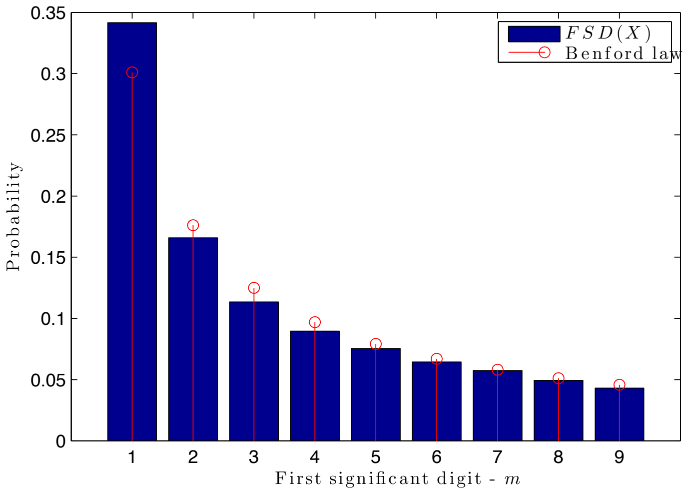
$$p(m) = K \log_{10}\left(1 + \frac{1}{\alpha + m^{\beta}}\right), \text{ with } m = 1, \ldots, 9.$$

  - When **multiple quantized**, this law does not hold!

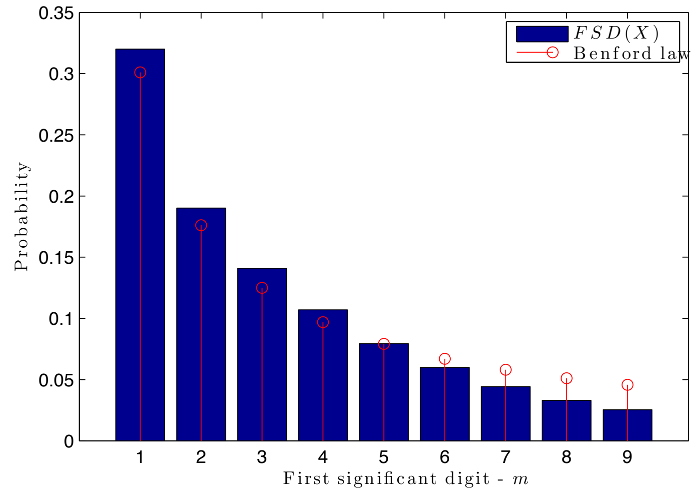
Coding-based footprints: multiple compression

- Single quantized:

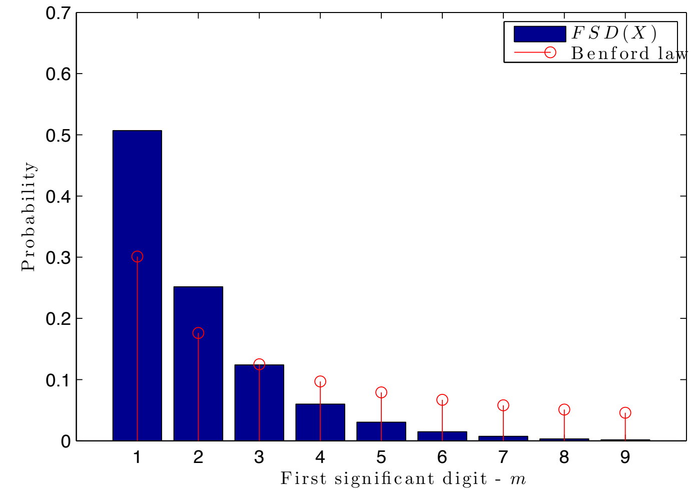
$\Delta = 0.1$



$\Delta = 0.2$

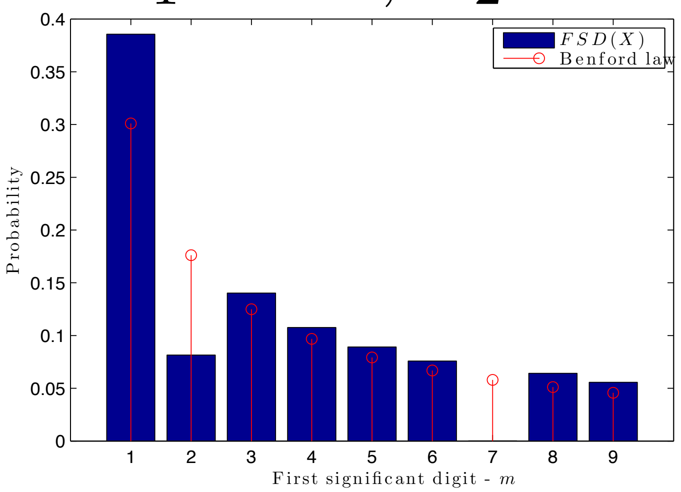


$\Delta = 0.5$

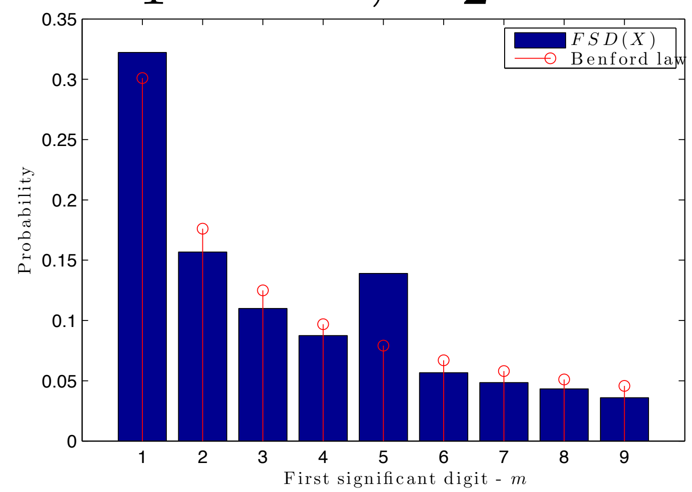


- Double quantized:

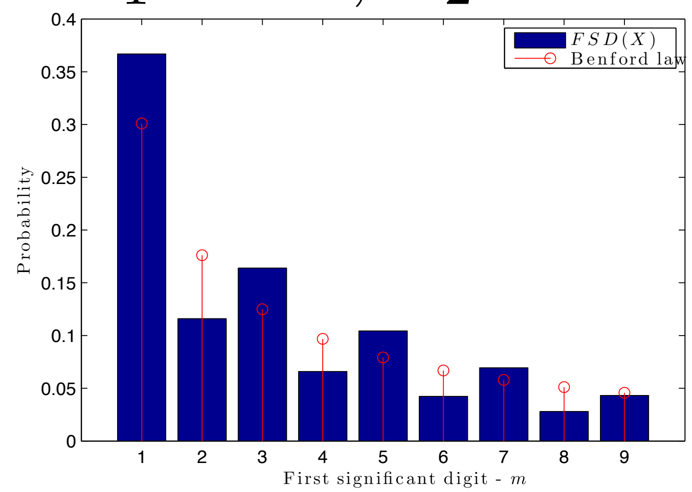
$\Delta_1 = 0.1, \Delta_2 = 0.8$



$\Delta_1 = 0.1, \Delta_2 = 0.11$

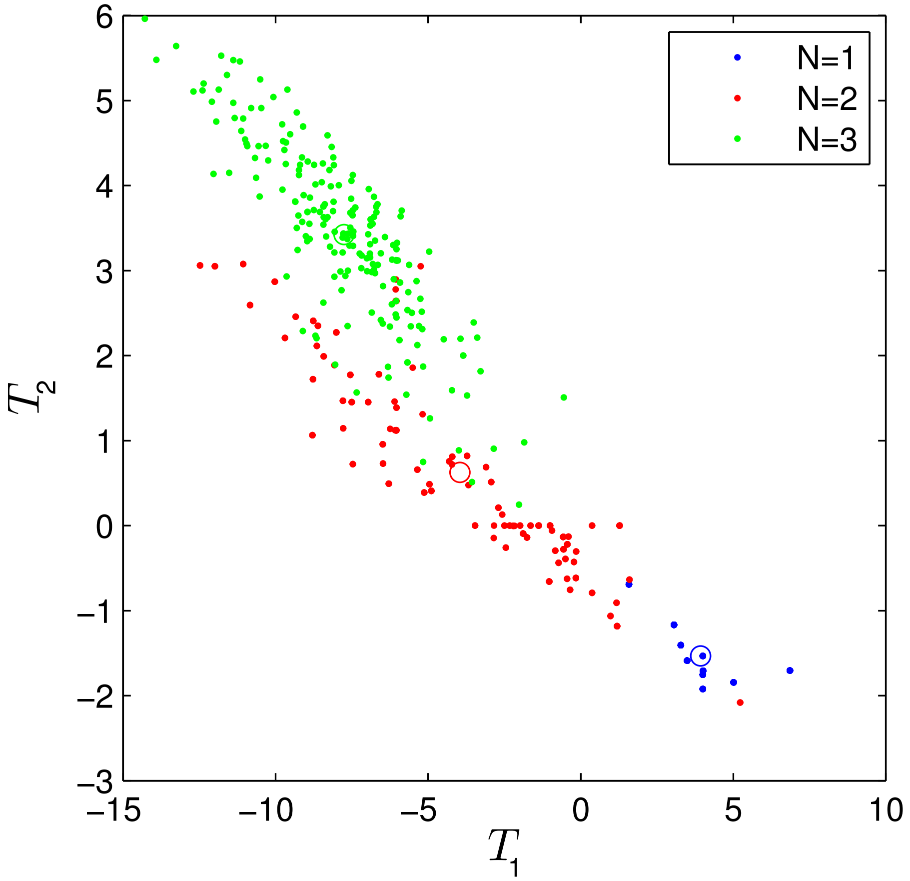


$\Delta_1 = 0.1, \Delta_2 = 0.15$



Coding-based footprints: multiple compression

- **Method 1:**
  - Compute FD histograms for a set of DCT frequencies.
  - Train a set of SVMs.
  - Combine SVMs outputs.



- **Results:**
  - Up to three compressions successfully detected

N , N*	1	2	3
1	100 %	0.00 %	0.00 %
2	0.00 %	73.89 %	26.11 %
3	0.00 %	22.22 %	77.78 %

Which codec has been used to encode a double-compressed video?

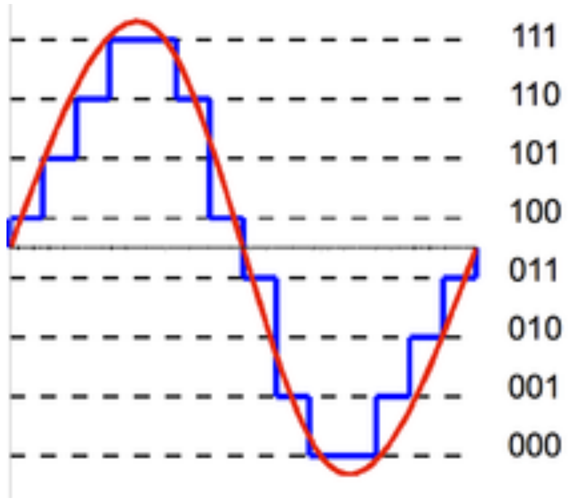
[Bestagini et al. **ICASSP 2012**]
[Bestagini et al. **EUVIP 2013**]
[Bestagini et al. **TIP 2016**]

Coding-based footprints: codec identification

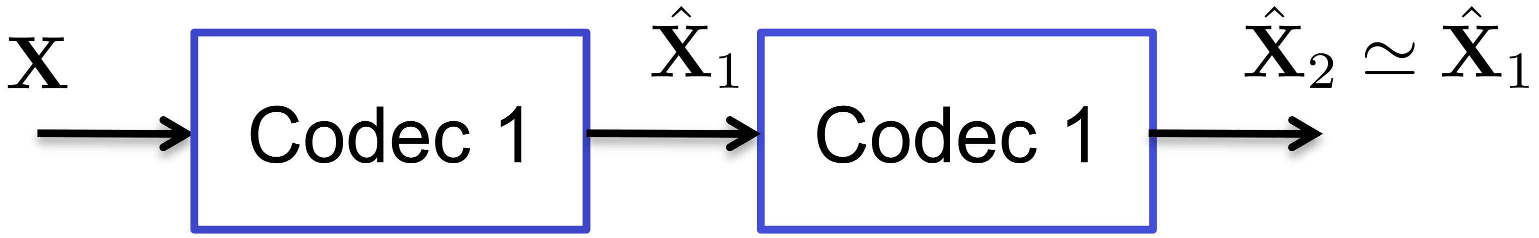
- **Idempotency property:**
  - if we re-quantize an already quantized signal with the same quantization step, the signal does not change

$$\hat{\mathbf{X}}_1 = Q_{\Delta_1}(\mathbf{X}) = \Delta_1 \left\lfloor \frac{\mathbf{X}}{\Delta_1} \right\rfloor$$

$$\hat{\mathbf{X}}_2 = Q_{\Delta_2}(\hat{\mathbf{X}}_1) = \Delta_1 \left\lfloor \frac{\Delta_1 \left\lfloor \frac{\mathbf{X}}{\Delta_1} \right\rfloor}{\Delta_1} \right\rfloor = \hat{\mathbf{X}}_1$$



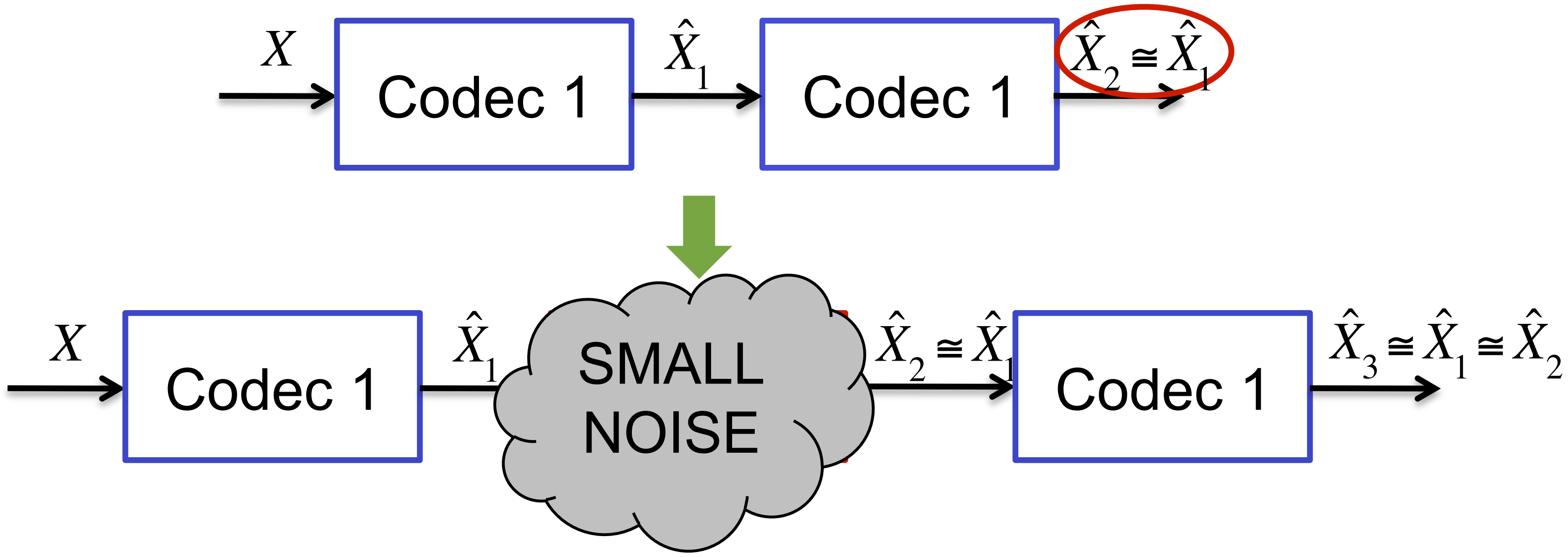
- This is partly true also for video codecs



$\mathbf{X}$ → Codec 1 → $\hat{\mathbf{X}}_1$ → Codec 1 → $\hat{\mathbf{X}}_2 \simeq \hat{\mathbf{X}}_1$

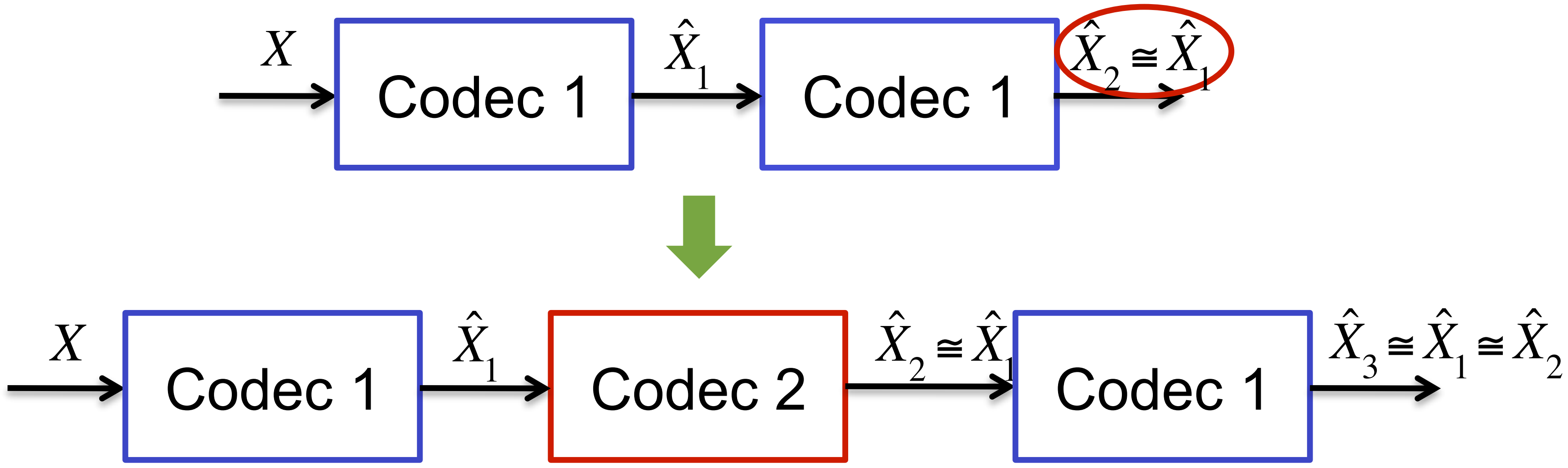
Coding-based footprints: codec identification

- **Main idea:**



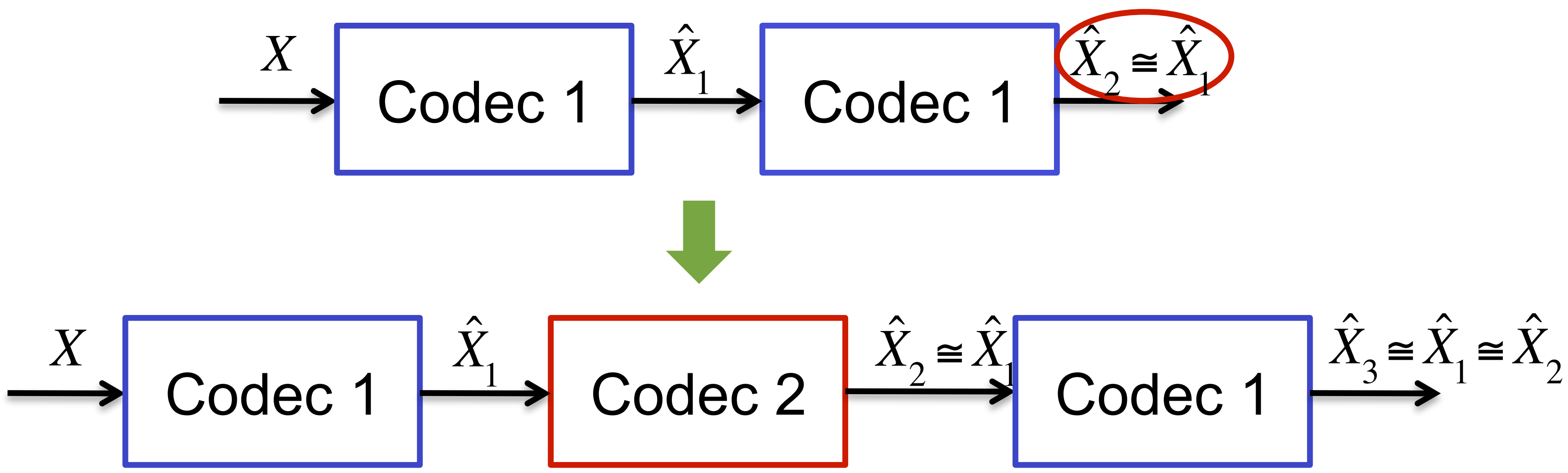
Coding-based footprints: codec identification

- **Main idea:**

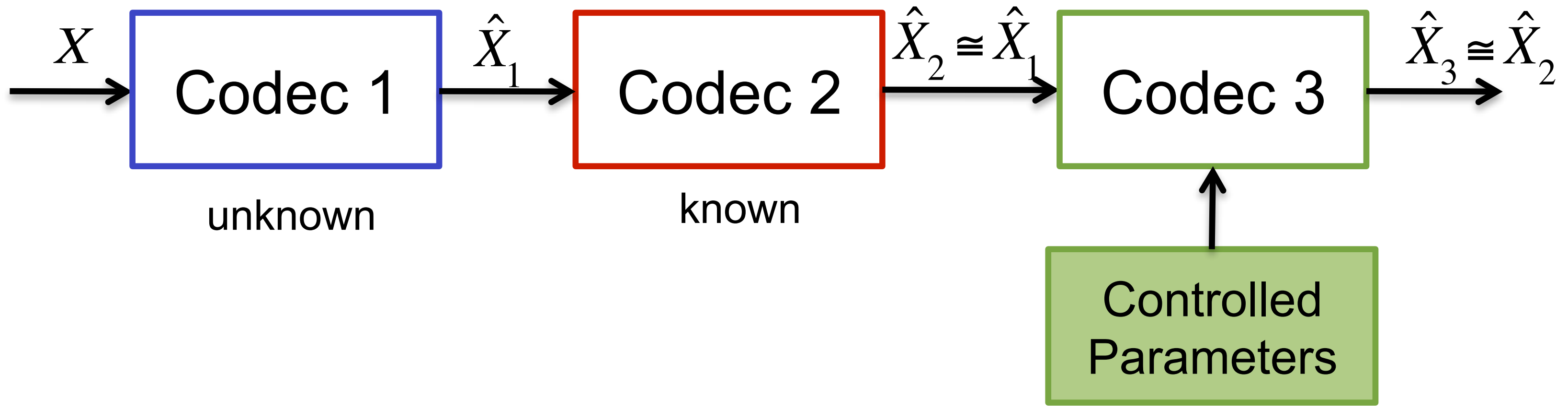


Coding-based footprints: codec identification

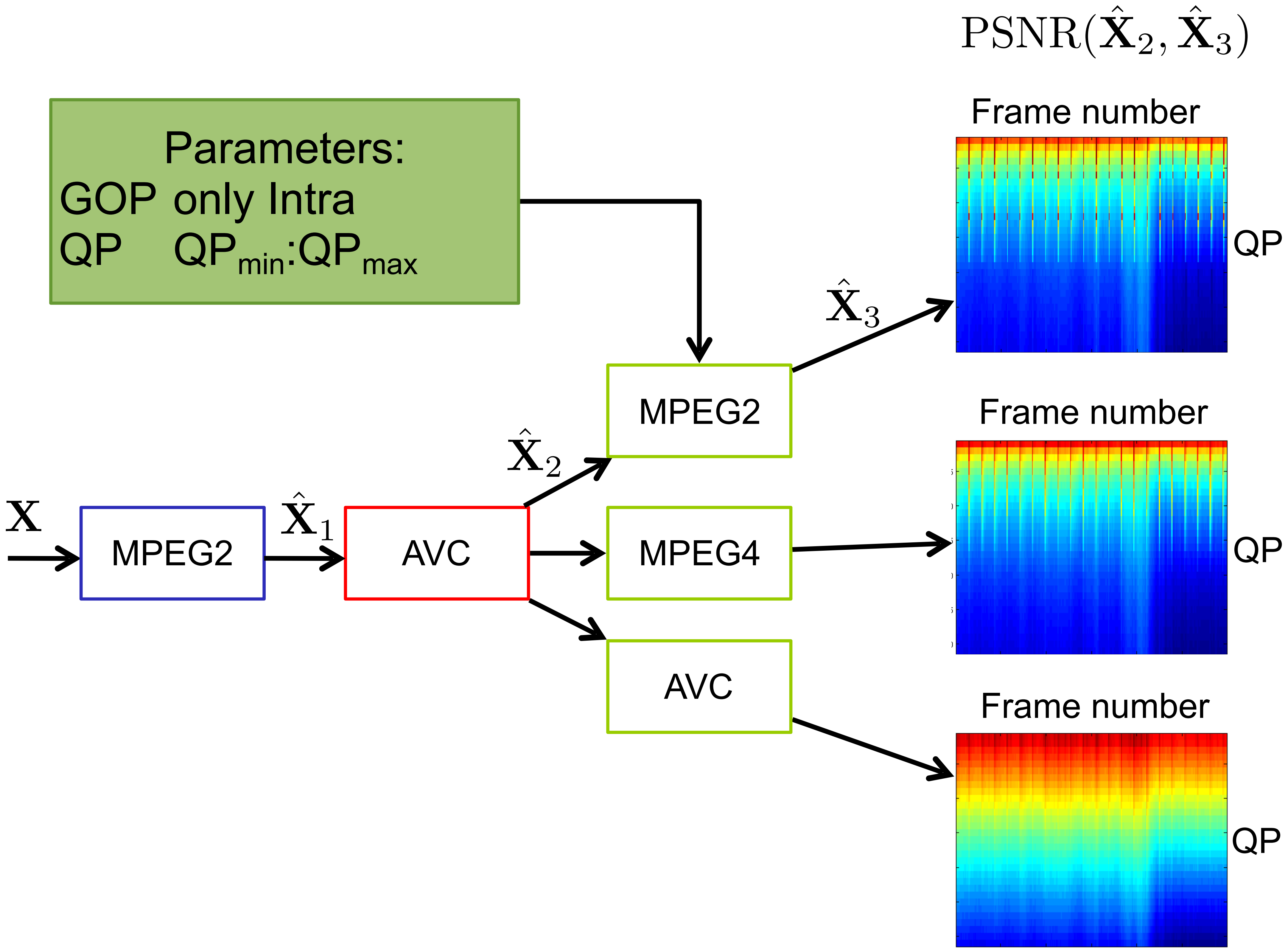
- **Main idea:**



- **Approach:**

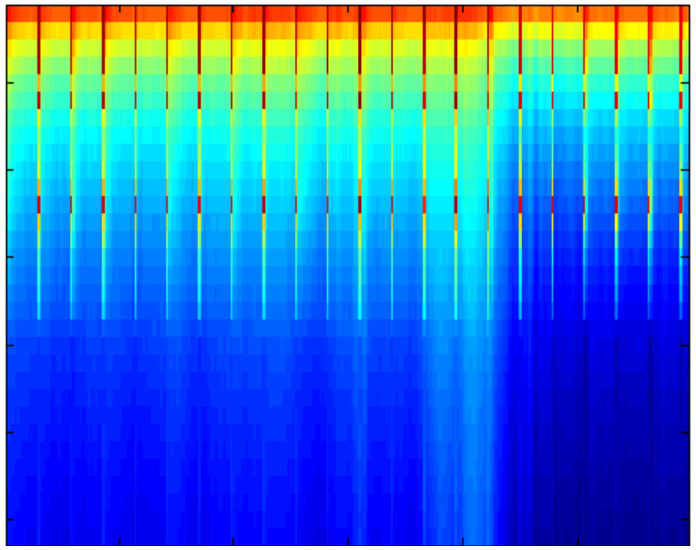


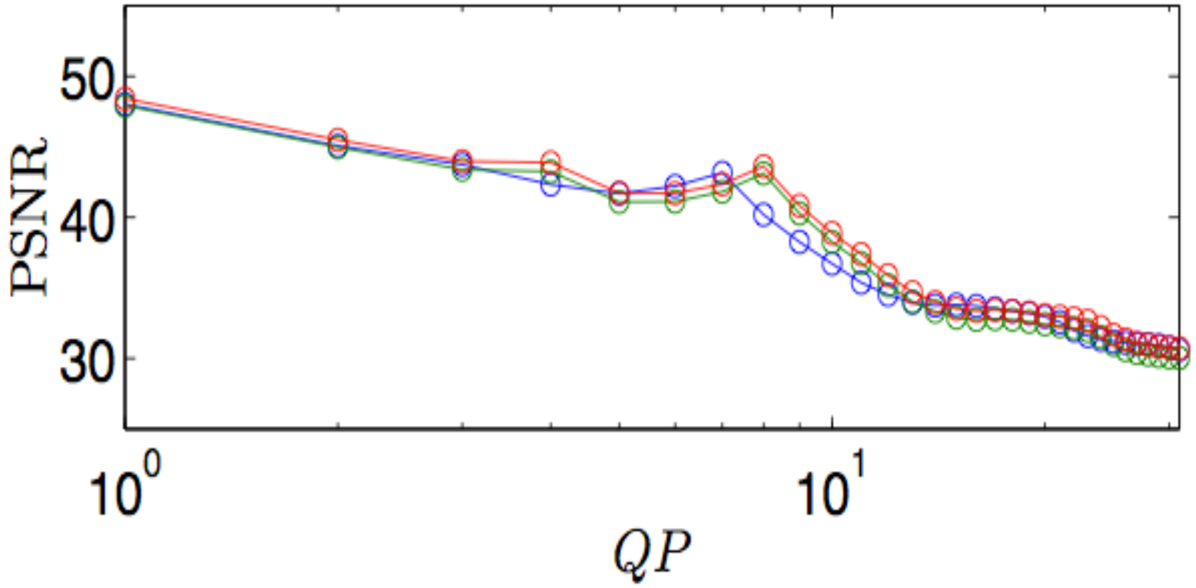
Coding-based footprints: codec identification



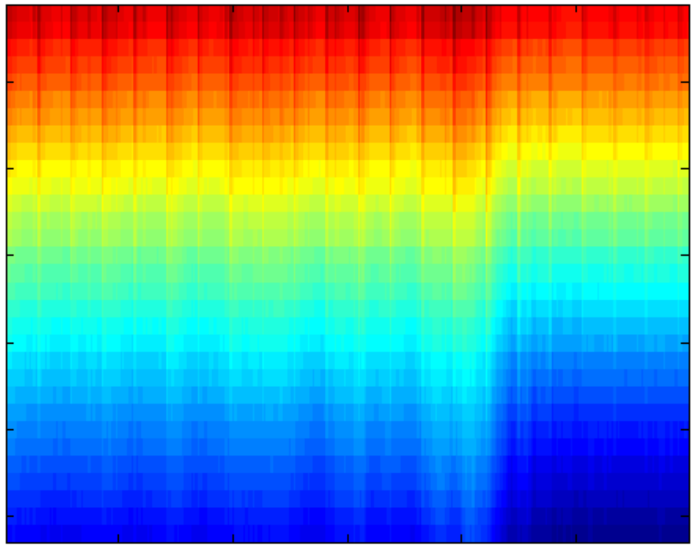
Coding-based footprints: codec identification

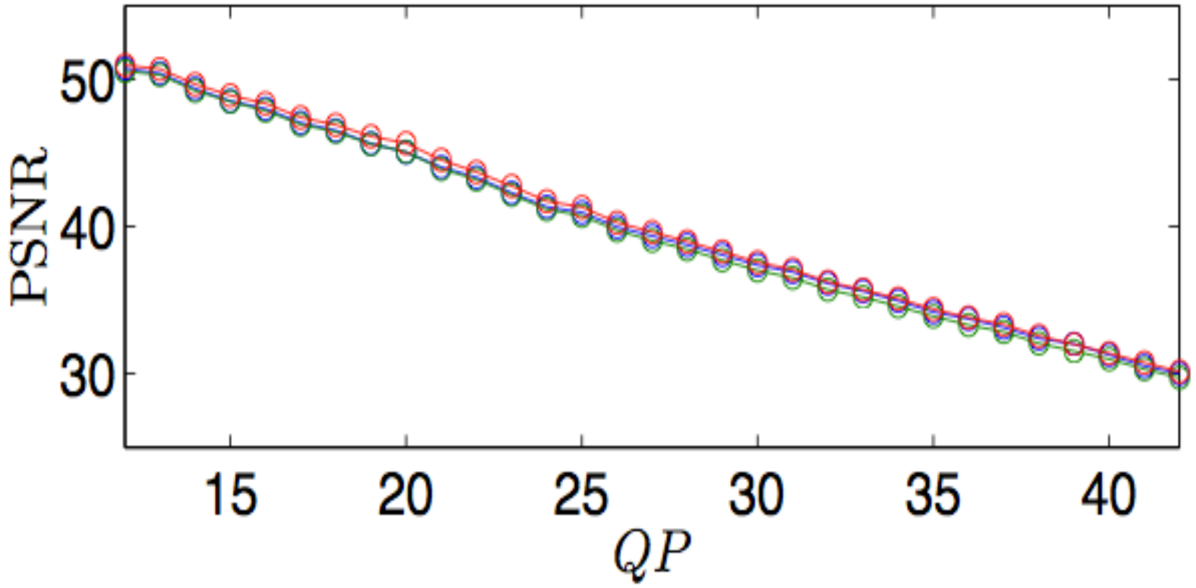
Correct



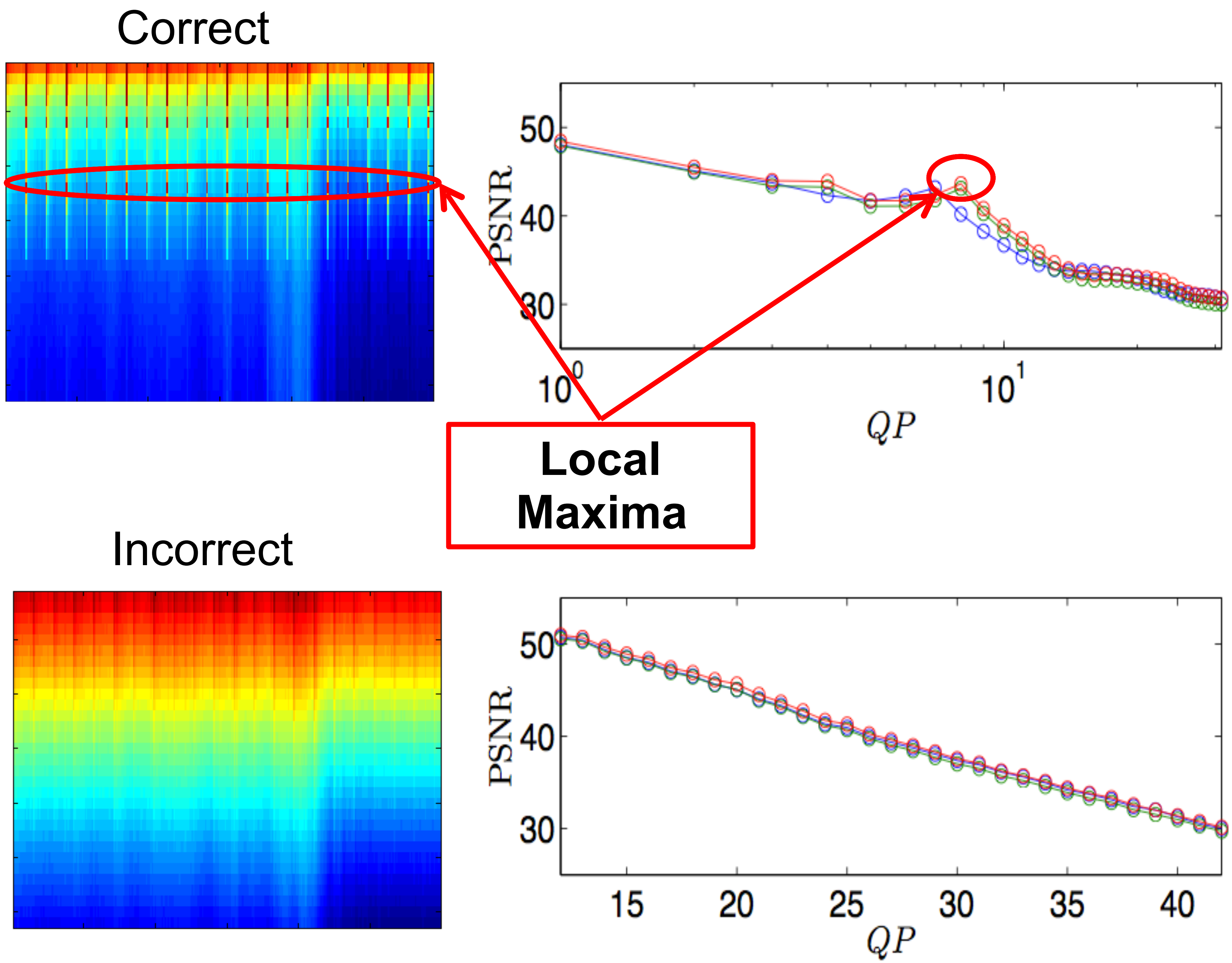


Incorrect





Coding-based footprints: codec identification



Coding-based footprints: codec identification

- **Results:**

c_1 \ $\hat{c}_1$	MPEG-2			MPEG-4			AVC			DIRAC		
c_2	MPEG-2	MPEG-4	AVC	MPEG-2	MPEG-4	AVC	MPEG-2	MPEG-4	AVC	MPEG-2	MPEG-4	AVC
MPEG-2	**0.94**	**0.96**	**0.96**	0.05	0.04	0.03	0	0	0.01	0.01	0	0
MPEG-4 (a)	0	0	0.04	**0.93**	**0.92**	**0.76**	0.02	0.02	0.2	0.06	0.06	0
MPEG-4 (b)	0.02	0.02	0.06	**0.87**	**0.87**	**0.69**	0.06	0.05	0.25	0.06	0.06	0
AVC (a)	0.01	0	0	0.14	0.24	0.06	**0.79**	**0.68**	**0.94**	0.06	0.08	0
AVC (b)	0	0.01	0	0.13	0.2	0.05	**0.81**	**0.69**	**0.94**	0.06	0.09	0.01
AVC (c)	0.06	0.06	0.15	0	0.03	0	**0.92**	**0.87**	**0.81**	0.02	0.05	0.04
DIRAC	0	0.02	0	0.09	0.12	0.01	0.13	0.12	0.21	**0.78**	**0.74**	**0.78**

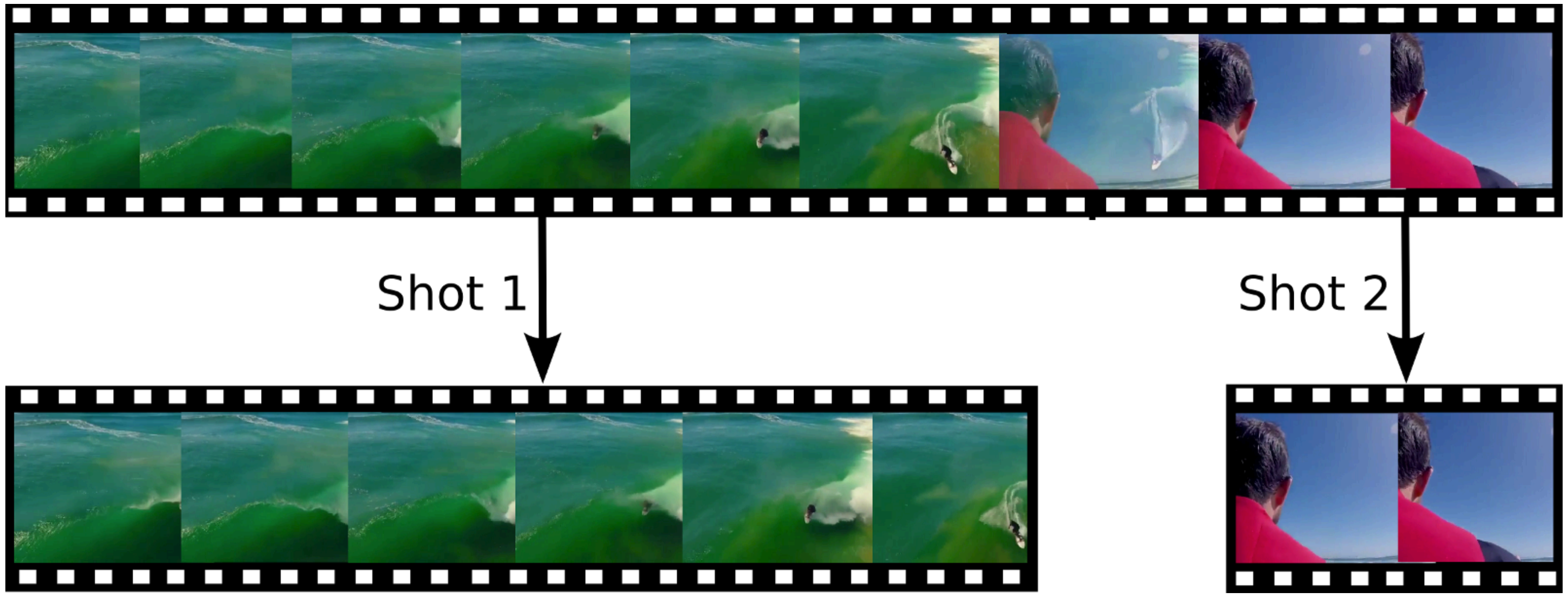
Are **codec** and **quality** coherent in time?

[Verde et al. **ICIP 2018**]

Coding-based footprints: temporal coherence

Problem

- Given a decoded video sequence, detect whether it is a compilation from multiple video shots.



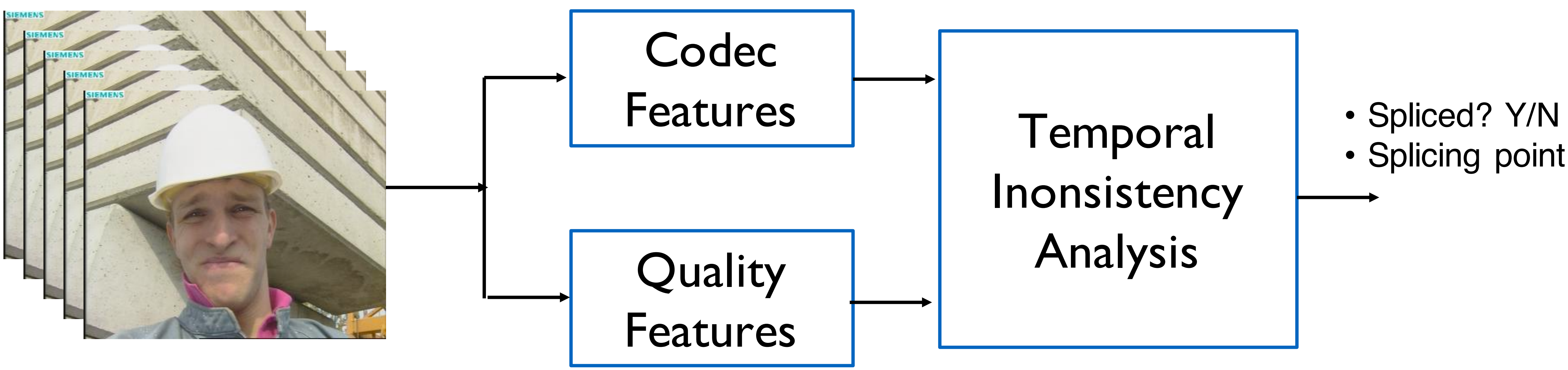
Assumptions

- Shots are seldom originally encoded with the exact same codec or parameters due to different sources and used software.

Coding-based footprints: temporal coherence

Main Pipeline

- Compute a frame-wise indicator of the used codec
- Compute a frame-wise indicator of the video quality
- Check inconsistency of these indicators frame-by-frame



Coding-based footprints: temporal coherence

Feature Extraction

Codec Features

- A CNN is trained to classifiy 4 different codecs (MPEG2, MPEG4, H264, H265)
- Feature vector is $\mathbf{f}_{\mathrm{C}}^{p}(n) = [f_{\mathrm{H264}}^{p}(n),\ f_{\mathrm{H265}}^{p}(n),\ f_{\mathrm{MPEG2}}^{p}(n),\ f_{\mathrm{MPEG4}}^{p}(n)]$

Quality Features

- A CNN is trained to classifiy 4 different qualities (H264 with QP=5, 10, 15, 20)
- Feature vector is $\mathbf{f}_{\mathrm{Q}}^{p}(n) = [f_{\mathrm{low}}^{p}(n),\ f_{\text{m-low}}^{p}(n),\ f_{\text{m-high}}^{p}(n),\ f_{\mathrm{high}}^{p}(n)]$

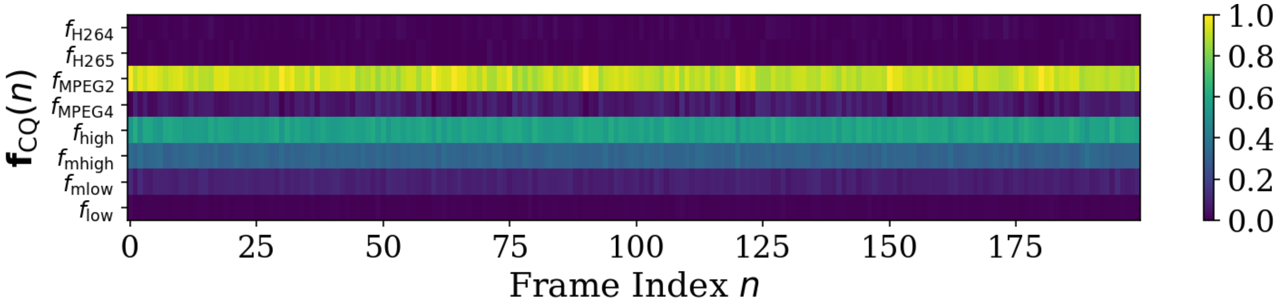
Coding-based footprints: temporal coherence

Temporal inconsistency analysis

Feature Merge

- Feature vectors are concatenated into a single one

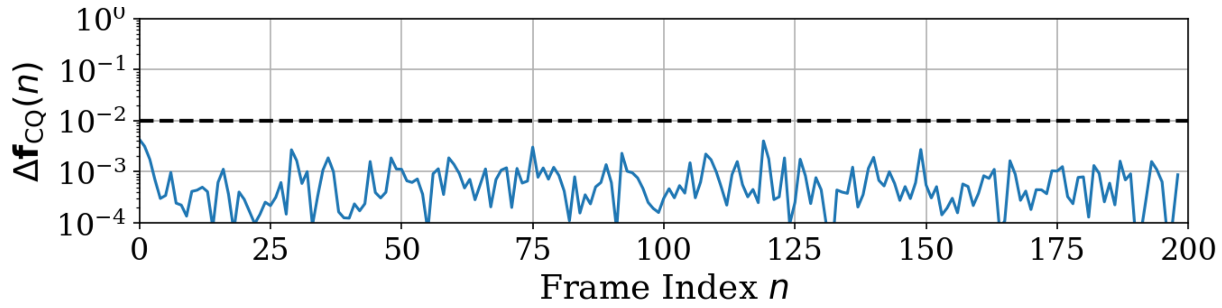
$$\mathbf{f}_{CQ}(n) = [\mathbf{f}_C(n),\ \mathbf{f}_Q(n)]$$



Time Analysis

- Compute MSE between feature pairs

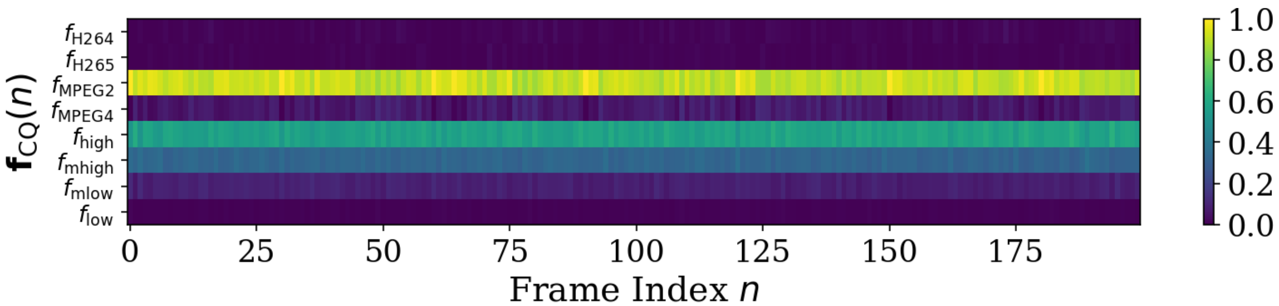
$$\Delta\mathbf{f}_{CQ}(n) = \mathrm{MSE}(\mathbf{f}_{CQ}(n),\ \mathbf{f}_{CQ}(n+1))$$



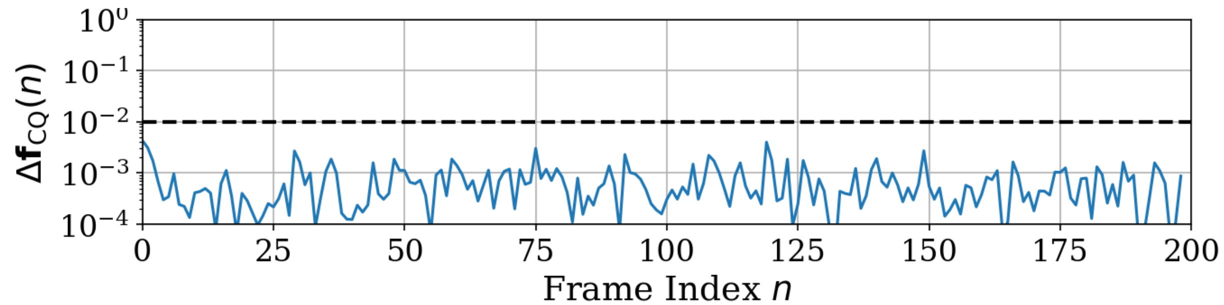
Threshold MSE

Coding-based footprints: temporal coherence

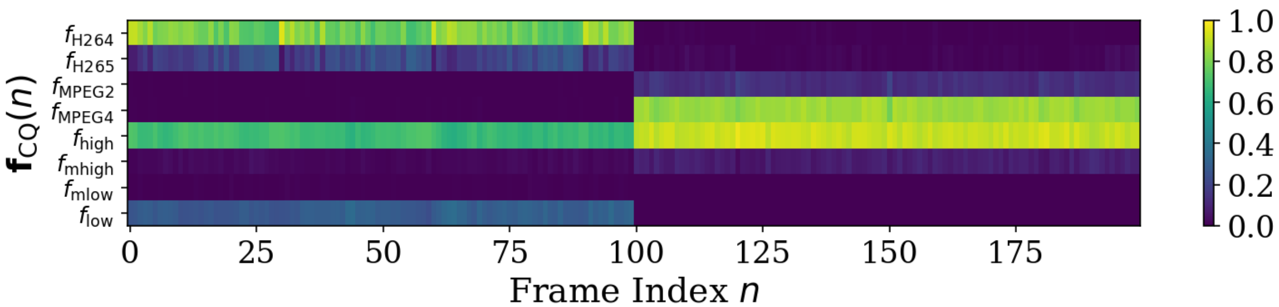
Visual example



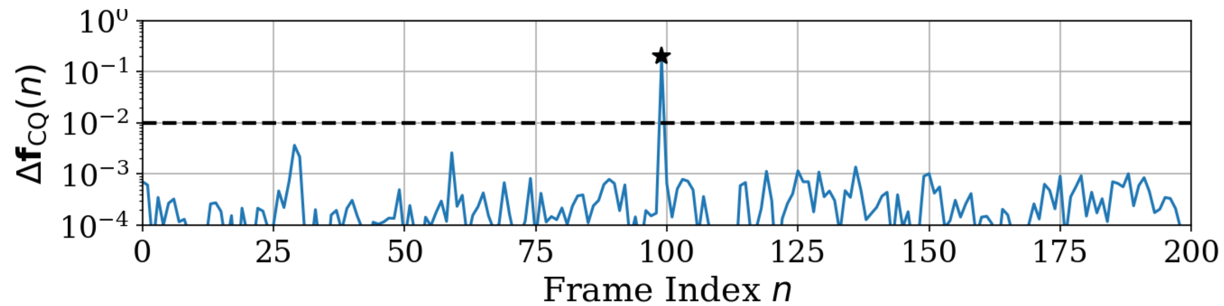
(a) Pristine



(a) Pristine



(b) Composition



(b) Composition

Coding-based footprints: temporal coherence

Challenging Example

Coding-based footprints: temporal coherence

Challenging Example

Coding-based footprints: temporal coherence

Video Codec Identification Results

	H264	H265	MPEG2	MPEG4
H264	0.76	0.20	0.02	0.02
H265	0.12	0.85	0.01	0.02
MPEG2	0.02	0.01	0.84	0.13
MPEG4	0.05	0.03	0.03	0.89

Predicted

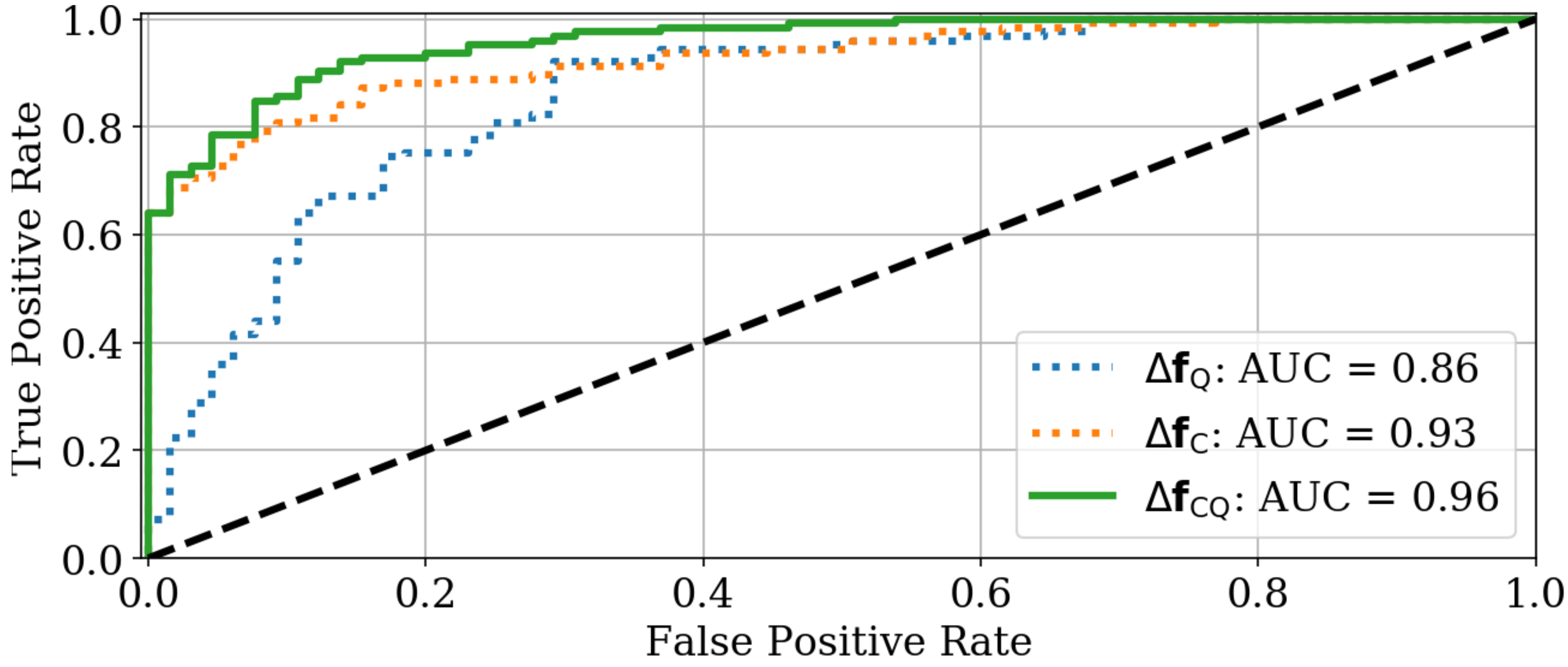
Video Quality Identification Results

	high	m-high	m-low	low
high	0.84	0.12	0.04	0.00
m-high	0.09	0.83	0.08	0.00
m-low	0.03	0.07	0.85	0.04
low	0.00	0.00	0.04	0.96

Predicted

Coding-based footprints: temporal coherence

Splicing Detection Results



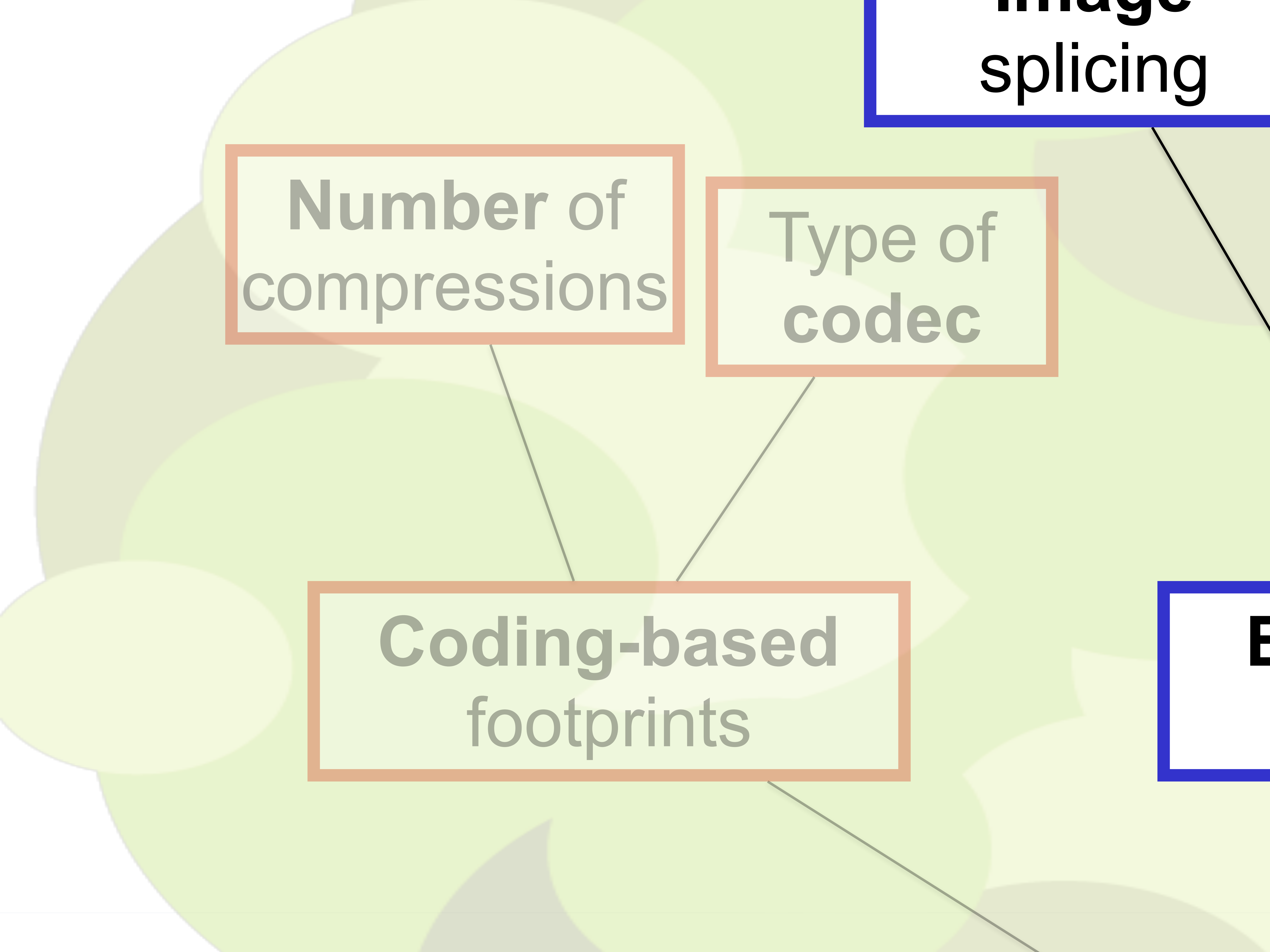


Image splicing
Number of compressions
Type of codec
Coding-based footprints

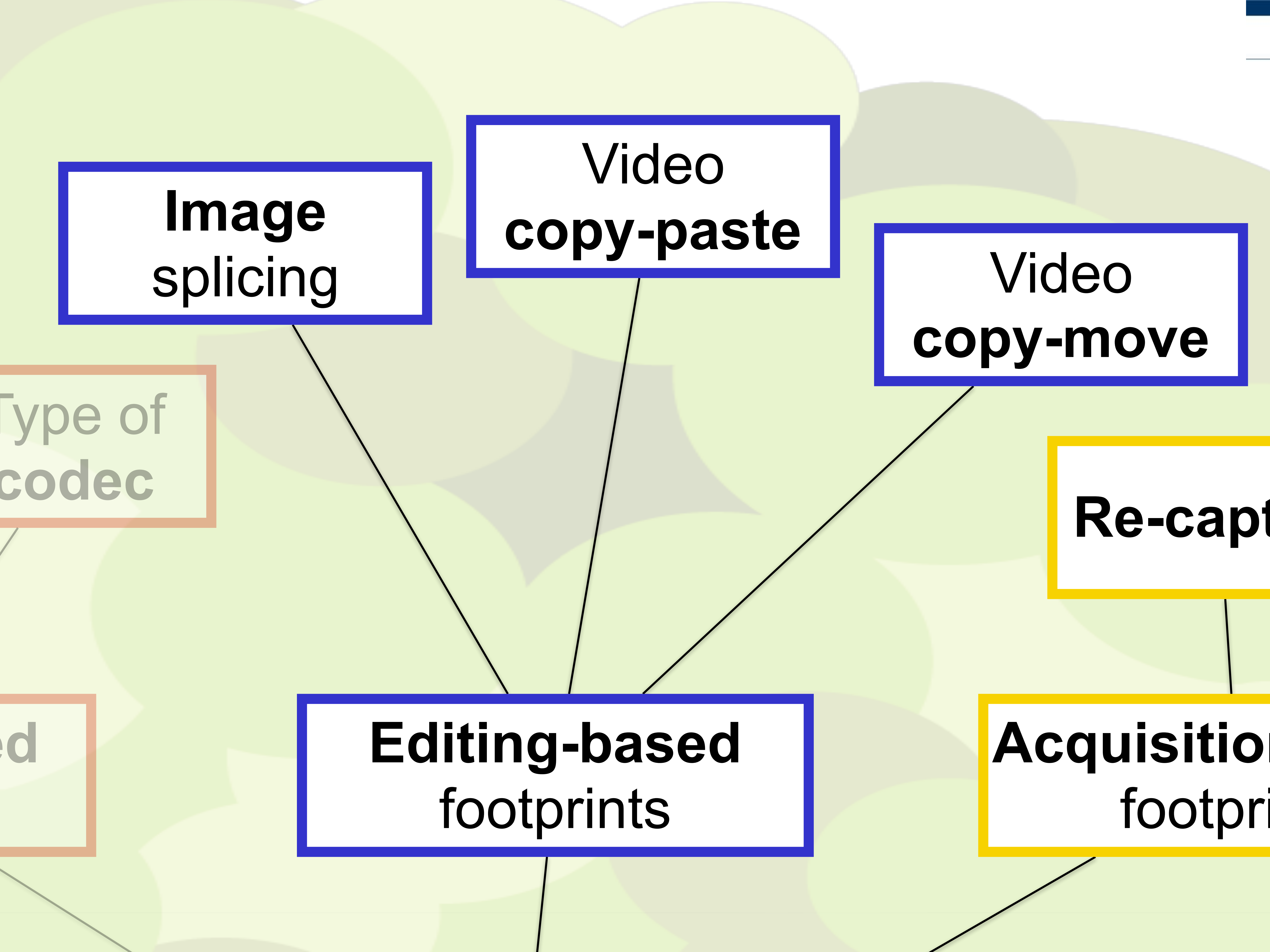


Image splicing
Video copy-paste
Video copy-move
Type of codec
Re-capt
ed
Editing-based footprints
Acquisitio footpri

Editing-based footprints

Editing-based footprints

Editing-based footprints

Editing-based footprints

Editing-based footprints

Forged

Original

Editing-based footprints

Forged

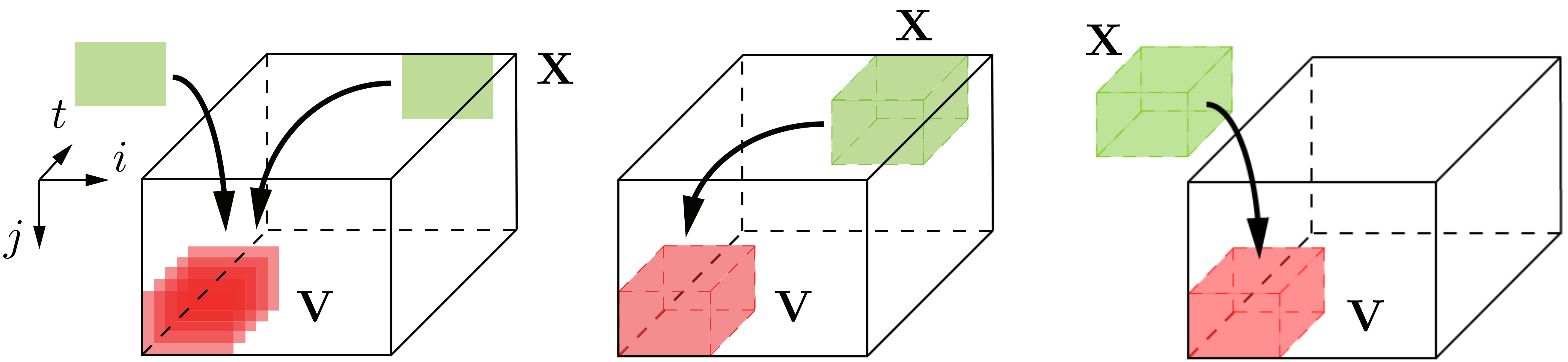
Original

Is the video **forged?**

[Bestagini et al. **ICASSP 2013**]
[Bestagini et al. **MMSP 2013**]

Editing-based footprints

- Video **forgeries** can be operated through different kinds of editing attacks
- We considered:
  1. insertion of a **still image**
  2. insertion of a portion of **video from the same source**
  3. insertion of a portion of **video from a different source**



Editing-based footprints: image copy-paste

- **Problem**:
  - An **image is inserted** and repeated in time
- **Method:**
  - Exploit characteristic residual between adjacent frames

Original Tampered

(a) Original residual (b) Tampered residual (c) Tampered and compressed residual

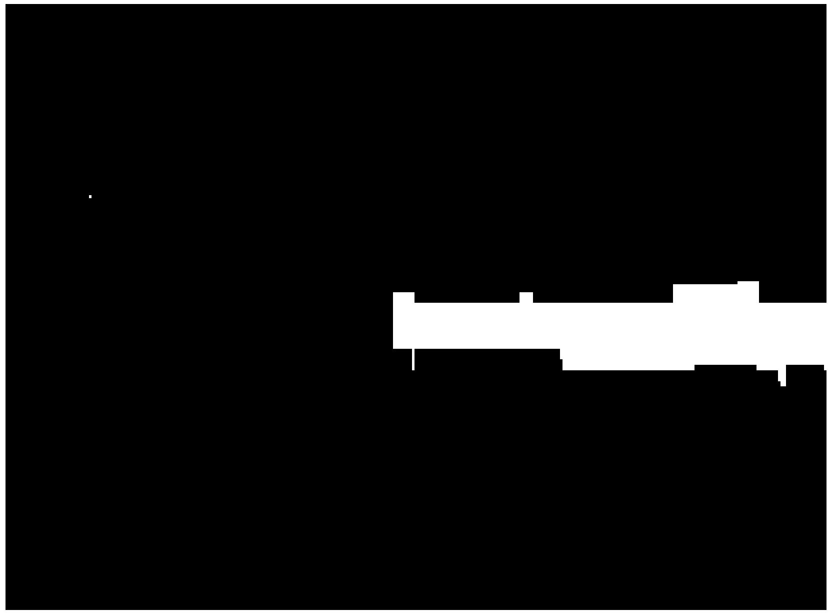
Editing-based footprints: image copy-paste

ALGORITHM:

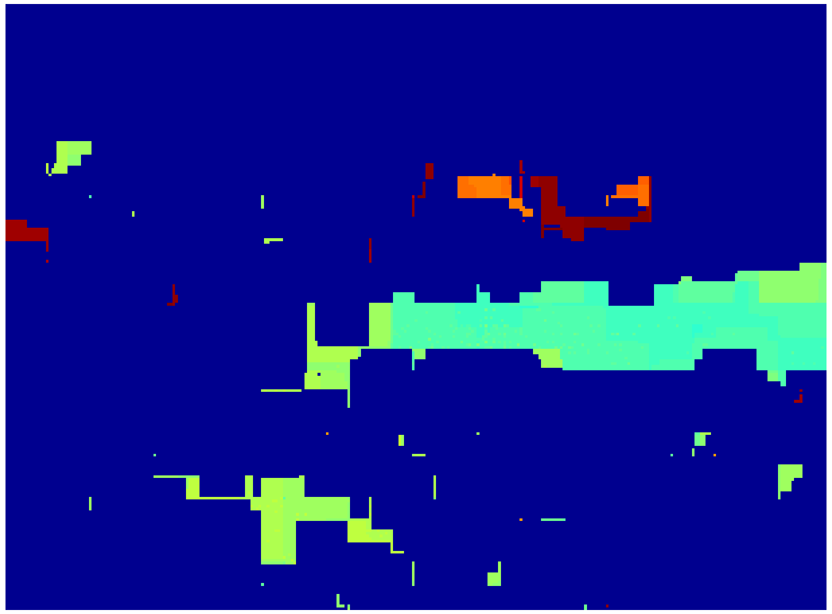
- Define the residual $r^t_{i,j} = x^t_{i,j} - x^{t+1}_{i,j}$
  - Zero for possibly tampered pixels
- Define the residual mask as $m^t_{i,j} = \begin{cases} 1 & \text{if } r^t_{i,j} = 0, \\ 0 & \text{otherwise,} \end{cases}$
  - One for possibly tampered pixels
- Apply an erosion with a Structuring Element $\mathbf{H}^{di,dj,dt}$ and obtain $\mathbf{E} = \{e^t_{i,j}\} = \mathbf{M} \ominus \mathbf{H}^{di,dj,dt}$,
  - Remove small areas
- Compute the feature vector $F_{i,j} = [f^1_{i,j}, f^2_{i,j}]$
  - $f^1_{i,j}$ cardinality of the longest set of ones in *(i,j)*
  - $f^2_{i,j}$ starting *t* value of the longest set of ones
- Search the longest set of ones starting from the same *t*

Editing-based footprints: image copy-paste

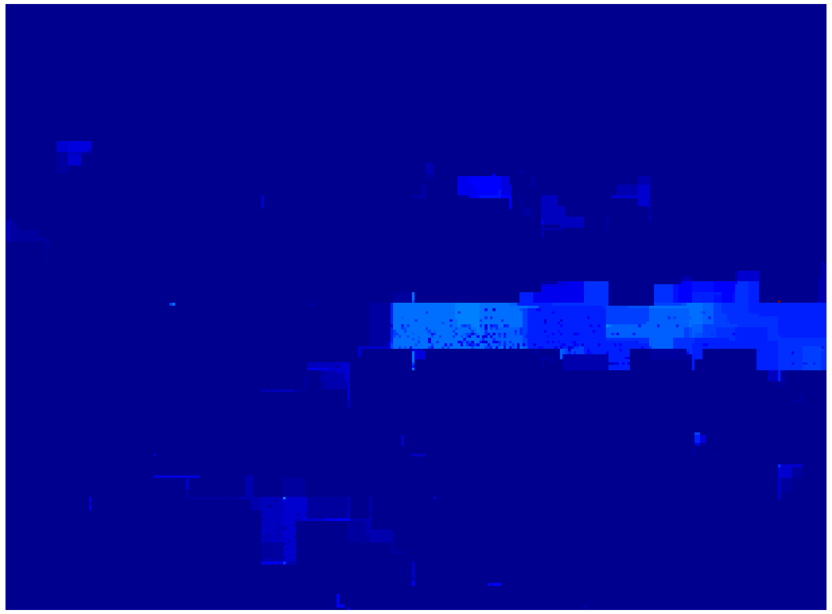
(a) Tampered frame



(b) Detected mask



(c) $f^1_{i,j}$



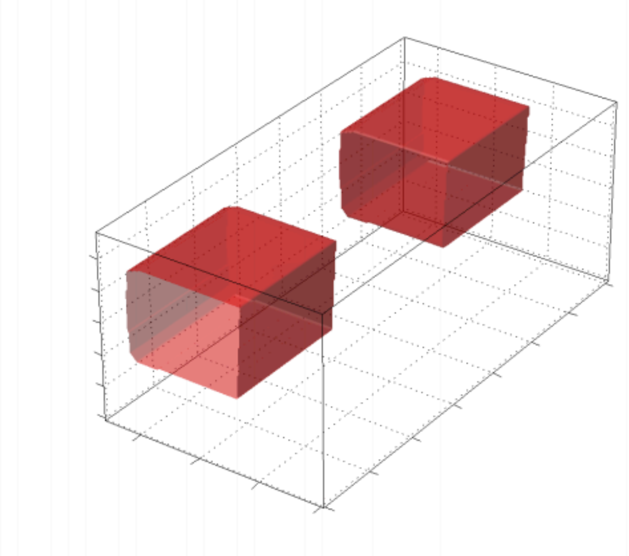
(d) $f^2_{i,j}$

Editing-based footprints: image copy-paste

- **Problem:**
  - A **video is inserted** from the same sequence

Method:

- Implementation of an automatic correlation analysis to detect local duplication

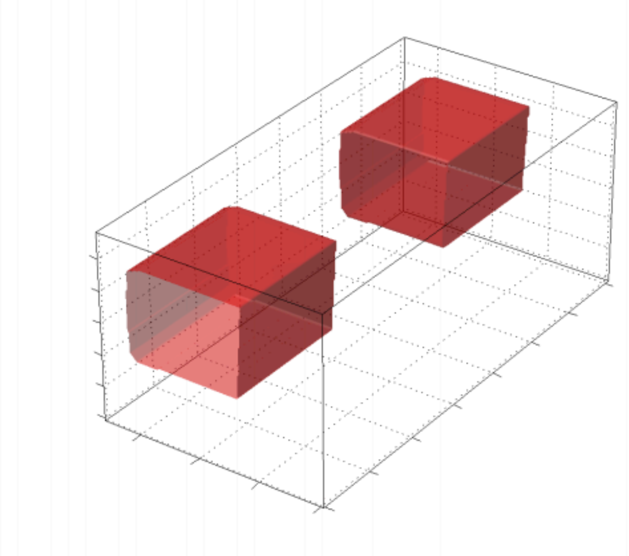


Editing-based footprints: image copy-paste

- **Problem:**
  - A **video is inserted** from the same sequence

Method:

- Implementation of an automatic correlation analysis to detect local duplication



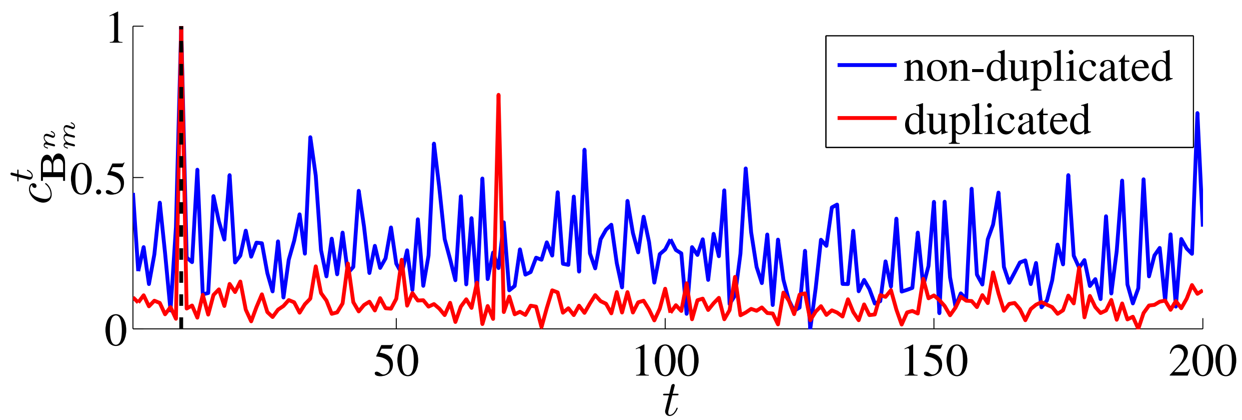
Editing-based footprints: image copy-paste

Algorithm:

- Compute the residual $\mathbf{R} = \{r_{i,j}^t\}$
- Divide the residual into non-overlapping 3D blocks $\mathbf{B}_m^n$
- Compute the phase correlation

$$\mathbf{C}_{i,j}^t(\mathbf{B}_m^n) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mathbf{B}_m^n)\mathcal{F}(\mathbf{R})^*}{|\mathcal{F}(\mathbf{B}_m^n)\mathcal{F}(\mathbf{R})^*|}\right)$$

- Compute the maximum correlation value for each time position $c_{\mathbf{B}_m^n}^t = \max_{i,j}\left(|\mathbf{C}_{i,j}^t(\mathbf{B}_m^n)|\right)$

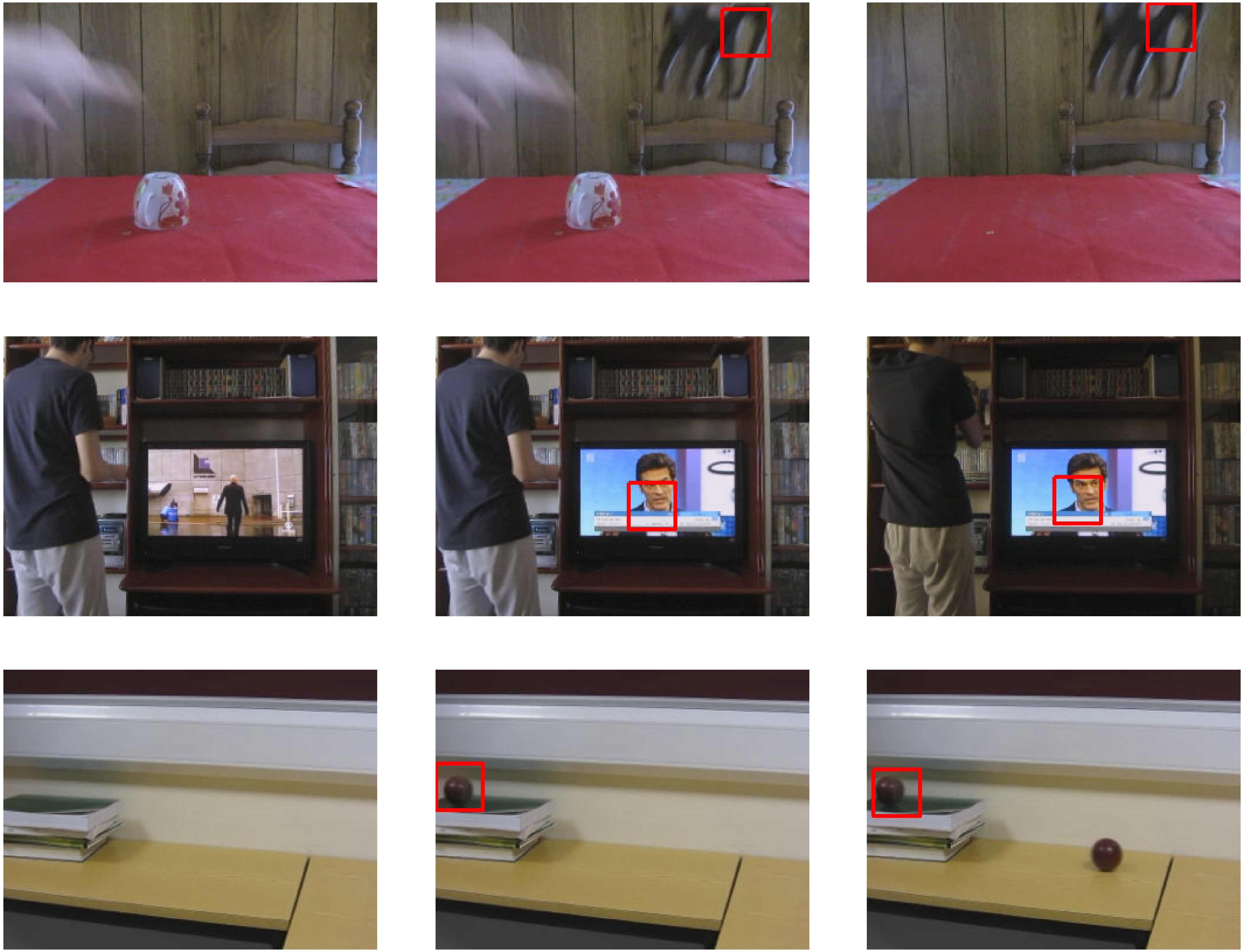


- Search for peaks indicating duplication by thresholding the max-mean ratio

$$p_{\mathbf{B}_m^n} = \frac{\max(c_{\mathbf{B}_m^n}^t)}{\frac{1}{(T-1)}\sum_t c_{\mathbf{B}_m^n}^t}$$

- Check if the detected duplicated block is similar to its original version (MSE)

Editing-based footprints: image copy-paste

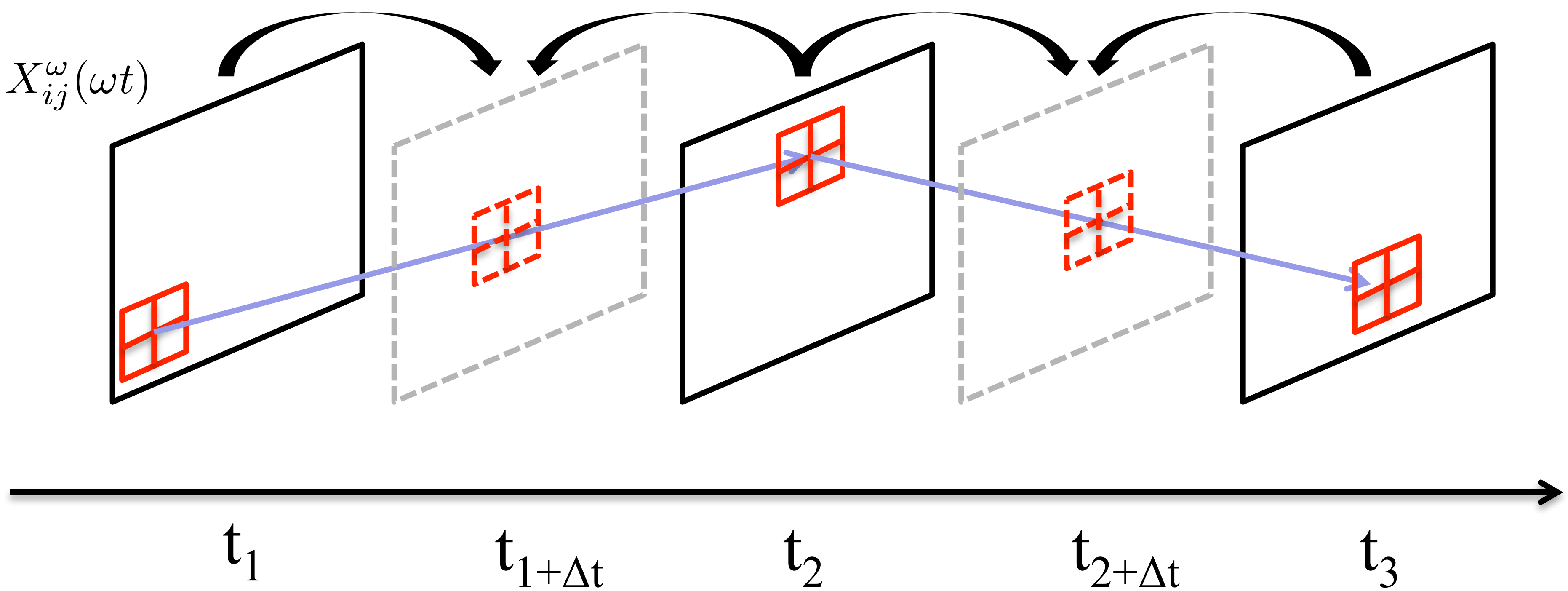


original | forged | detected duplication

Editing-based footprints: image copy-paste

- **Problem:**
  - A **video is inserted** from ~~the same~~ a different sequence at with different frame-rate

- **Method:**
  - Search for traces left by frame-rate equalisation
  - Up-sampling and down-sampling leave a characteristic pixel correlation in time



Editing-based footprints: image copy-paste

Algorithm:

- Estimate each frame from their neighbors
  - Compute motion vectors
  - Average frames

- Compute the prediction error
  - Original frames ➔ high error
  - Predicted frames ➔ low error

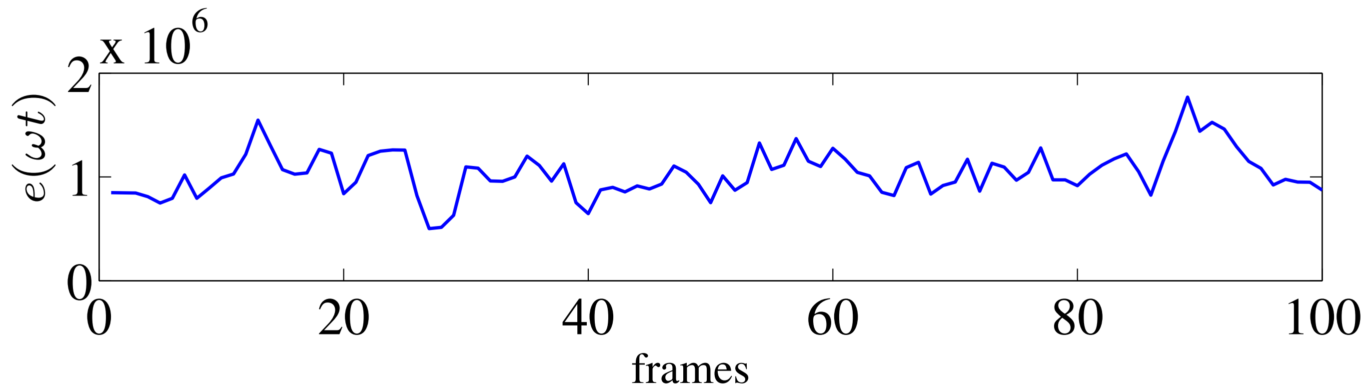
$$e_{ij}(\omega t) = X_{ij}^{\omega}(\omega t) - \sum_{k=-K}^{K} h_k^* \cdot X_{m_{t,i,j} n_{t,i,j}}^{\omega}(\omega t + \omega k)$$

Analysis filter (h_k^*)   Estimated MVs ($m_{t,i,j} n_{t,i,j}$)

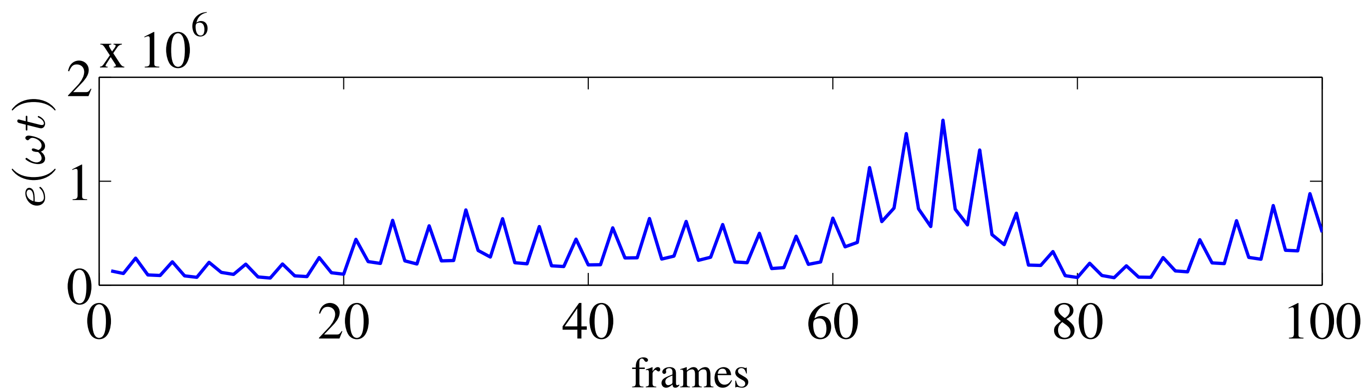
$$e(\omega t) = \sum_{ij} |e_{ij}(\omega t)|^2$$

- Estimate error periodicity (spectral analysis)
  - If non-periodic ➔ not interpolated
  - If periodic ➔ interpolated

Editing-based footprints: image copy-paste

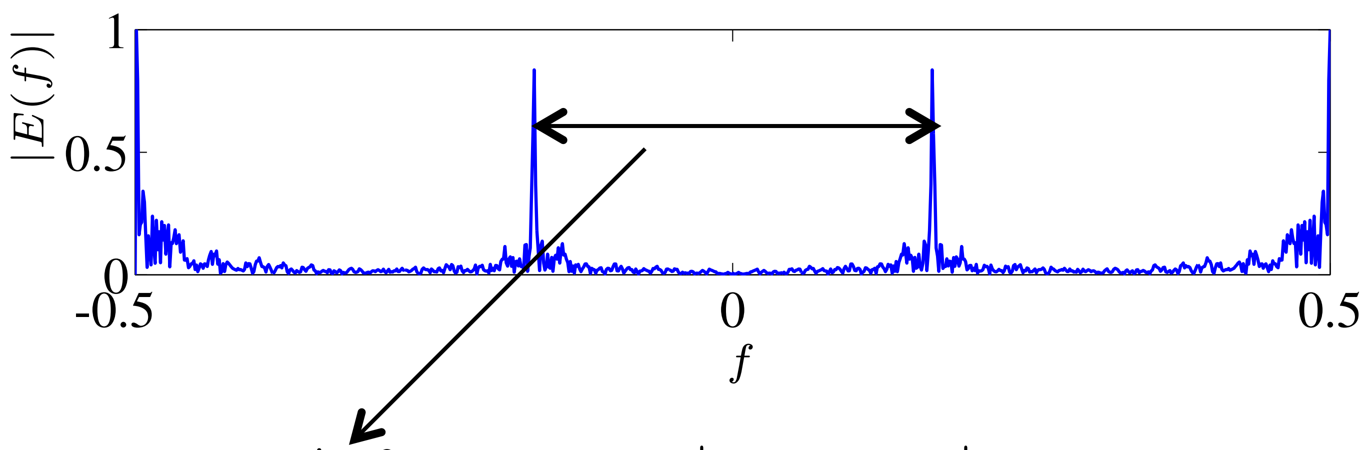


Foreman (original)



Foreman (interpolated)

Fps: 30 ➔90



Foreman (interpolated)
Spectral analysis

$$\Delta f = 0.5 - |\omega - 0.5|,$$

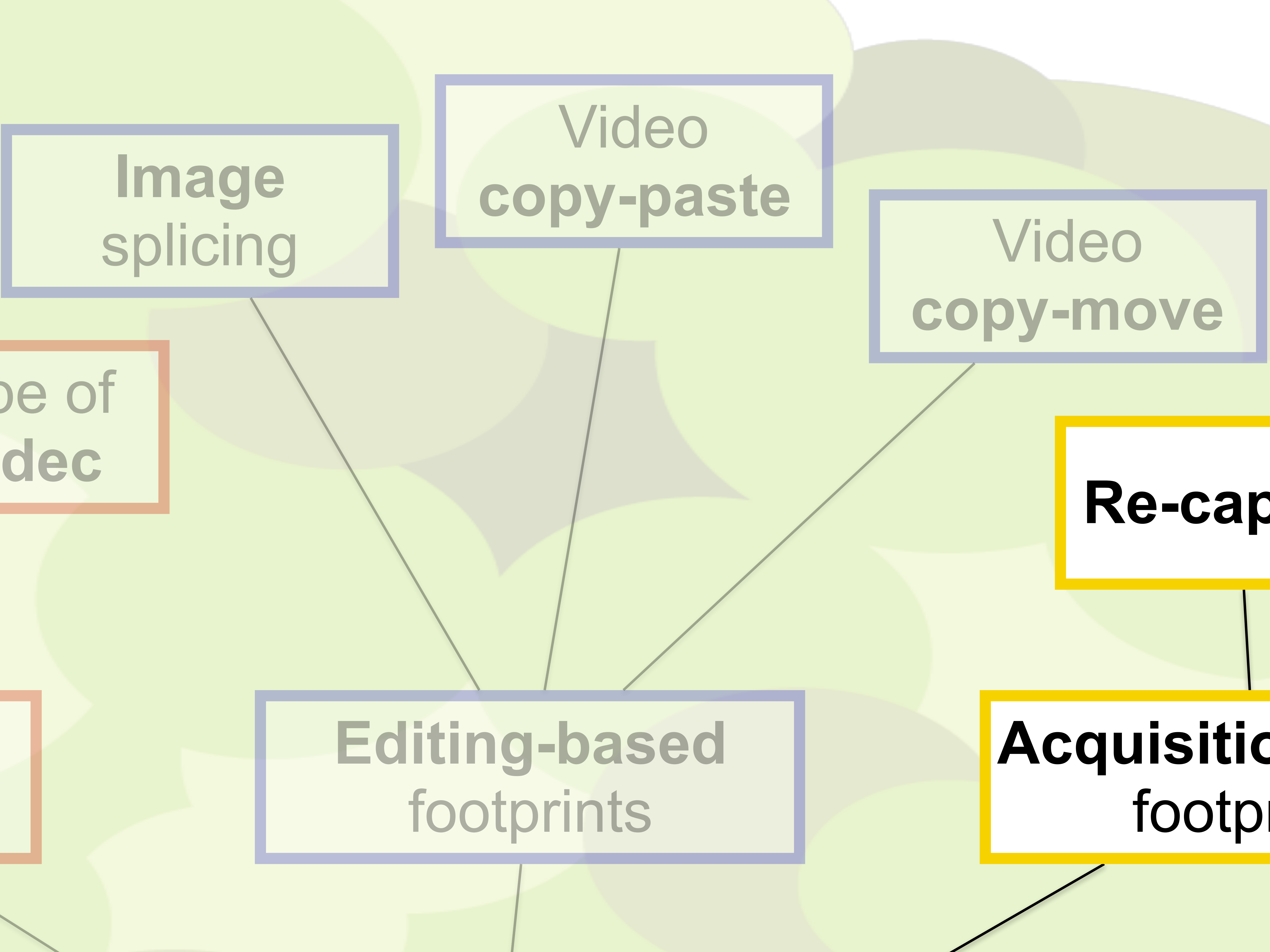
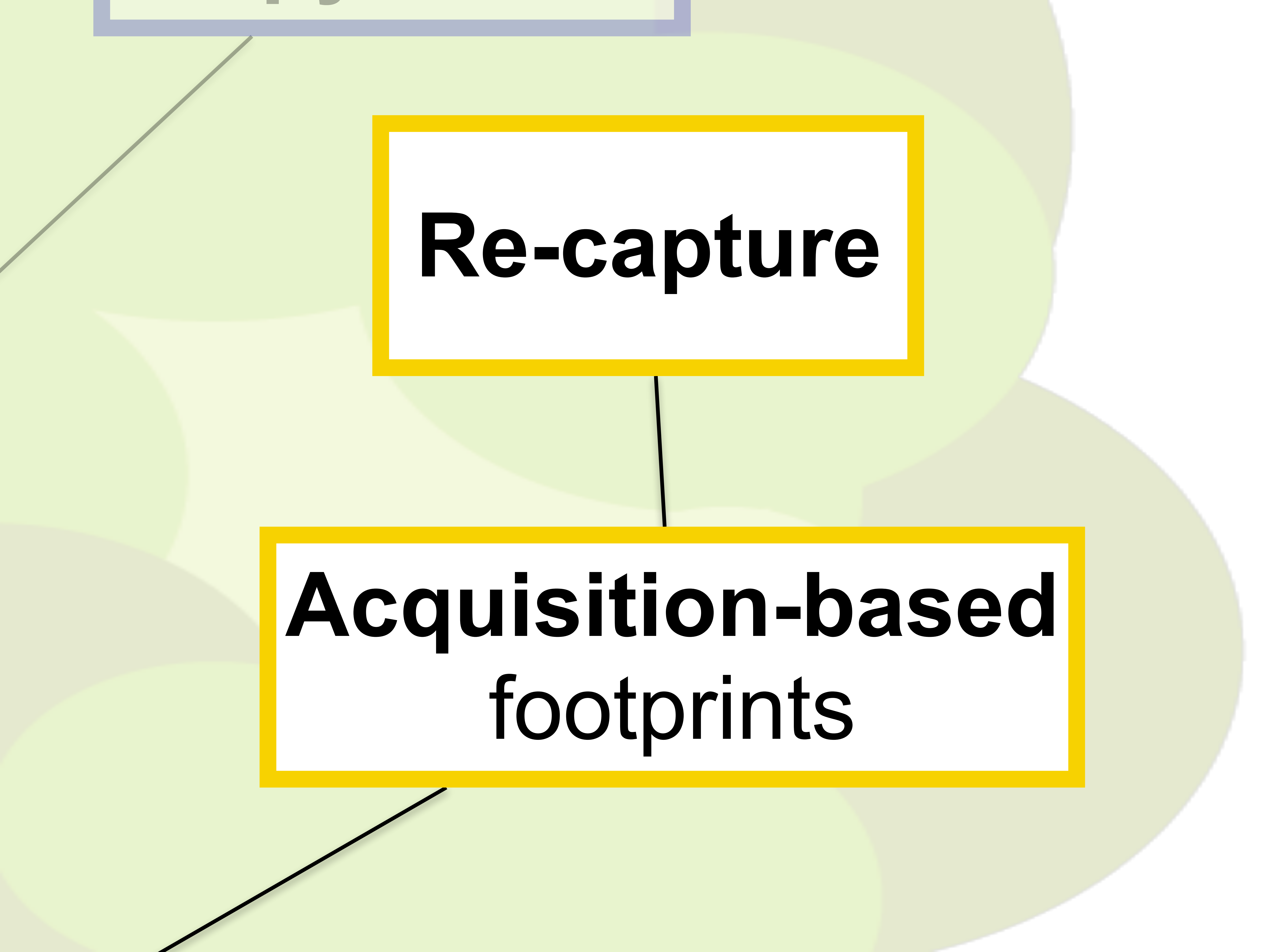


Image splicing
Video copy-paste
Video copy-move
pe of
dec
Re-cap
Editing-based footprints
Acquisitic
footpr


Re-capture
Acquisition-based footprints

Acquisition-based footprints

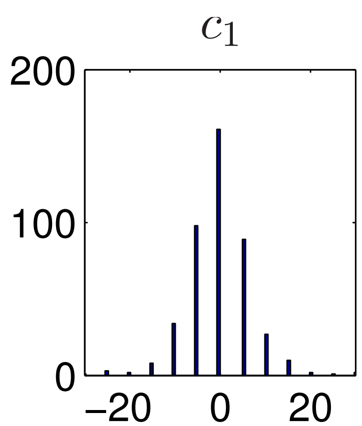
Acquisition-based footprints

Acquisition-based footprints

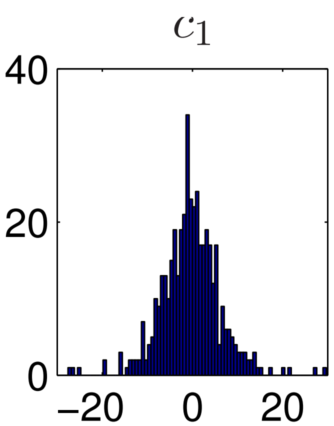
- Re-acquisition is a powerful **anti-forensic** tool

- Re-acquired videos are visually **similar to the originals**

- Many footprints are **masked**
  - Detectors can be-fooled



original: DCT statistics



recaptured: DCT statistics

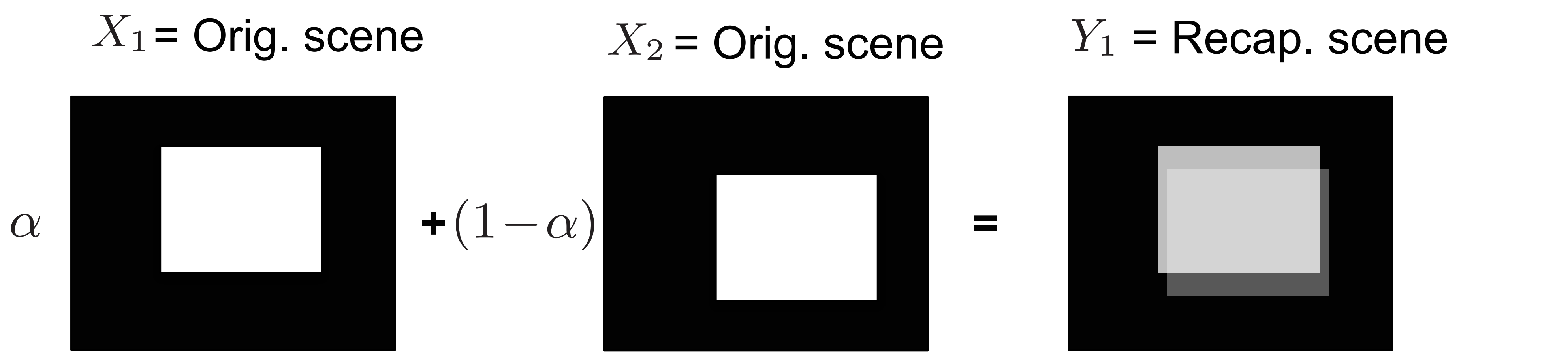
Is the video **recaptured?**

[Bestagini et al. **ICIP 2013**]

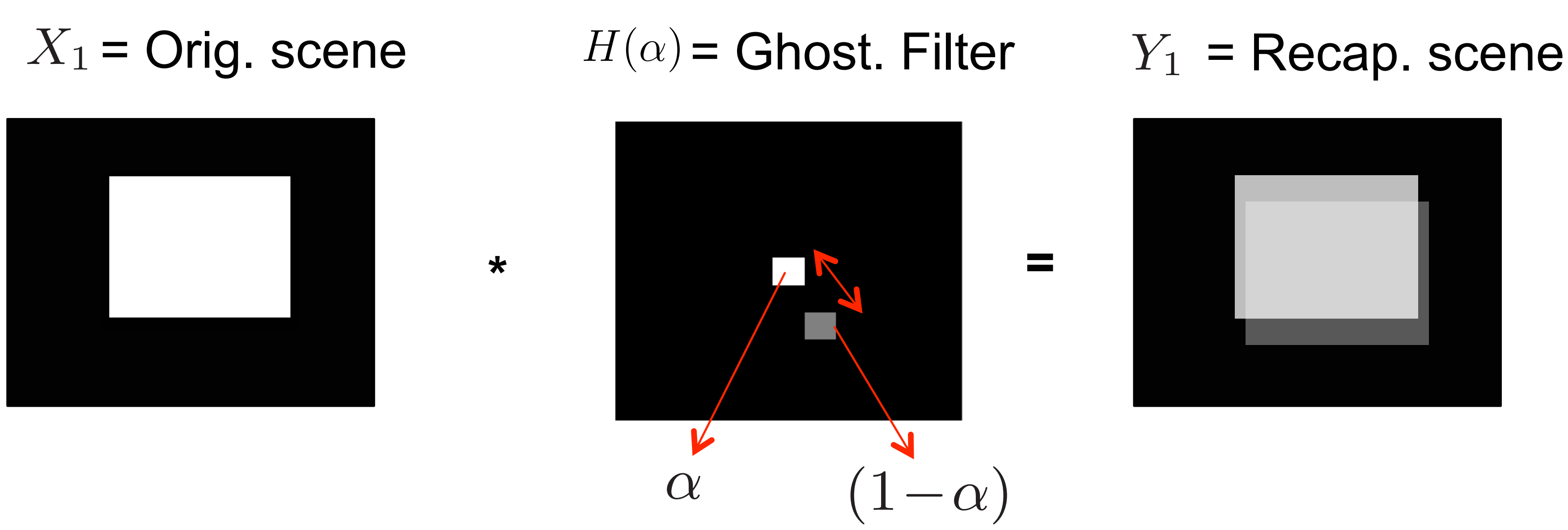
Acquisition-based footprints: re-capture

- **Setup:**
  - A video is **re-captured from a LCD** monitor

- **Ghosting as filtering:**

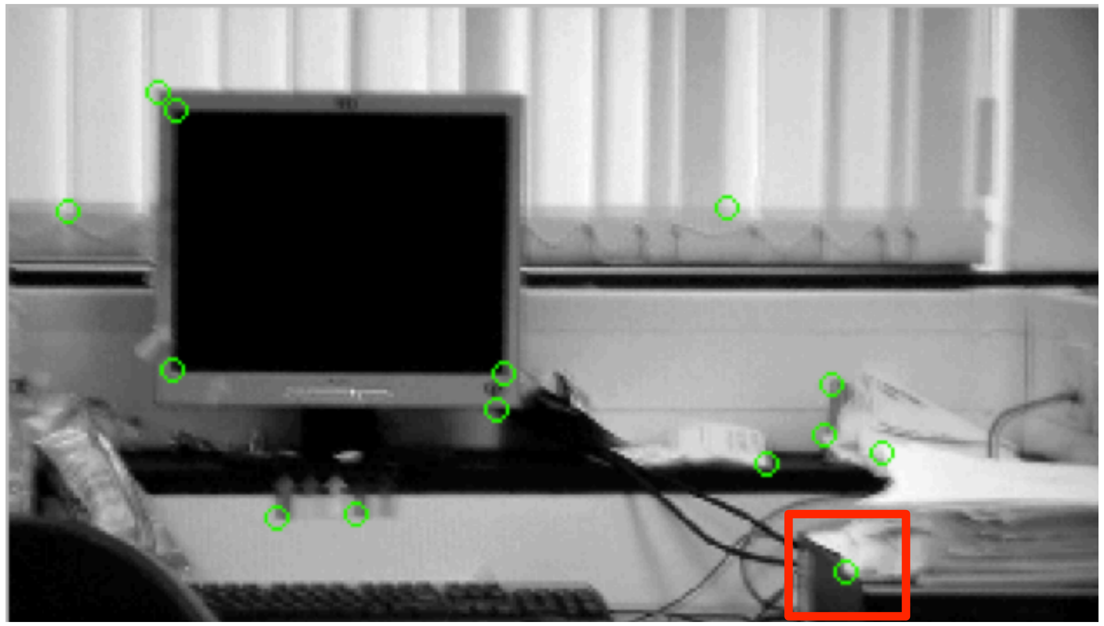


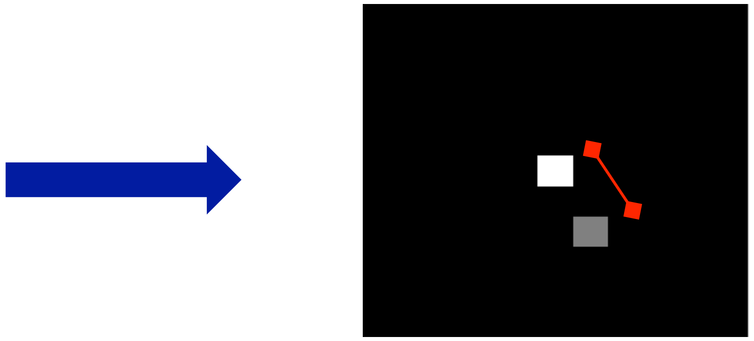
Approximating the motion as a translation between adjacent frames:



Acquisition-based footprints: re-capture

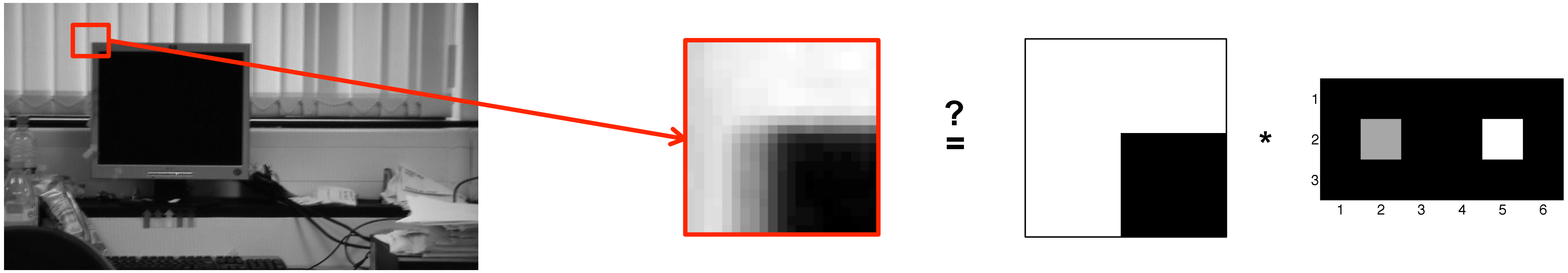
- Filter shape is derived from motion estimation





Acquisition-based footprints: re-capture

- **Method:**
  - Minimize a cost function to detect whether key-points underwent "ghosting filtering"



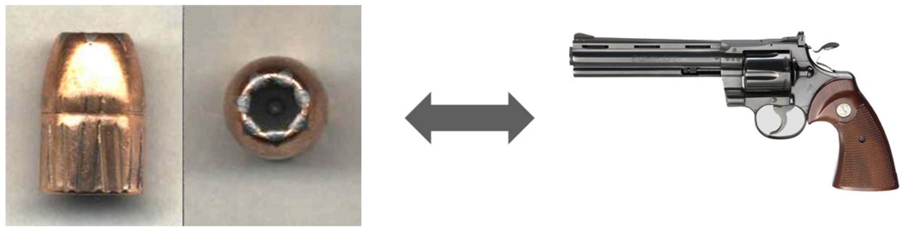
- **Results:**
  - Detection accuracy over 91%

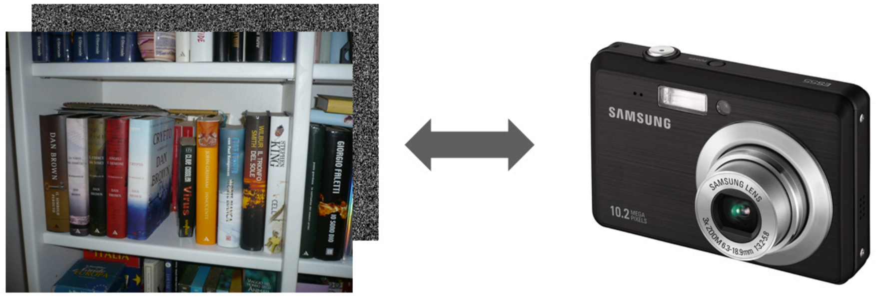
From **which camera** does the video come from?

[Mandelli et al. **EUSIPCO 2018**]
[Mandelli et al. **to be submitted**]

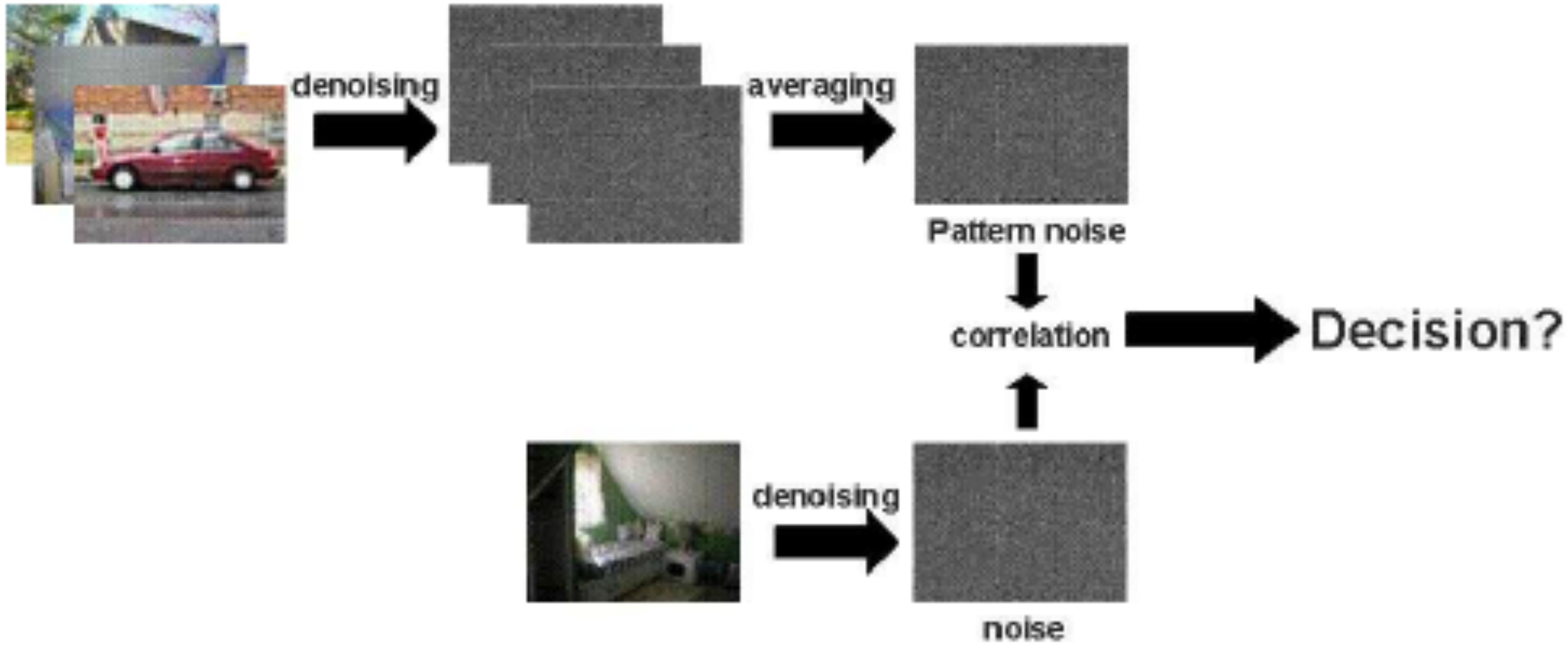
Acquisition-based footprints: camera attribution

- **Photo Response Non Uniformity:**
  - It enables linking images to devices



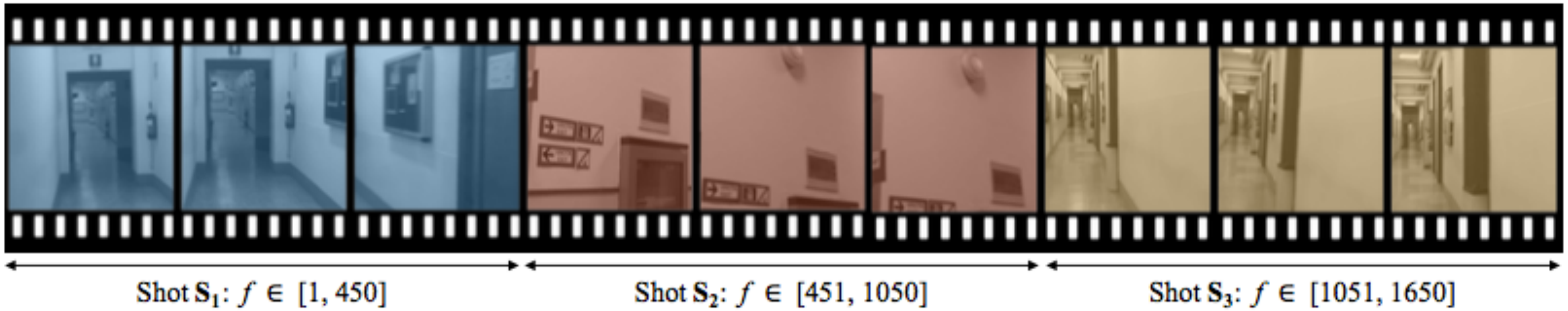


- **How to:**
  - Extract noise pattern from images
  - Compute correlation

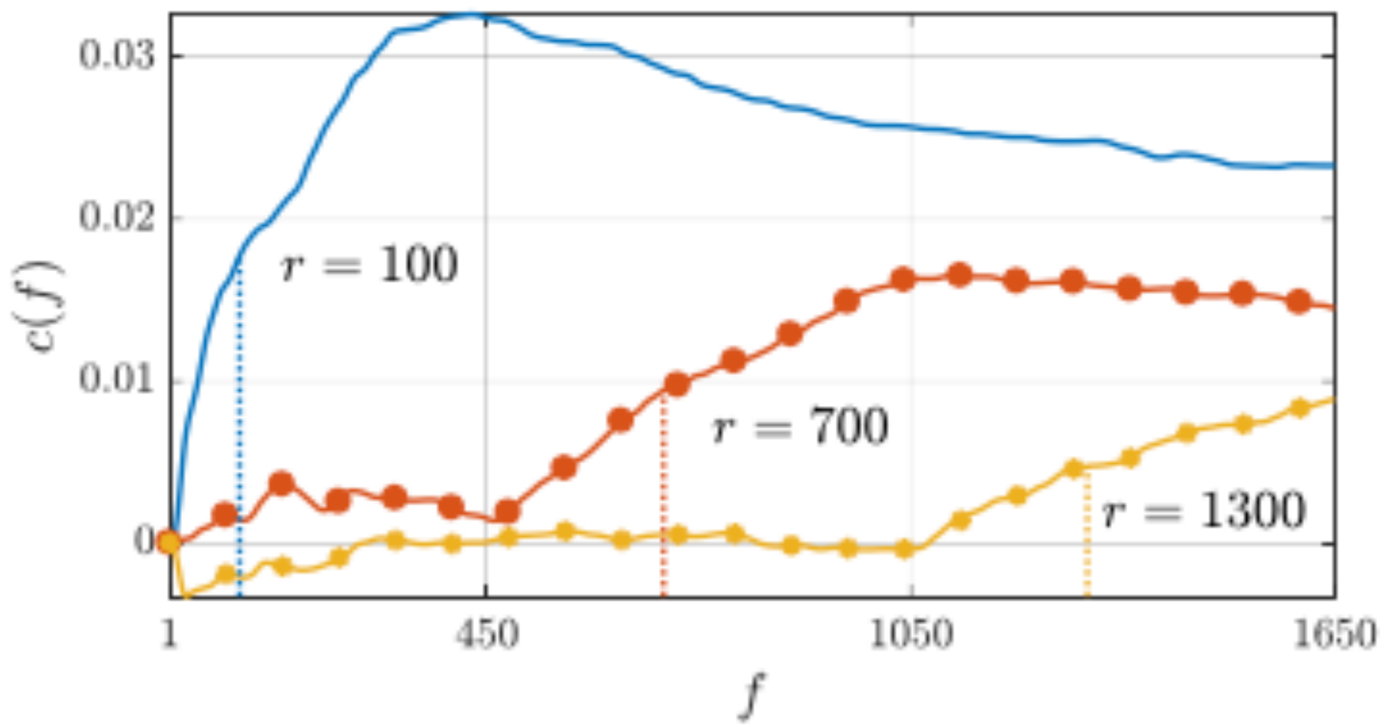


Acquisition-based footprints: camera attribution

- **Application:**
  - Video compilation detection and segmentation

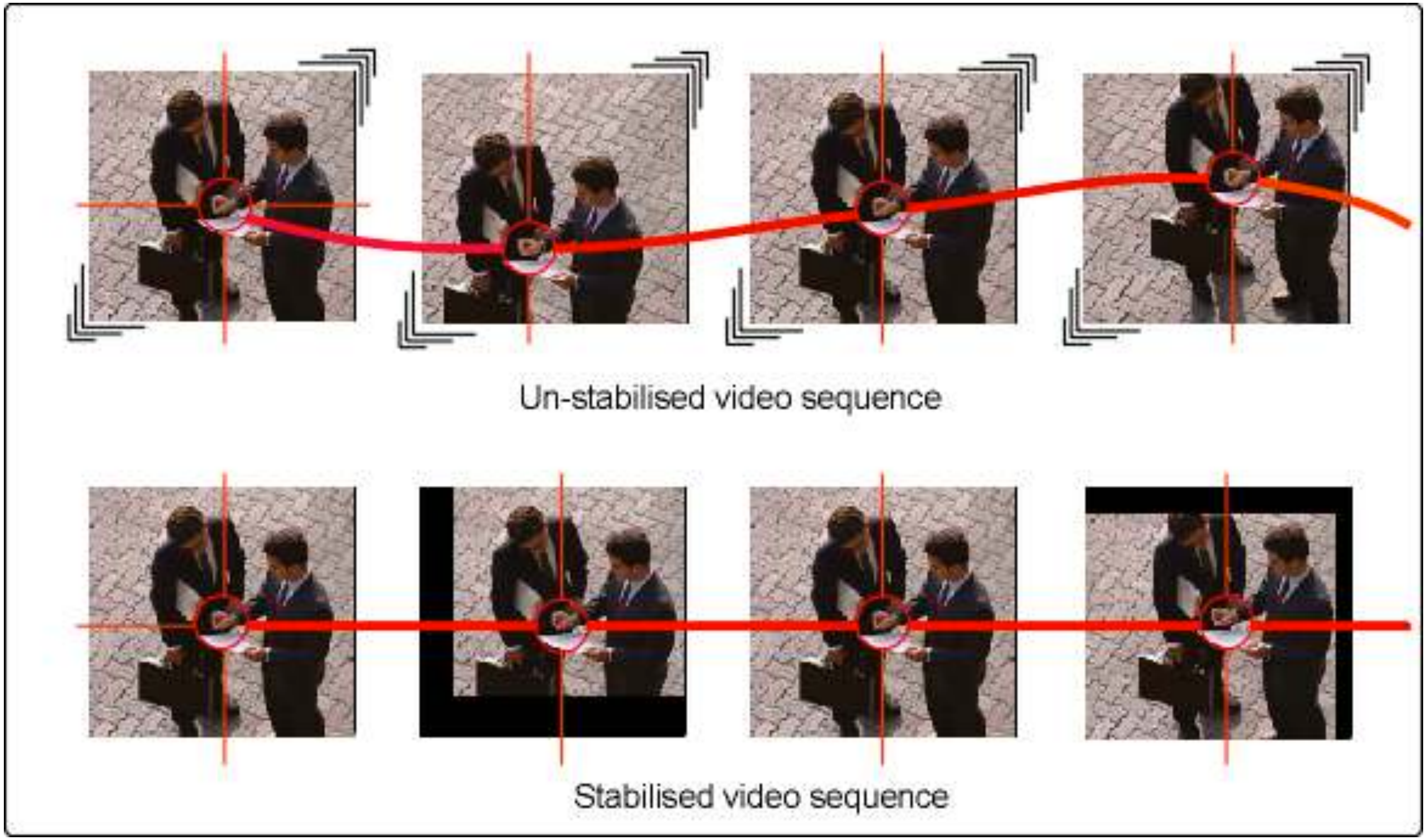


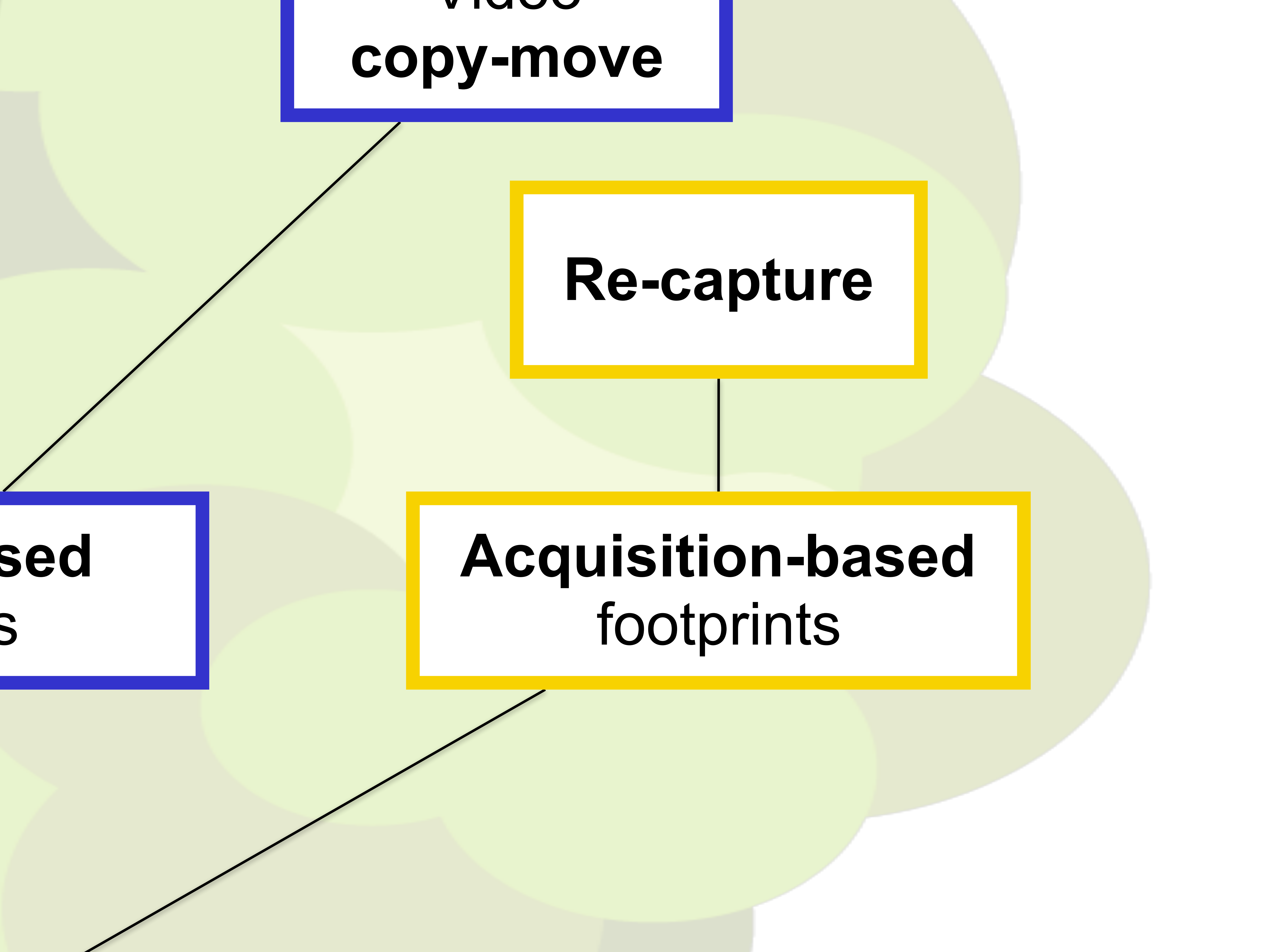
- **Method:**
  - Compute cumulative correlation $c(f) = \rho(\mathbf{W}_r, \overline{\mathbf{W}}(f)\mathbf{I}_r)$



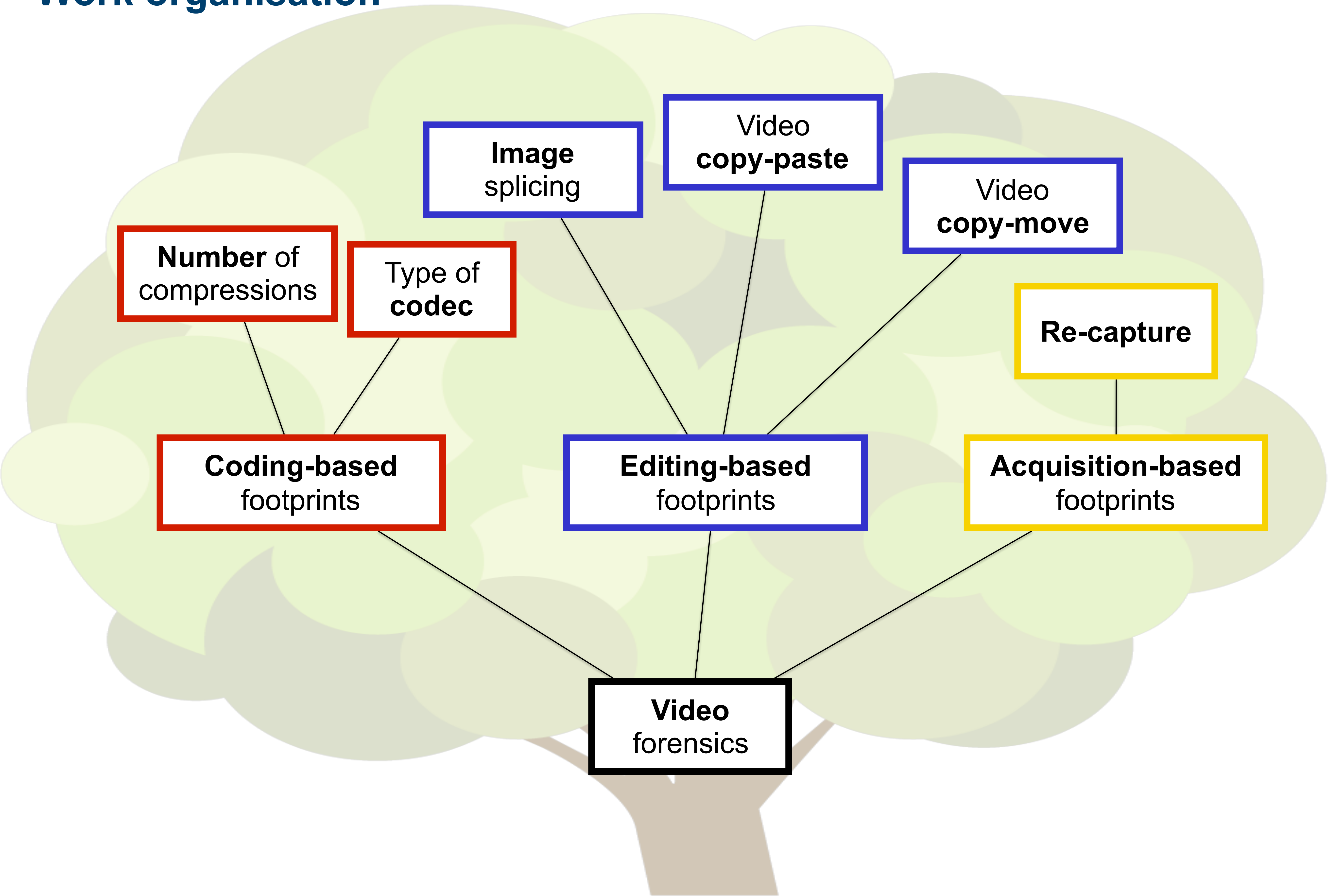
Acquisition-based footprints: camera attribution

- **Challenges:**
  - Aggressive coding
  - Digital video stabilizaion

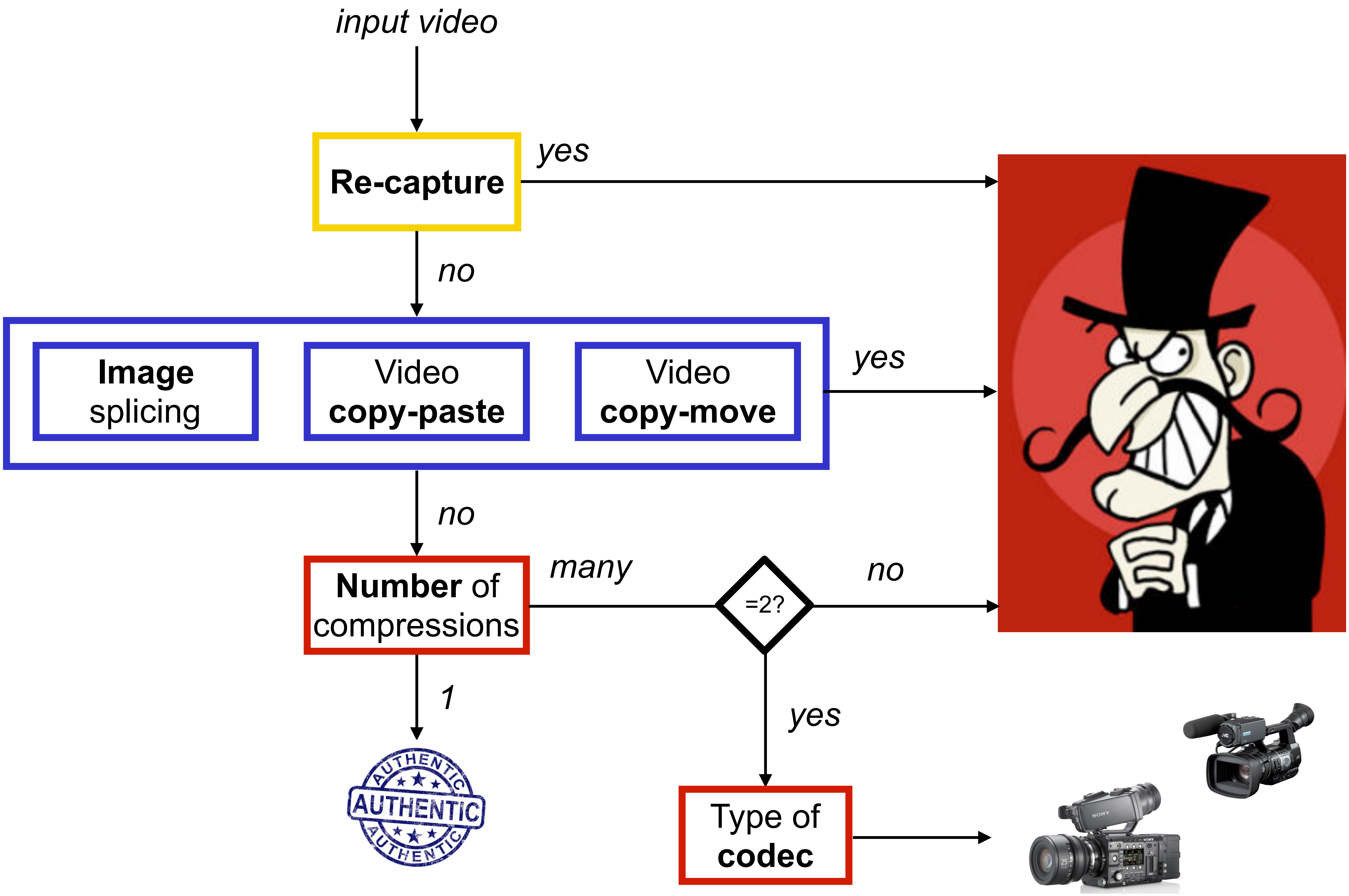



copy-move
Re-capture
sed
s
Acquisition-based footprints

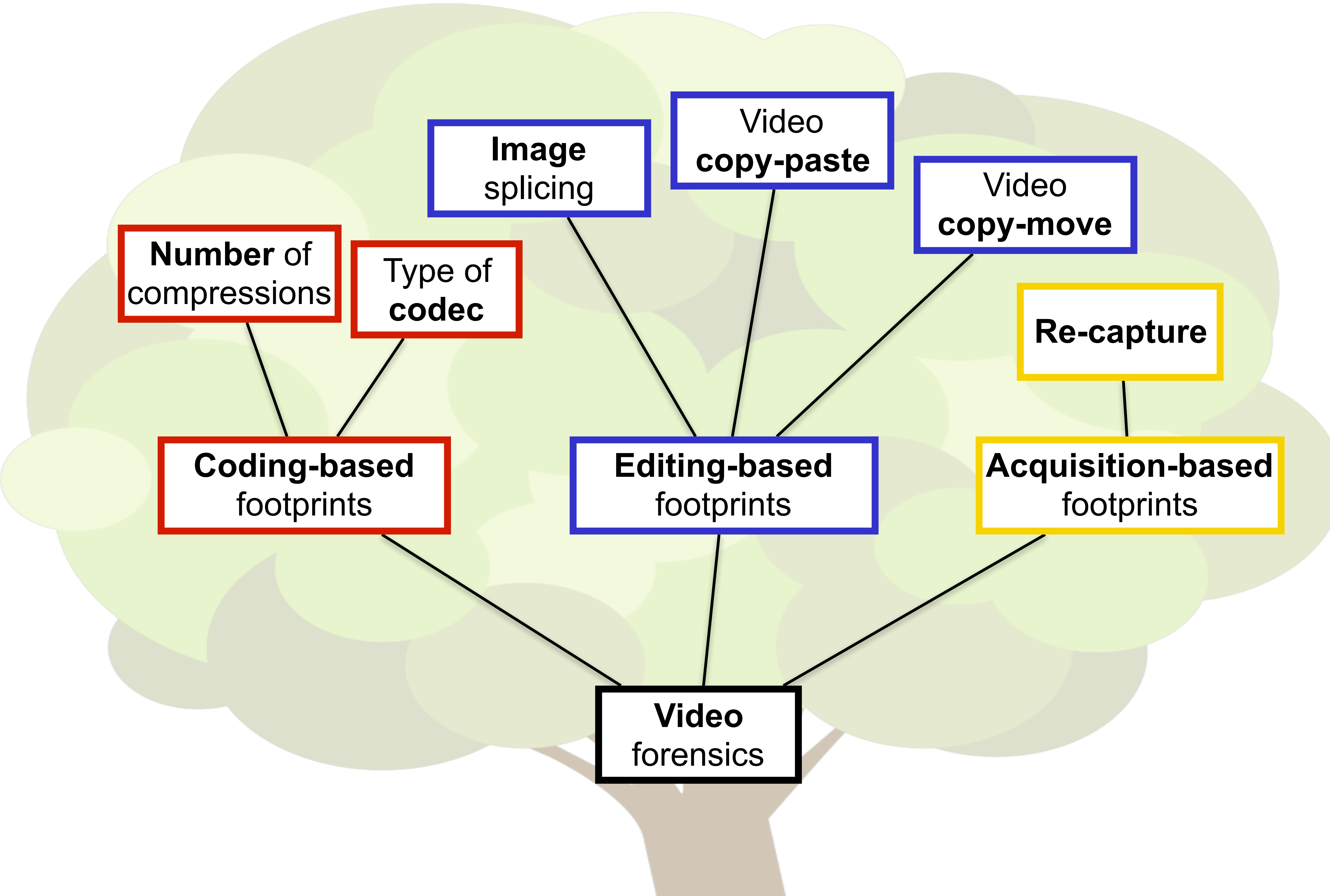
Work organisation



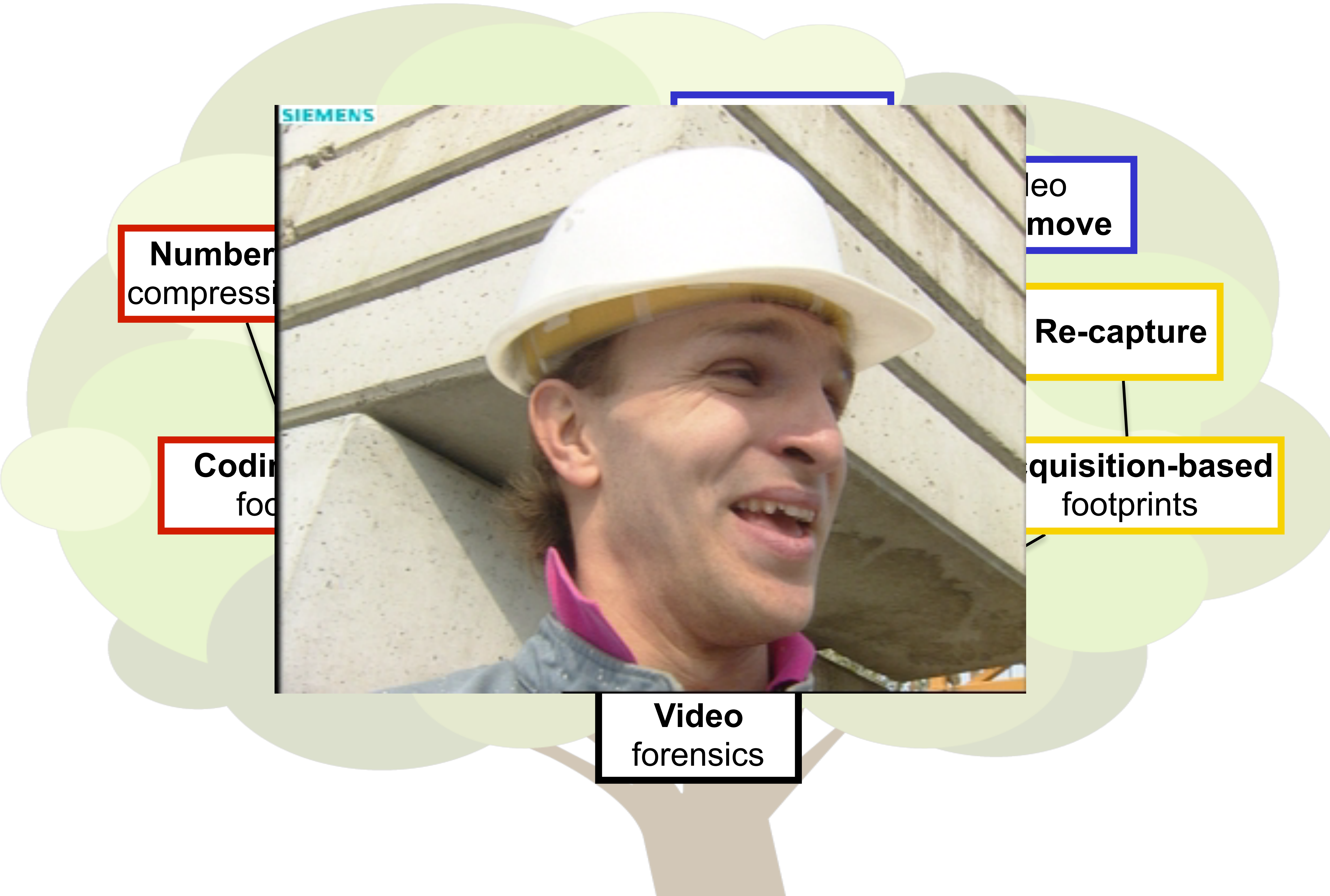
Put everything together



From one to many

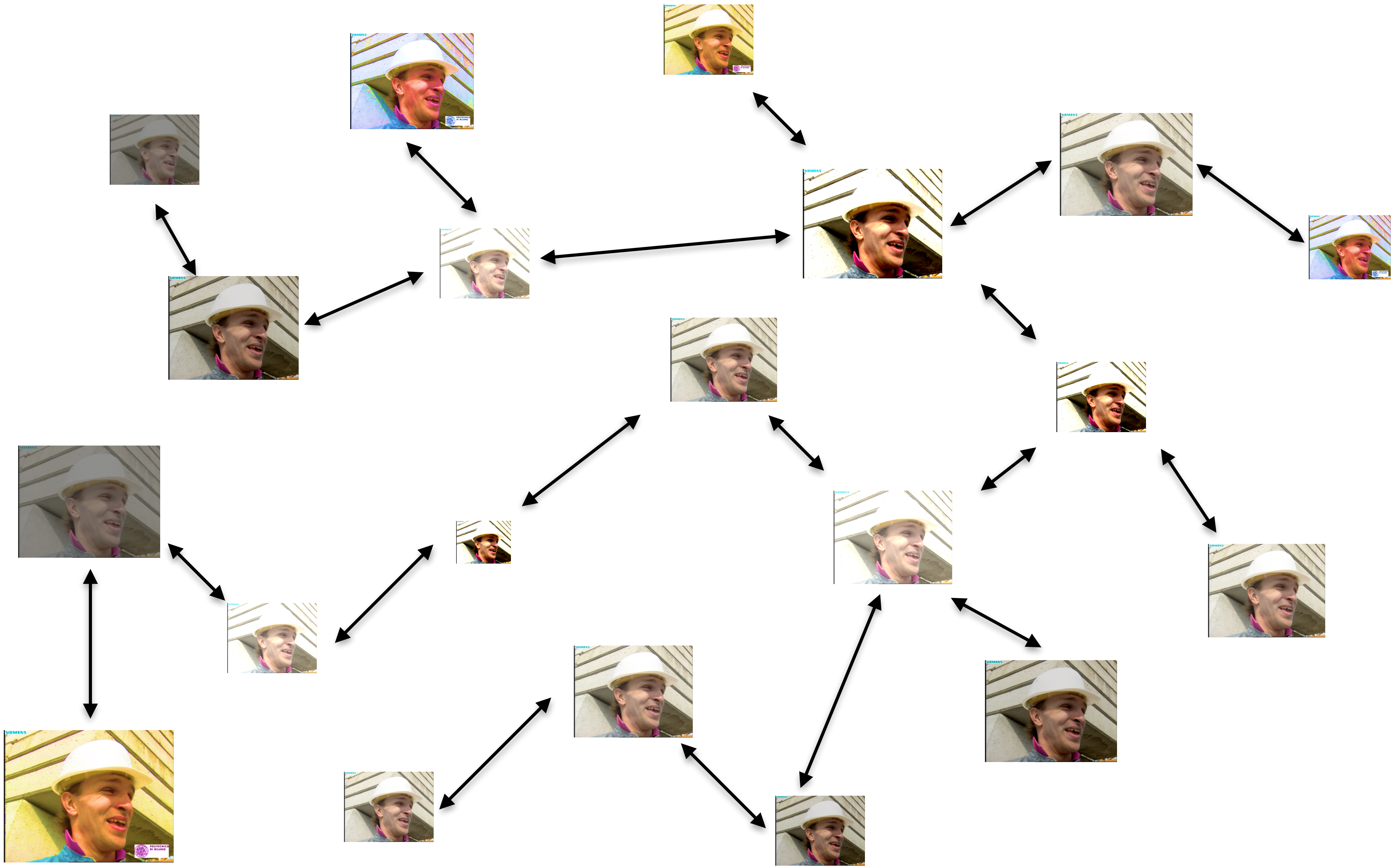


From one to many



From one to many

From one to many

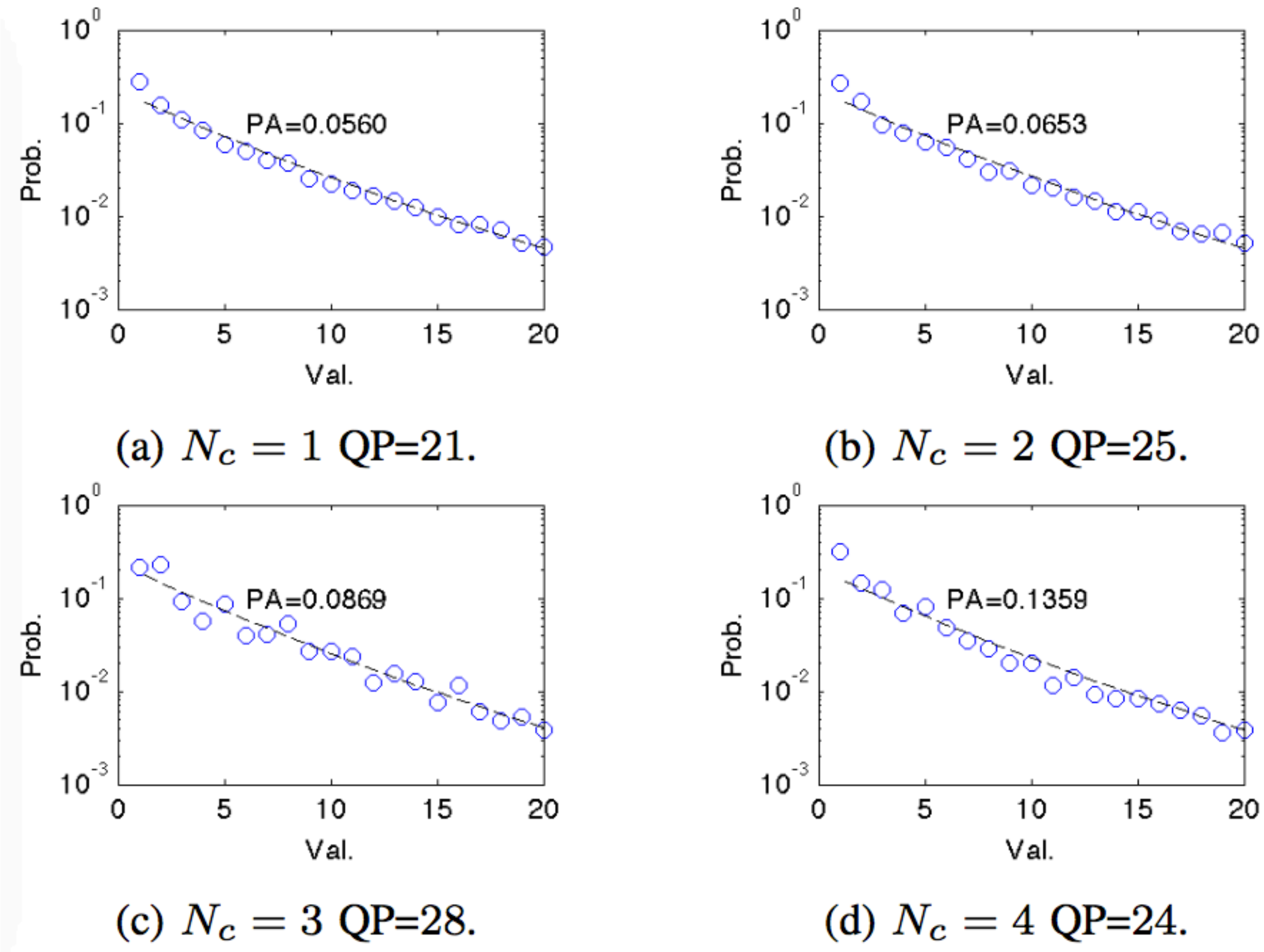


Applications

Coding-based applications

[Milani et al. **EUSIPCO 2017**]

- **Which video has been more processed?**
  - Extend Benford's law to base-N first digits
  - Fit logarithmic curve
  - Check goodness of fit (processing age)
  - The better, the younger!



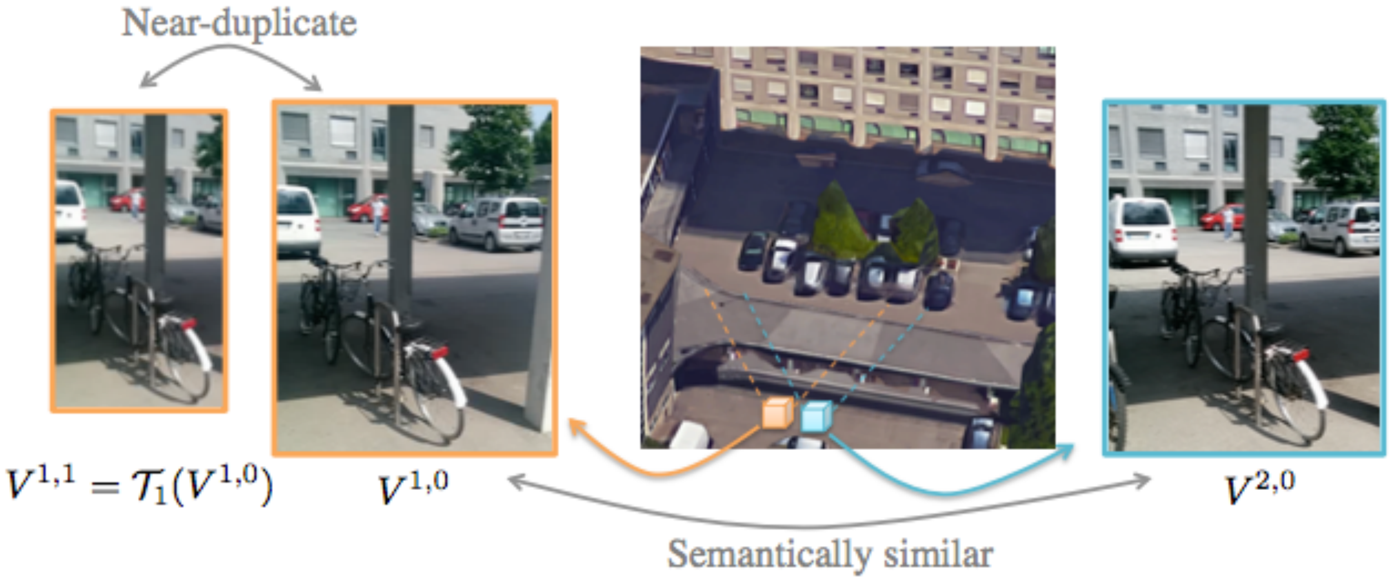
(a) $N_c = 1$ QP=21. (b) $N_c = 2$ QP=25.

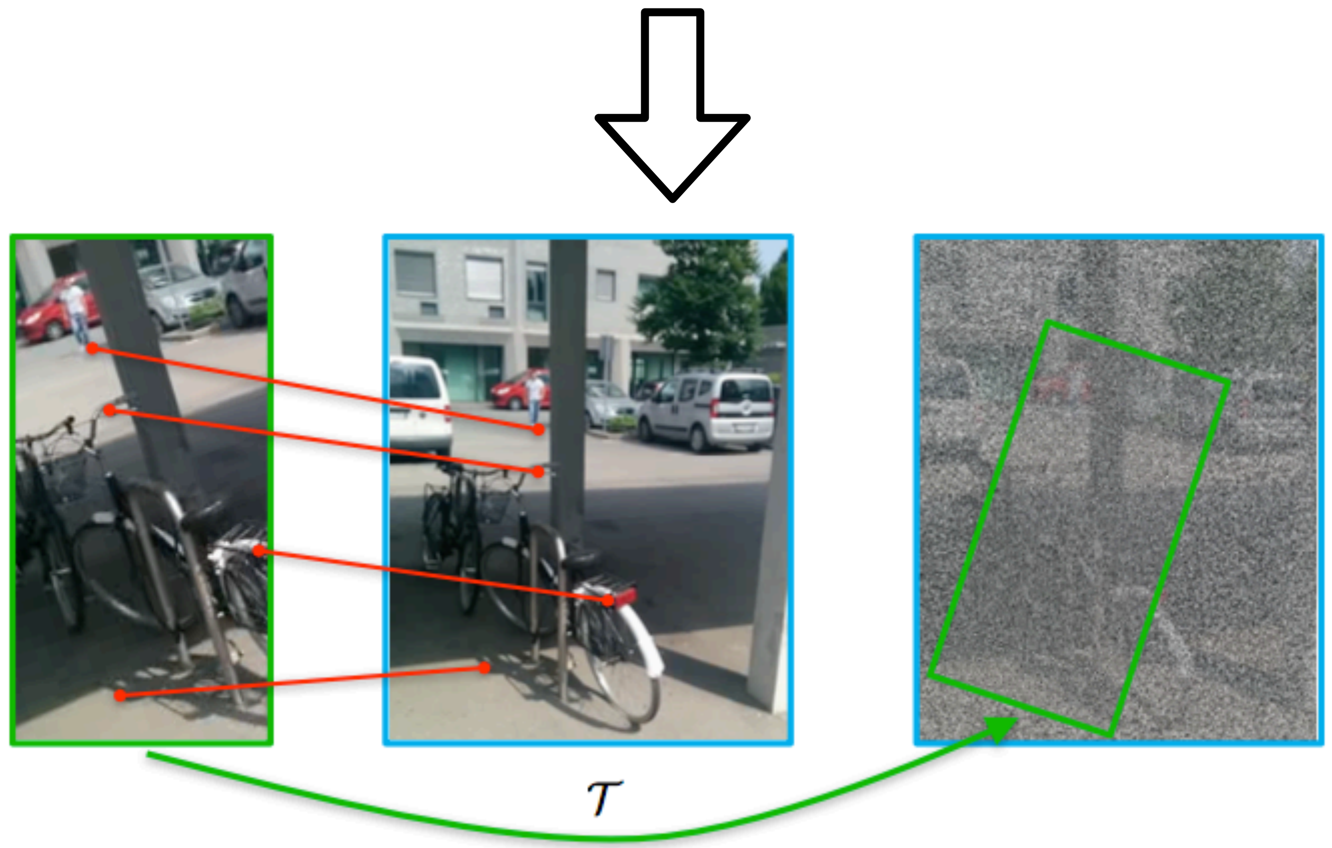
(c) $N_c = 3$ QP=28. (d) $N_c = 4$ QP=24.

Acquisition-based applications

[Lameri et al. **ICIP 2017**]

- **Which views are the redundant?**
  - match video PRNUs to detect those from same device





Parent reconstruction

[Bestagini et al. **ICIP 2014**]

- **Who is my parent?**
- It is possible that we are analysing a short shot (**child**) of a longer sequence (**parent**)
  - e.g., a VIP speech

- Can we find other (partially overlapping) **child** sequences to reconstruct the **parent**?

Parent reconstruction

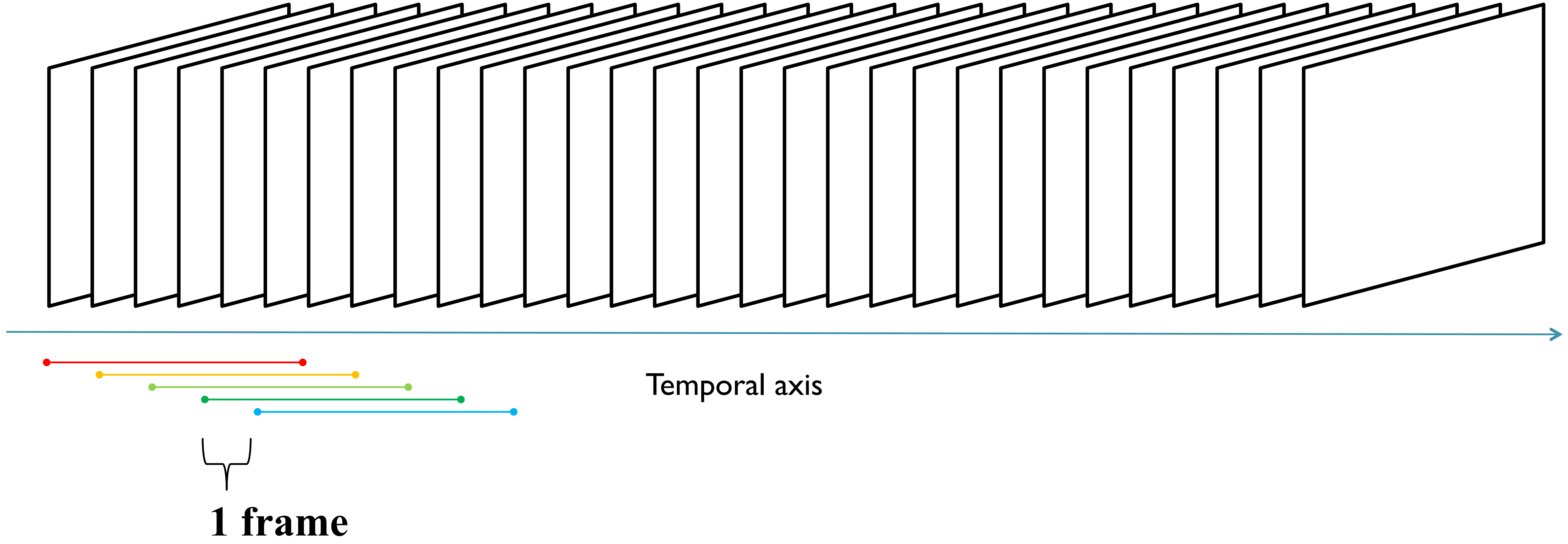
- Download a set of videos related to the topic under analysis:

Parent reconstruction

- Download a set of videos related to the topic under analysis:

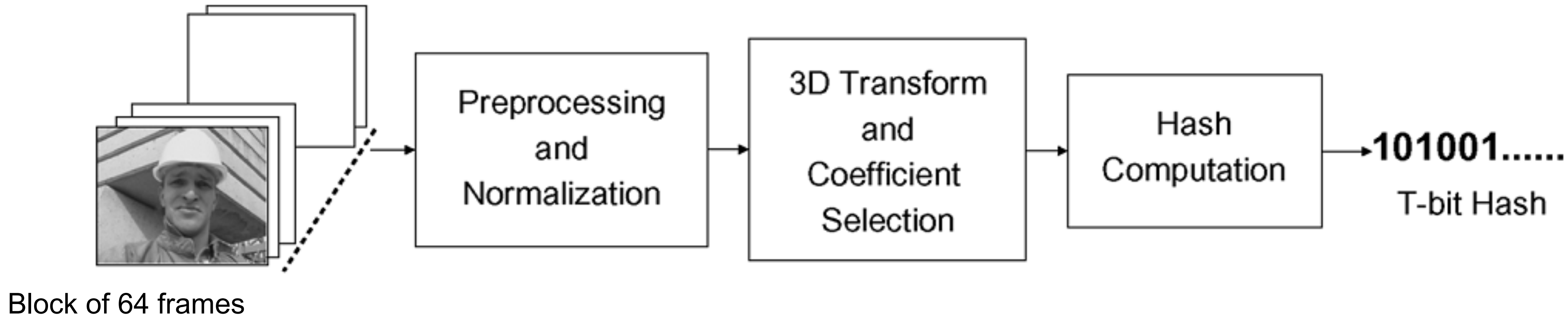
Parent reconstruction

- Analyse each pair of sequences exploiting a **robust hash** algorithm
  - A sequence is split in overlapping time segments of 64 frames each



Parent reconstruction

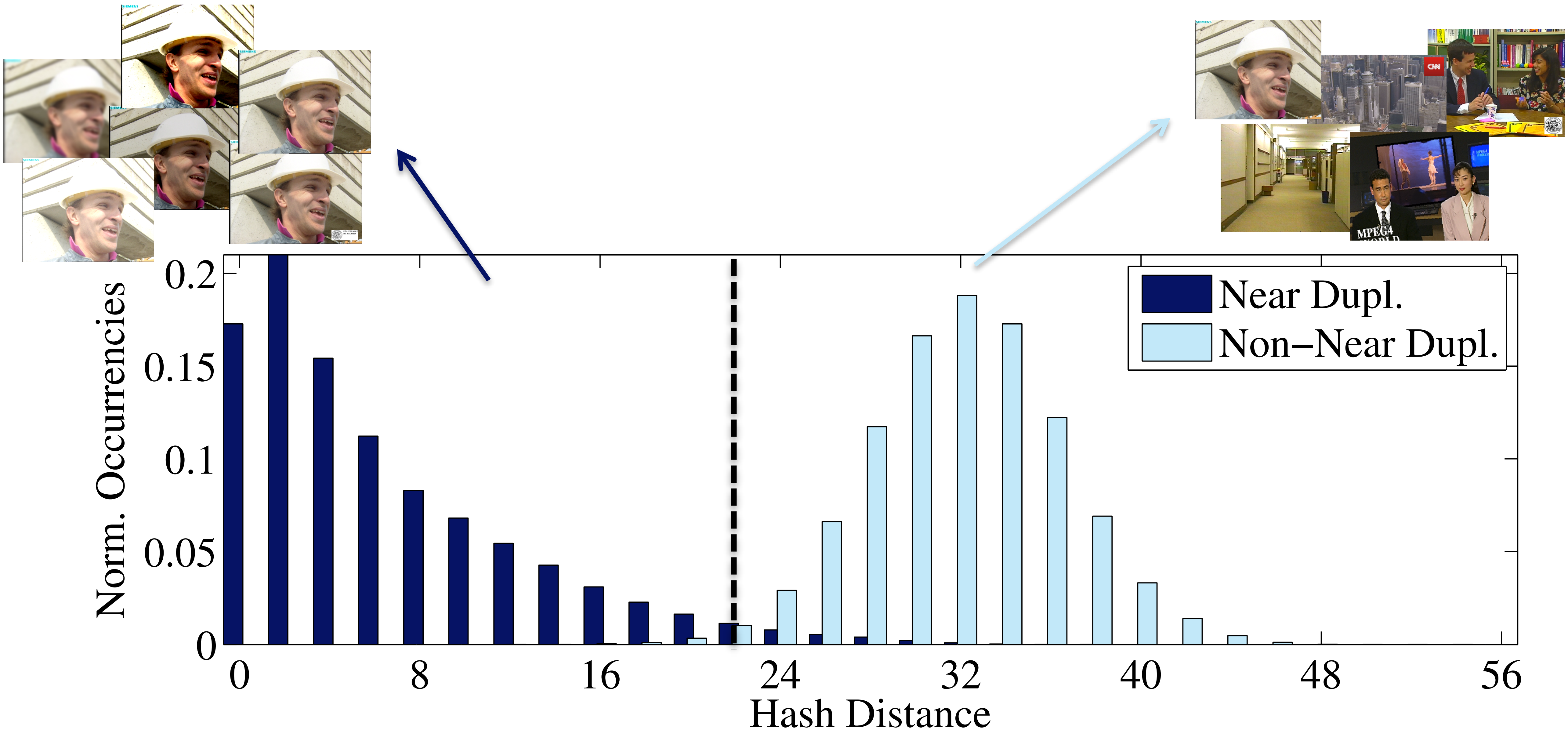
- Analyse each pair of sequences exploiting a **robust hash** algorithm
  - each block is described by a binary hash



- Every frame in the block is **spatially resized** to 32x32 pixels
  - The block now measures 32x32x64 pixels
- **3D DCT** is applied to the block
- **64** DCT **coefficients** are selected
- This 64 DCT coefficients are **binarized** according to their median value
  - 32 are set to zero, 32 are set to 1
- The **hash** is this 64 binary string

Parent reconstruction

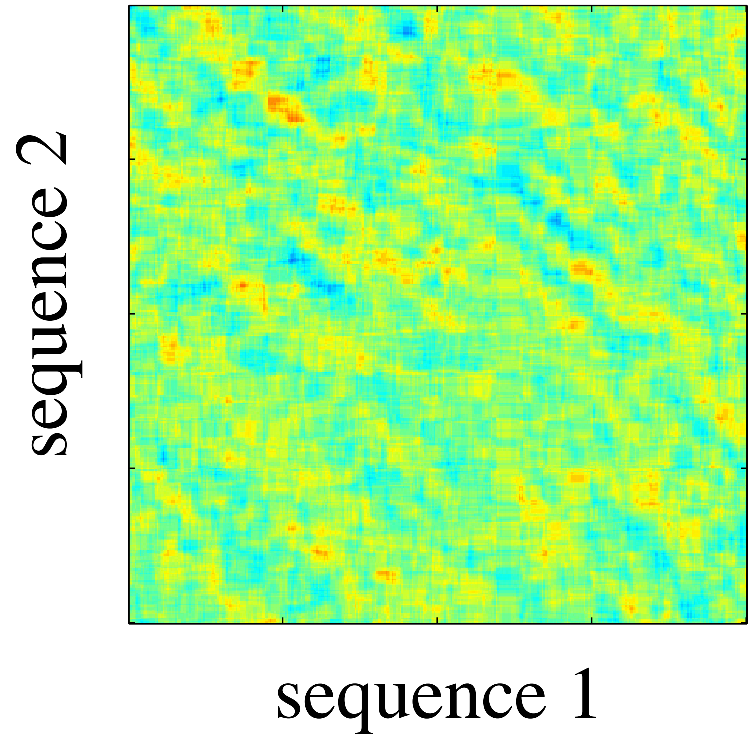
- Analyse each pair of sequences exploiting a **robust hash** algorithm
  - Hashes of different blocks are compared by computing **hamming distance**



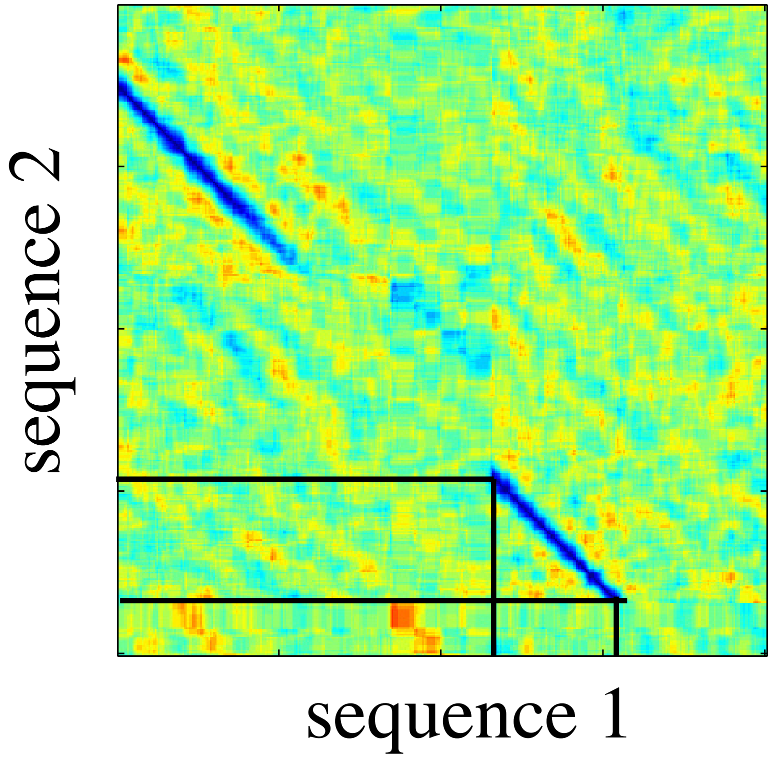
Parent reconstruction

- Compute the **distance between every block** of *sequence1* and every block of *sequence2*

- **Non-near duplicates**
  - High distance
  - No regular patterns

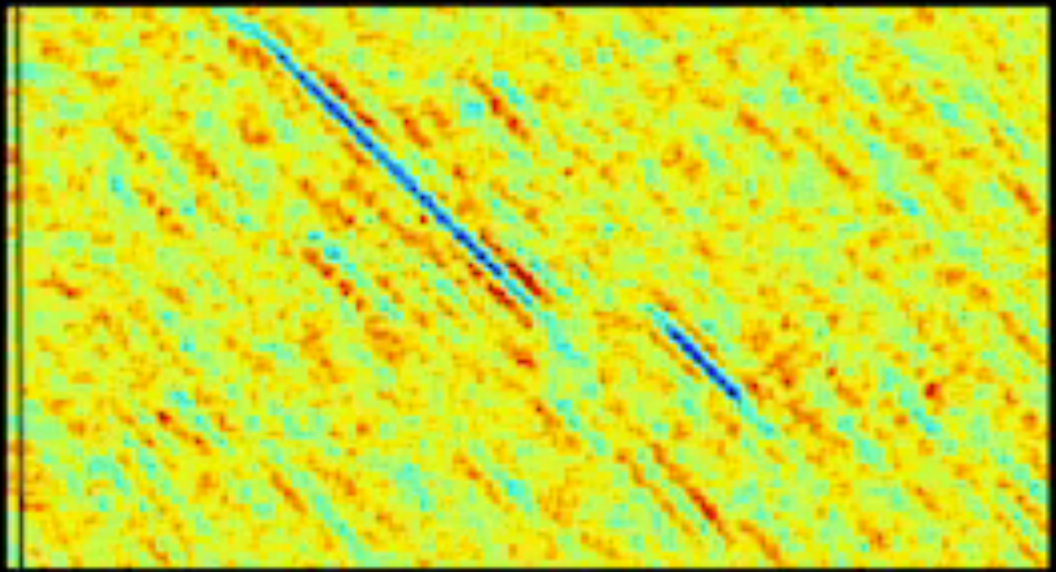


- **Near duplicates**
  - Low distances = matching
  - Start and end points used for alignment



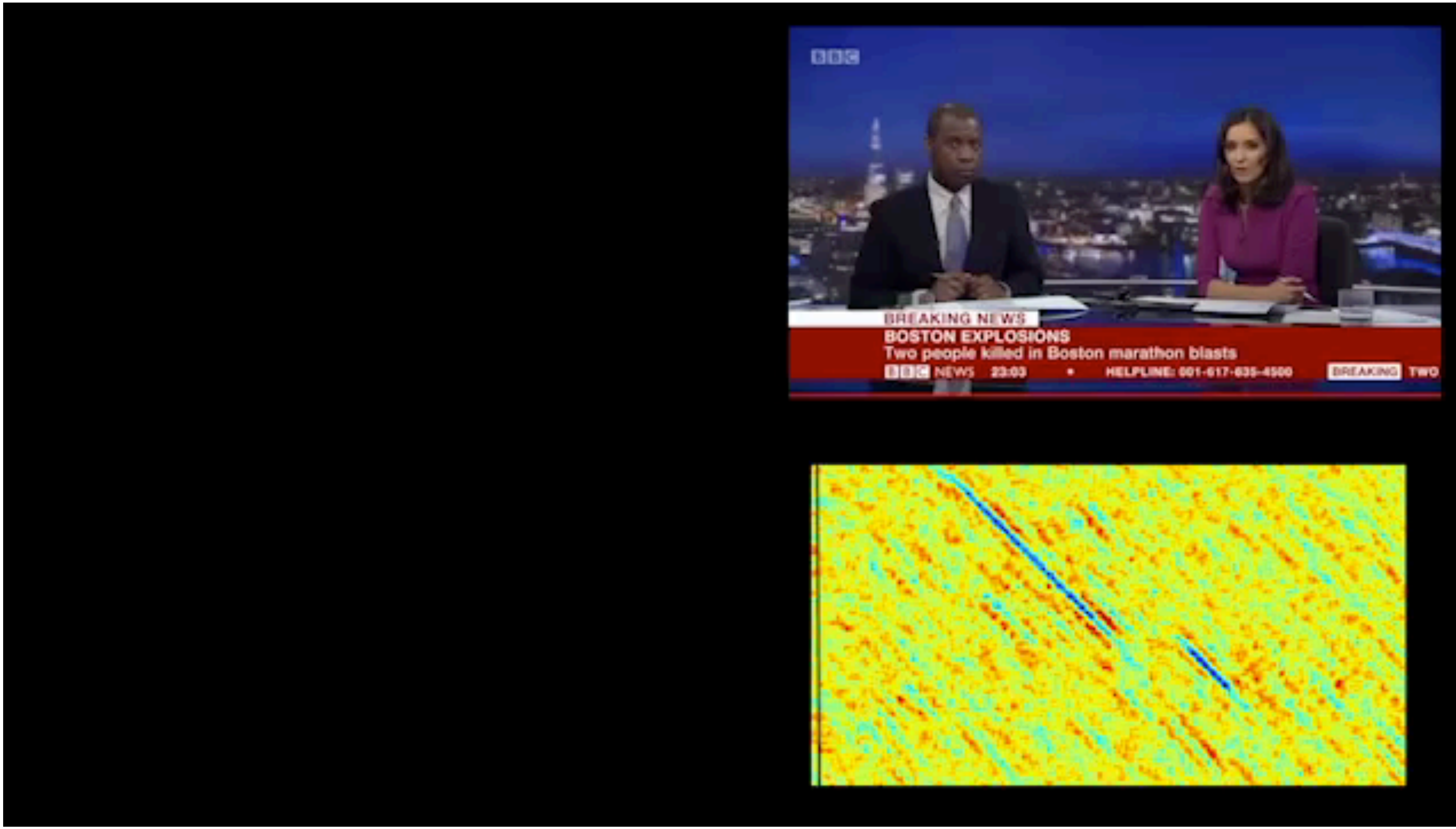
Parent reconstruction

- Analyse each pair of sequences exploiting a robust hash algorithm:

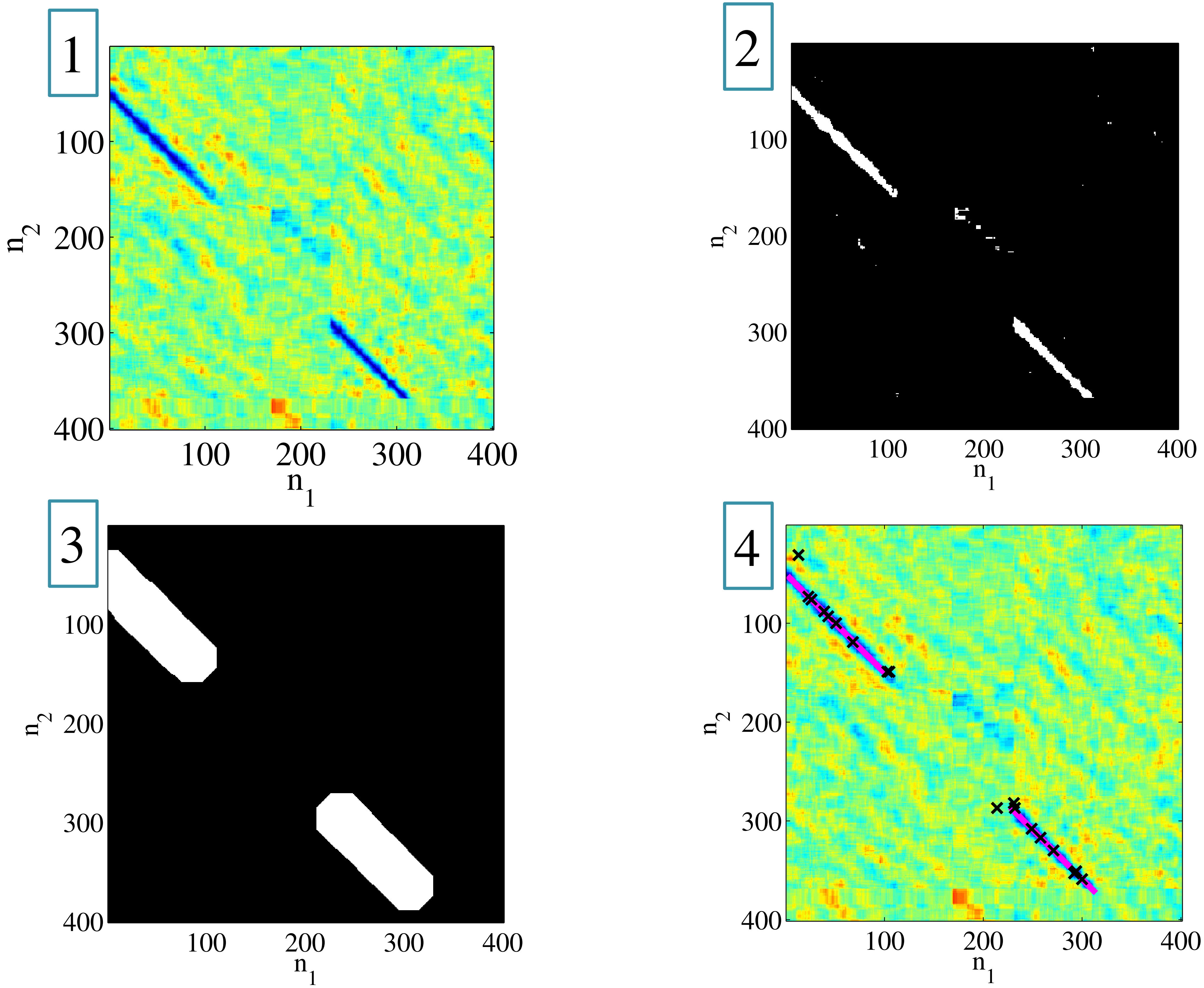


Parent reconstruction

- Analyse each pair of sequences exploiting a robust hash algorithm:



Parent reconstruction



Parent reconstruction

- Segment each sequence according to the matching shots:

Parent reconstruction

- Segment each sequence according to the matching shots:

Parent reconstruction

- Reconstruct the most part of the **parent** sequence for the analysis:

Parent reconstruction

- Reconstruct the most part of the **parent** sequence for the analysis:

Parent reconstruction

- Being able to **reconstruct the parent** from the children enables to shed very interesting insights on the way content is reused:

  1. Analyse the **context** from which a child sequence was taken
  2. **Reconstruct** sequences no longer available online in their totality
  3. Establish causal **relationship** between children

Conclusions

- **Remarks**
  - Forensics vs. Anti-forensics
  - Single video analysis is just part of the problem
  - Multiple video analysis paves the way to the development of novel applications

- **Open questions**
  - Merge results from *content*- and *context*-aware detectors
    - Do metadata match the video content?
  - Deal with *big data*
    - Time-consuming algorithms need optimisation
  - Deep learning
    - Still under-investigated in video forensics (space-time?)
    - Training data hardly available…

References

- S. Verde, L. Bondi, P. Bestagini, S. Milani, G. Calvagno, S. Tubaro, **"Video Codec Forensics Based on Convolutional Neural Networks"**, IEEE International Conference on Image Processing (ICIP), Athens, Greece, 2018
- S. Mandelli, D. Cozzolino, P. Bestagini, L. Verdoliva, S. Tubaro, **"Blind Detection and Localization of Video Temporal Splicing Exploiting Sensor-Based Footprints"**, European Signal Processing Conference (EUSIPCO), Rome, Italy, 2018
- S. Lameri, L. Bondi, P. Bestagini, S. Tubaro, **"Near-Duplicate Video Detection Exploiting Noise Residual Traces"**, IEEE International Conference on Image Processing (ICIP), Beijing, China, 2017
- S. Milani, P. Bestagini, S. Tubaro, "**Video Phylogeny Tree Reconstruction Using Aging Measures"**, European Signal Processing Conference (EUSIPCO), Kos, Greece, 2017
- F. Costa, S. Lameri, P. Bestagini, Z. Dias, S. Tubaro, A. Rocha, **"Hash-Based Frame Selection for Video Phylogeny"**, IEEE International Workshop on Information Forensics and Security (WIFS), Abu Dhabi, UAE, 2016
- S. Lameri, P. Bestagini, S. Tubaro, **"Video alignment for phylogenetic analysis"**, European Signal Processing Conference (EUSIPCO), Budapest, Hungary, 2016
- P. Bestagini, S. Milani, M. Tagliasacchi, S. Tubaro, "**Codec and GOP Identification in Double Compressed Videos**", IEEE Transactions on Image Processing, vol.25, no.5, pp.2298-2310, 2016
- F. O. Costa, S. Lameri, P. Bestagini, Z. Dias, A. Rocha, M. Tagliasacchi, S. Tubaro, **"Phylogeny reconstruction for misaligned and compressed video sequences"**, IEEE International Conference on Image Processing (ICIP), Québec City, Canada, 2015

References

- S. Lameri, P. Bestagini, A. Melloni, S. Milani, A. Rocha, M. Tagliasacchi, S. Tubaro, "**Who is my parent? Reconstructing video sequences from partially matching shots**" , IEEE International Conference on Image Processing (ICIP), Paris, France, 2014
- P. Bestagini, S. Milani, M. Tagliasacchi, S. Tubaro, "**Local tampering detection in video sequences**" , IEEE International Workshop on Multimedia Signal Processing (MMSP), Pula, Italy, 2013
- P. Bestagini, M. Visentini-Scarzanella, M. Tagliasacchi, P. L. Dragotti, S. Tubaro, "**Video recapture detection based on ghosting artifact analysis**" , IEEE International Conference on Image Processing (ICIP), Melbourne, Australia, 2013
- P. Bestagini, S. Milani, M. Tagliasacchi, S. Tubaro, "**Video codec identification extending the idempotency property**" , European Workshop on Visual Information Processing (EUVIP), Paris, France, 2013
- P. Bestagini, S. Battaglia, S. Milani, M. Tagliasacchi, S. Tubaro, "**Detection of temporal interpolation in video sequences**" , IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, Canada, 2013
- S. Milani, P. Bestagini, M. Tagliasacchi, S. Tubaro, "**Multiple Compression Detection for Video Sequences**" , IEEE International Workshop on Multimedia Signal Processing (MMSP), Banff, Canada, 2012
- S. Milani, M. Fontani, P. Bestagini, M. Barni, A. Piva, M. Tagliasacchi, S. Tubaro, "**An overview on video forensics**" , APSIPA Transactions on Signal and Information Processing, vol.1, 2012
- P. Bestagini, A. Allam, S. Milani, M. Tagliasacchi, S. Tubaro, "**Video codec identification**" , IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 2012

Thank you for the attention!

Any questions?