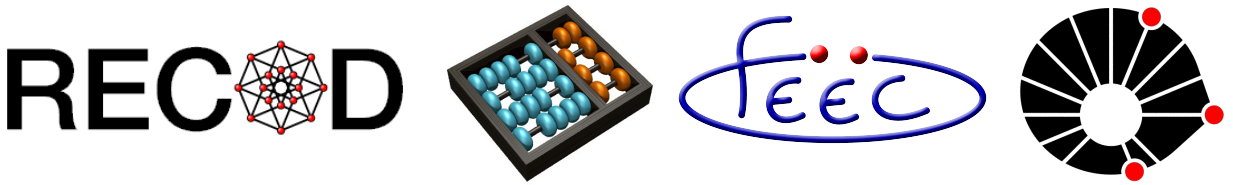
Sensitive Media Analysis (Pornography Detection)

Digital Forensics – MO447 / MC919

Sandra Avila
Postdoctoral Researcher
sandra@dca.fee.unicamp.br



Agenda

- Motivation
- Pornography … what is it?
- Existing Solutions
- Recent Works
- Conclusions

Pornography Detection: why do we care?

www.huffingtonpost.com


childrenvoice.com

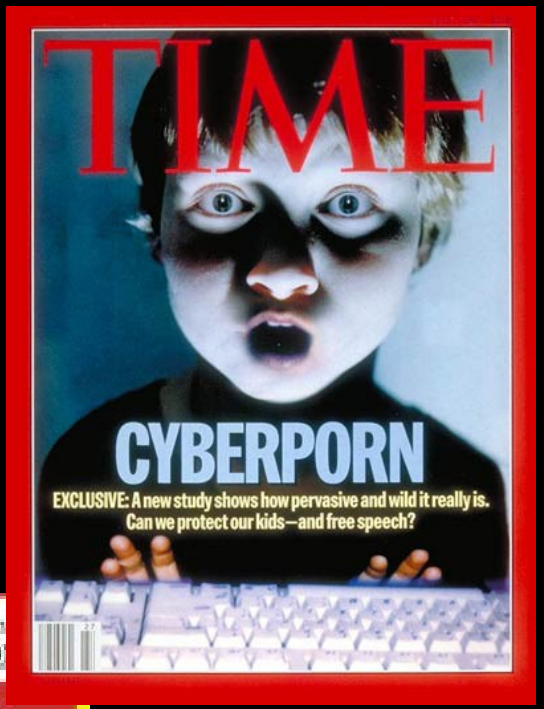

Google, GE & the Smart Home
Adam Carolla Uncaged
FAST COMPANY
ONE BANK TO LOVE
FORD REBOOTS
IRAQ'S OIL WAR
"With these devices, kids explore by themselves and figure it out."
—PAUL KIM
The REAL SMART PHONE REVOLUTION
How Tech Is Making Kids Smarter Everywhere
By Anya Kamenetz
amazon.com

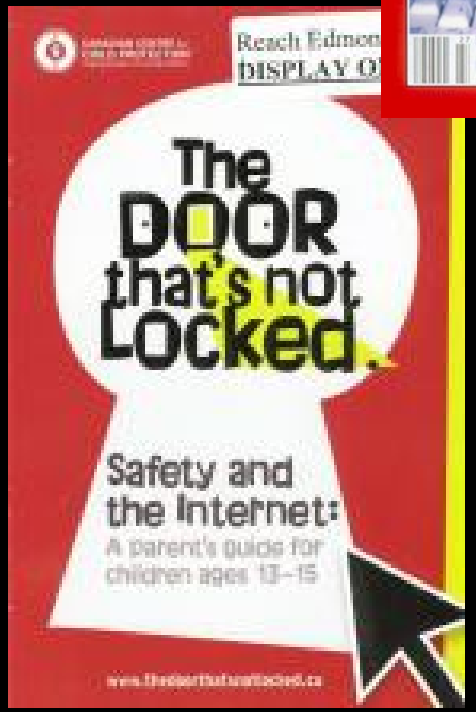
But ...

Are they safe?


TIME
COMPUTER GENERATION
A New Breed of Whiz Kids


Tea Party Under Fire / Rescuing Gorillas / Evernote
TIME
THE SECRET WEB
WHERE DRUGS, PORN AND MURDER HIDE ONLINE
BY LEV GROSSMAN & JAY NEWTON-SMALL


TIME
CYBERPORN
EXCLUSIVE: A new study shows how pervasive and wild it really is.
Can we protect our kids—and free speech?


The DOOR that's not Locked.
Safety and the Internet:
A parent's guide for children ages 13–15

Pornography ... what is it?

Pornography?

No, it isn't.

Pornography?

No, it isn't! Your (malicious) thoughts.

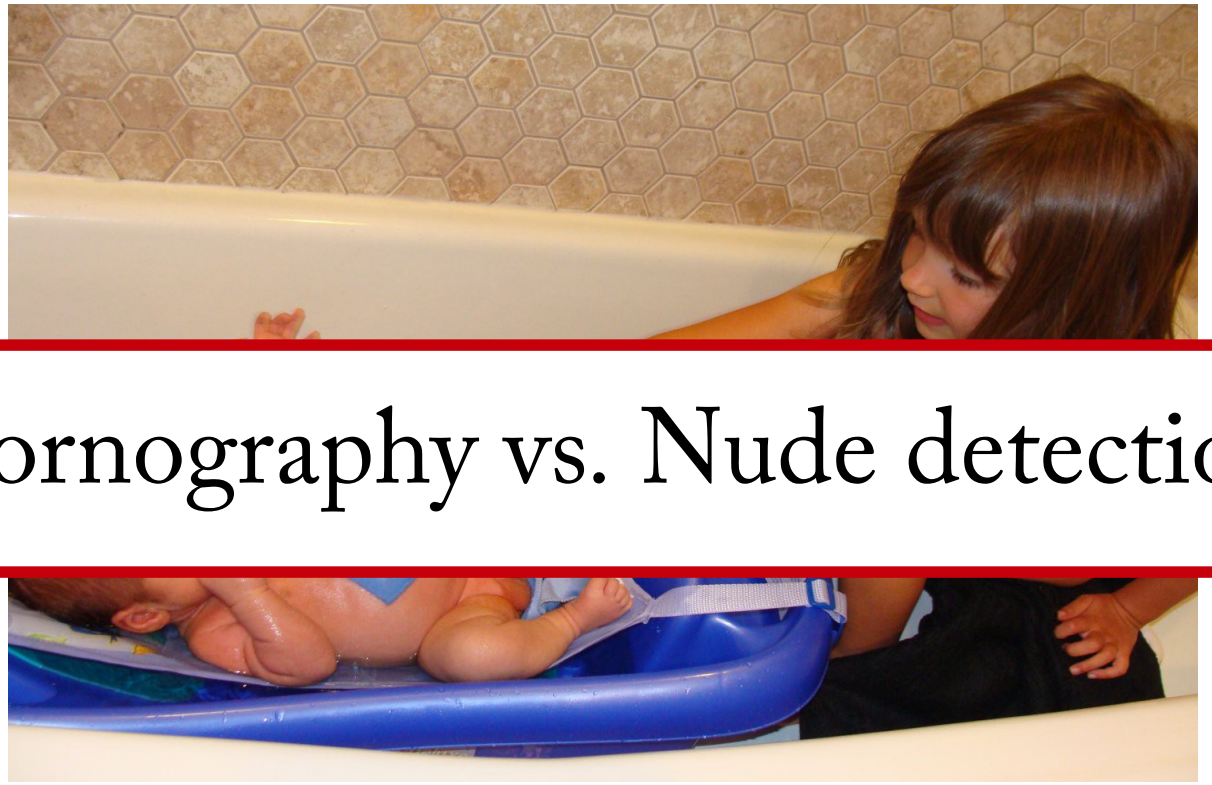
Pornography?

Ok, it is!

Pornography?

No, it isn't.

Pornography?



Pornography vs. Nude detection

No, it isn't.

Pornography is “any sexually **explicit** material with the **aim** of sexual arousal or fantasy.” [Short et al., 2012]

Short et al. (2012). A review of internet pornography use research: methodology and content from the past 10 years. Cyberpsychology, Behavior, and Social Networking, 15(1):13–23.

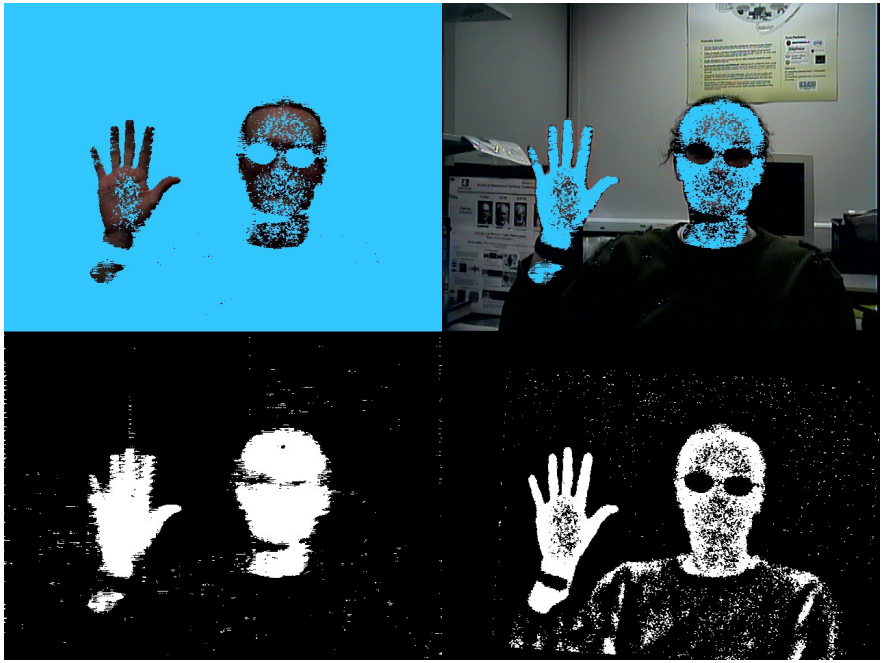
Existing solutions explore ...

In a nutshell ...

- Most work regarding the detection of pornographic material has been done for the **image domain**.
- The vast majority of those works is based on the detection of **human skin**.
- Few works have extracted **audio/spatiotemporal features**.
- Very recent methods have explored other possibilities, like **bag-of-words models**.

Existing solutions explore ...

Skin Detection



[Fleck et al., 1996]
[Forsyth and Fleck, 1999]
[Jones and Rehg, 2002]
[Rowley et al., 2006]
[Lee et al., 2007]
[Zuo et al., 2010]
[Hu et al., 2011]
[Ries and Lienhart, 2012]
[Kia et al., 2014]

Existing solutions explore ...

Skin Detection

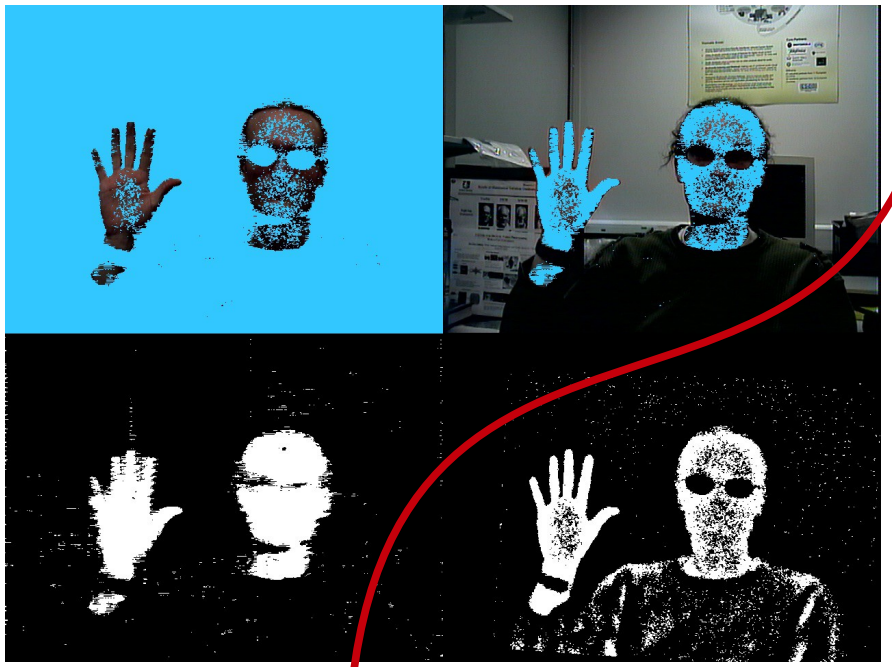


Image domain

[Fleck et al., 1996]
[Forsyth and Fleck, 1999]
[Jones and Rehg, 2002]
[Rowley et al., 2006]
[Lee et al., 2007]
[Zuo et al., 2010]
[Hu et al., 2011]
[Ries and Lienhart, 2012]
[Kia et al., 2014]

Skin detection

- Fleck et al. [1996], Forsyth and Fleck [1999] proposed to detect skin regions in an image and match them with human bodies by applying geometric grouping rules.

- Jones and Rehg [2002] focused on the detection of human skin by constructing RGB color histograms from a large database of skin and non-skin pixels.

Skin detection

- Rowley et al. [2006] used Jones and Rehg' skin color histograms in a system installed in Google's Safe Search.

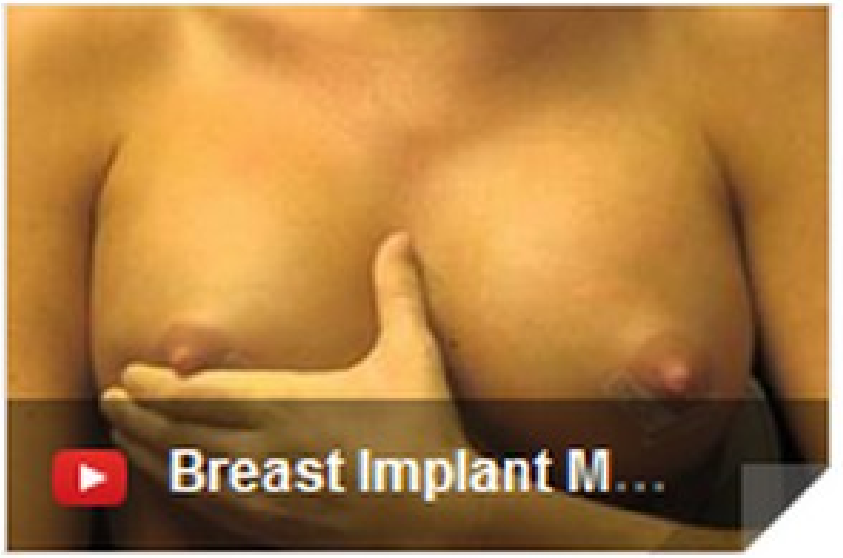
- Zuo et al. [2010] proposed a patch-based skin color detection that verifies whether all the pixels in a small patch correspond to human skin tone.

Skin detection

- Ries and Lienhart [2012] provided an **overview of state-of-the-art approaches** to visual adult image recognition, including human skin-based methods.

- Kia et al. [2014] extracted Fourier descriptors and signature of boundary of skin regions as shape features.

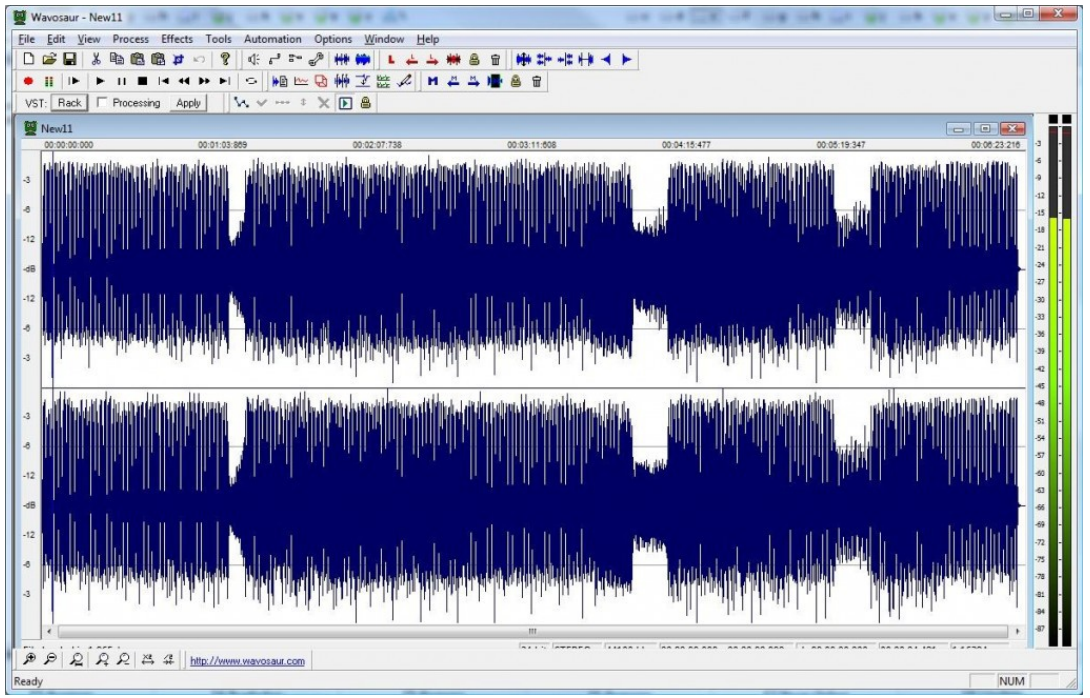
Skin detection: Problem?



False Positives!

Existing solutions explore ...

Audio Features



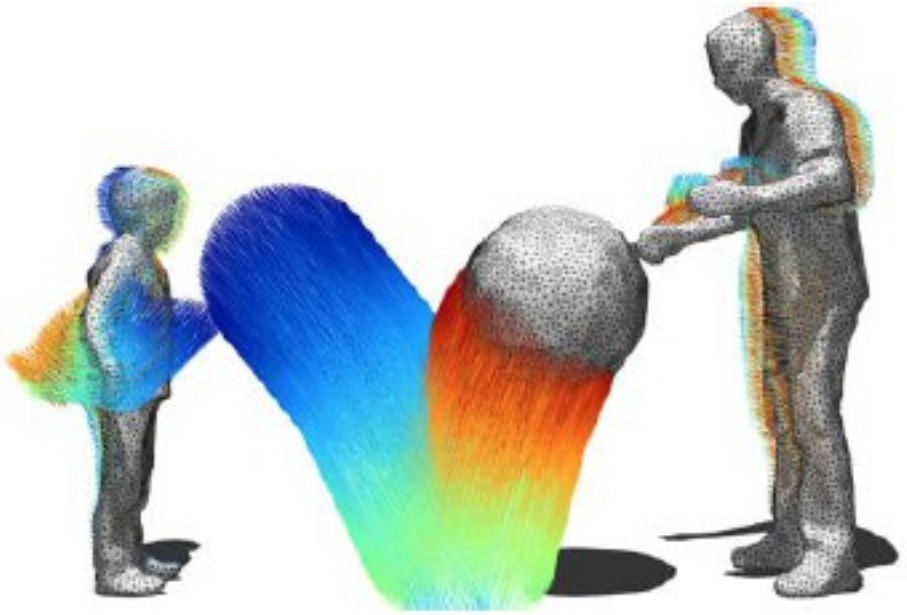
[Rea et al., 2006]
[Liu et al., 2011]
[Ulges et al., 2012]

Audio features

- Rea et al. [2006] combined skin color estimation with the detection of periodic patterns in a video's audio signal.
- Liu et al. [2011] demonstrated improvements by fusion visual features (color moments and edge histograms) with “audio words”.
- Ulges et al. [2012] proposed an approach of late fusing motion histograms with “audio words”.

Existing solutions explore ...

Spatiotemporal Features



[Tong et al., 2005]
[Endeshaw et al., 2008]
[Jansohn et al., 2009]
[Valle et al., 2012]
[Souza et al., 2012]

Spatiotemporal features

- Tong et al. [2005] proposed a method to estimate the period of a signal to classify periodic motion patterns.

- Endeshaw et al. [2008] developed a fast method for detection of indecent video content using repetitive movement analysis.

- Jansohn et al. [2009] introduced a framework that combines keyframe-based methods with a statistical analysis of MPEG-4 motion vectors.

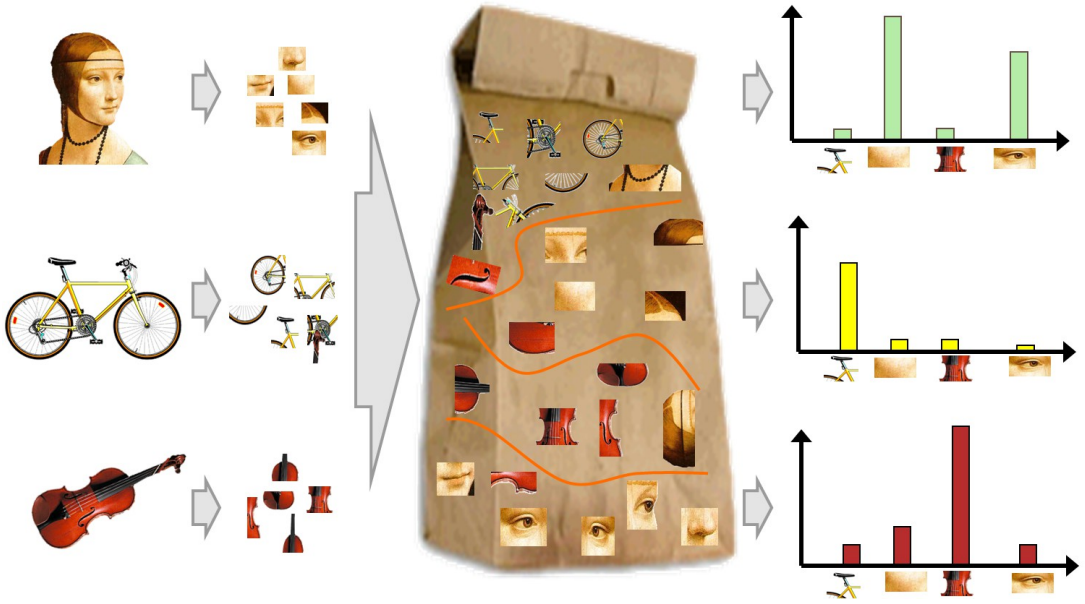
Spatiotemporal features

- Valle et al. [2012] compared the use of several features, including spatiotemporal local descriptors (STIP descriptors), in the pornography detection.

- Souza et al. [2012] evaluated the performance of the family of color-based STIPs in the pornography detection.

Existing solutions explore ...

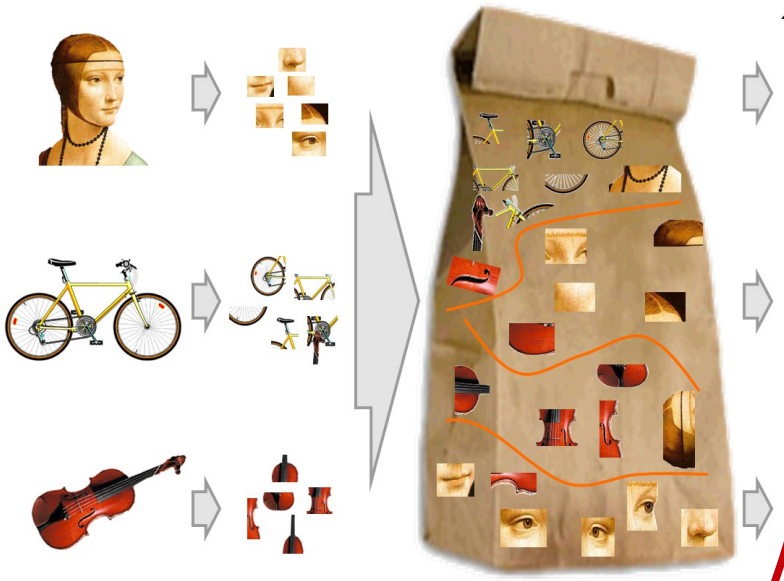
Bag-of-words



[Deselaers et al., 2008]
[Lopes et al., 2009a,b]
[Avila et al., 2011, 2013]
[Ulges and Stahl, 2011]
[Steel, 2012]
[Caetano et al., 2014a,b]

Existing solutions explore ...

Bag-of-words



[Deselaers et al., 2008]
[Lopes et al., 2009a,b]
[Avila et al., 2011, 2013]
[Ulges and Stahl, 2011]
[Steel, 2012]
[Caetano et al., 2014a,b]

BossaNova

BossaNova
Video Descriptor

Bag-of-words models

- Deselaers et al. [2008] first proposed a BoW model to filter pornographic images, which greatly improved the efficiency of the identification of pornographic images.

- Lopes et al. [2009ab] developed a BoW-based approach, which used the HueSIFT color descriptor, to classify images [Lopes et al., 2009b] and videos [Lopes et al., 2009a] of pornography.

Bag-of-words models

- Ulges and Stahl [2011] introduced a color enhanced visual word features in YUV color space to classify child pornography.

- Steel [2012] proposed a pornographic images recognition method based on visual words, by using mask-SIFT in a cascading classification system.

Bag-of-words models

- Avila et al. [2011, 2013] introduced the **BossaNova representation** (a BoW-based approach) and extracted the HueSIFT descriptors to classify pornographic videos.

- Caetano et al. [2014ab] extended the **BossaNova for video representation** and applied the **binary descriptors** (e.g., BRIEF, BRISK, BinBoost) for pornography detection.

Bag-of-Words Models

Bag-of-words (BoW) Models

[Sivic and Zisserman, 2003; Csurka et al., 2004; Boureau et al., 2010]

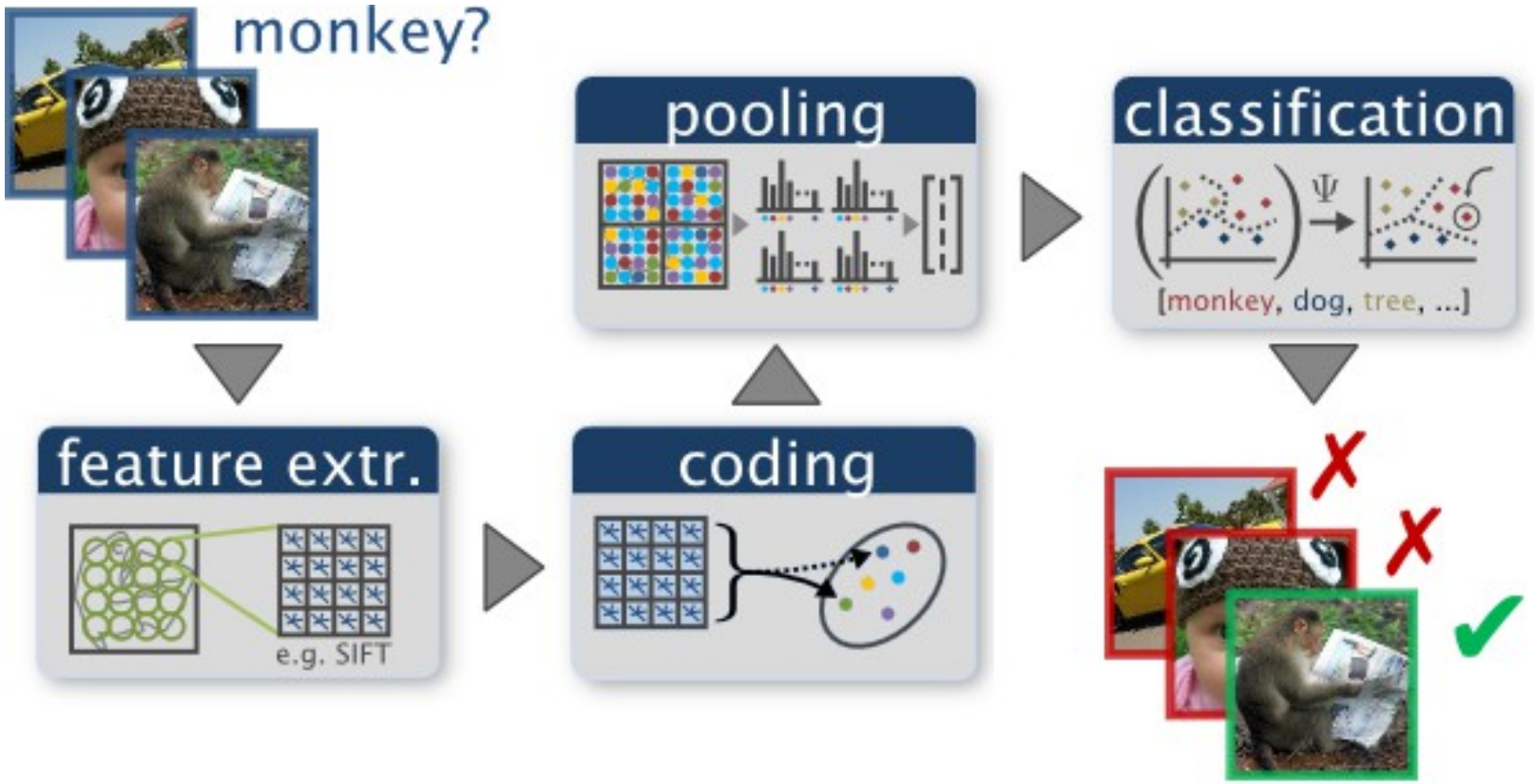
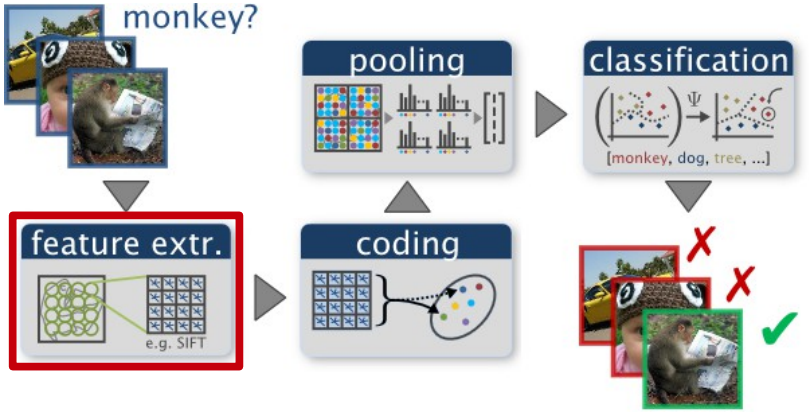
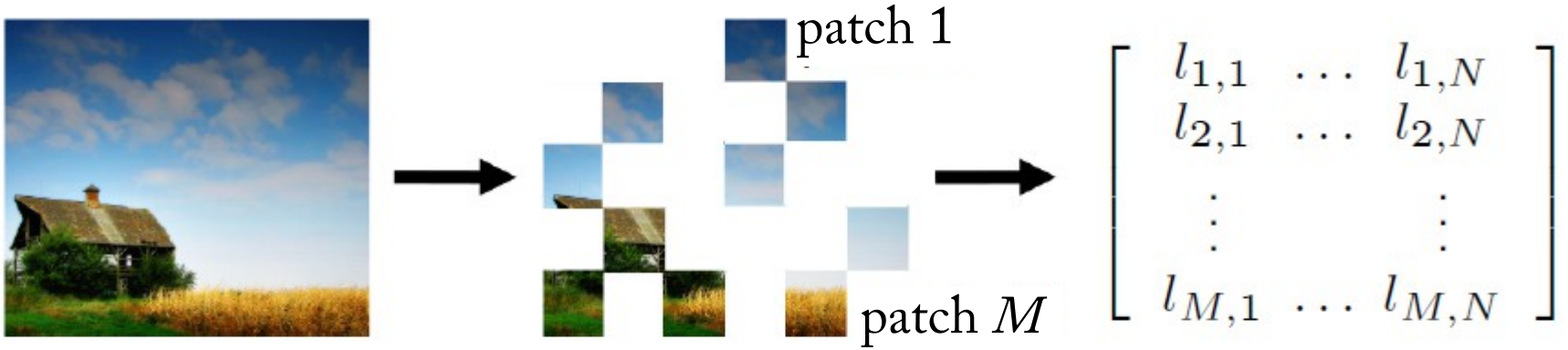


Figure credit: Ken Chatfield

Low-level Visual Feature Extraction

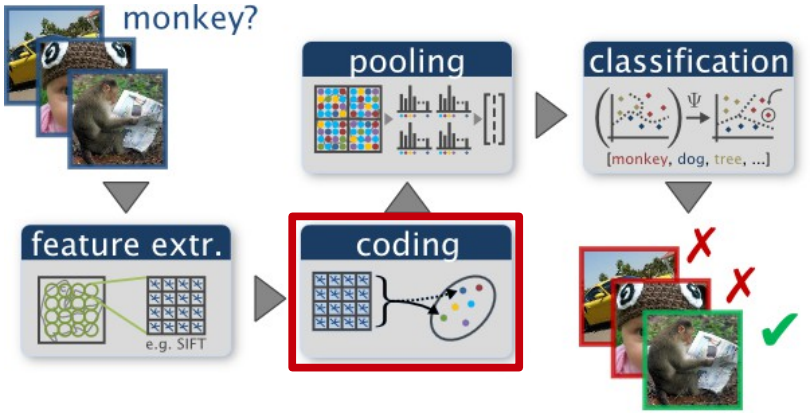




Local feature extraction

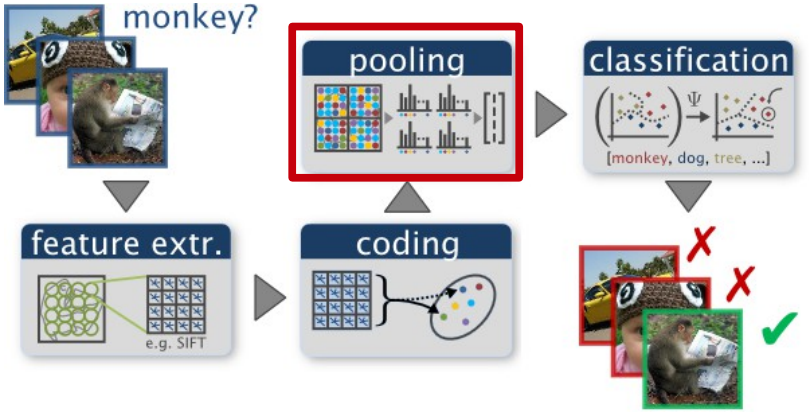
- **Patch detection:** interest points, dense sampling
- **Feature extraction:** SIFT [Lowe, 2004], SURF [Bay et al., 2008]

Visual Codebook Coding Step

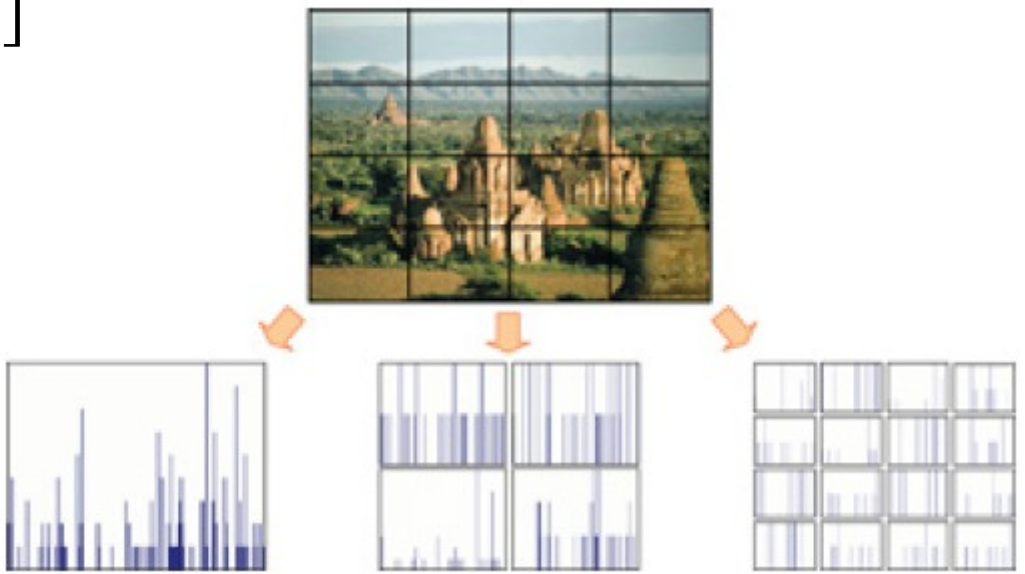


- **Visual codebook learning**: random, unsupervised (e.g., k-means, GMM), supervised [Perronnin et al., 2006; Goh et al., 2012]
- **Coding:** hard-assignment, soft-assignment [van Gemert et al., 2008, 2010], sparse coding [Yang et al., 2009; Boureau et al., 2010]
- **Feature coding based on the vector difference:** VLAD [Jégou et al., 2010], SVC [Zhou et al., 2010], VLAT [Picard et al., 2011], Fisher Vector [Perronnin et al., 2010]

Pooling Step

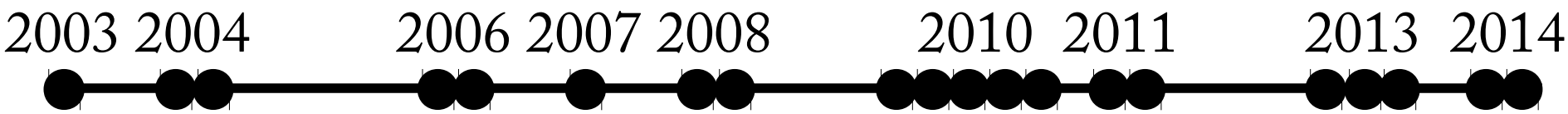


- **Pooling:** sum/average-pooling, max-pooling [Yang et al., 2009]
- **Spatial pooling:** spatial pyramid matching [Lazebnik et al., 2006], [Jia et al., 2012]



Spatial Pyramid Matching

Overview of BoW literature



[Sivic and Zisserman, 2003]
[Csurka et al., 2004]
[Lowe, 2004]
[Lazebnik et al., 2006]
[Bay et al., 2006]
[Perronnin and Dance, 2007]
[Tuytelaars and Mikolajczyk, 2008]
[Li and Allinson, 2008]
[Boureau et al., 2010]
[Perronninet al., 2010]
[Zhou et al., 2010]
[Jégou et al., 2010]
[van Gemert et al., 2010]
[Chatfield et al., 2011]
[Picard and Gosselin, 2011]
[Koniusz et al., 2013]
[Arandjelovic et al., 2013]
[Avila et al., 2013]
[Caetano et al., 2014]
[Murray and Perronnin, 2014]

Representing Local Binary Descriptors with BossaNova for Visual Recognition

Carlos Caetano, Sandra Avila, Silvio Guimarães, and Arnaldo Araújo

ACM Symposium on Applied Computing (SAC 2014)

Pornography Detection using BossaNova Video Descriptor

Carlos Caetano, Sandra Avila, Silvio Guimarães, and Arnaldo Araújo

European Signal Processing Conference (EUSIPCO 2014)

Representing Local Binary Descriptors with BossaNova for Visual Recognition

Carlos Caetano, Sandra Avila, Silvio Guimarães, and Arnaldo Araújo

ACM Symposium on Applied Computing (SAC 2014)

Binary Descriptors + BossaNova + Pornography

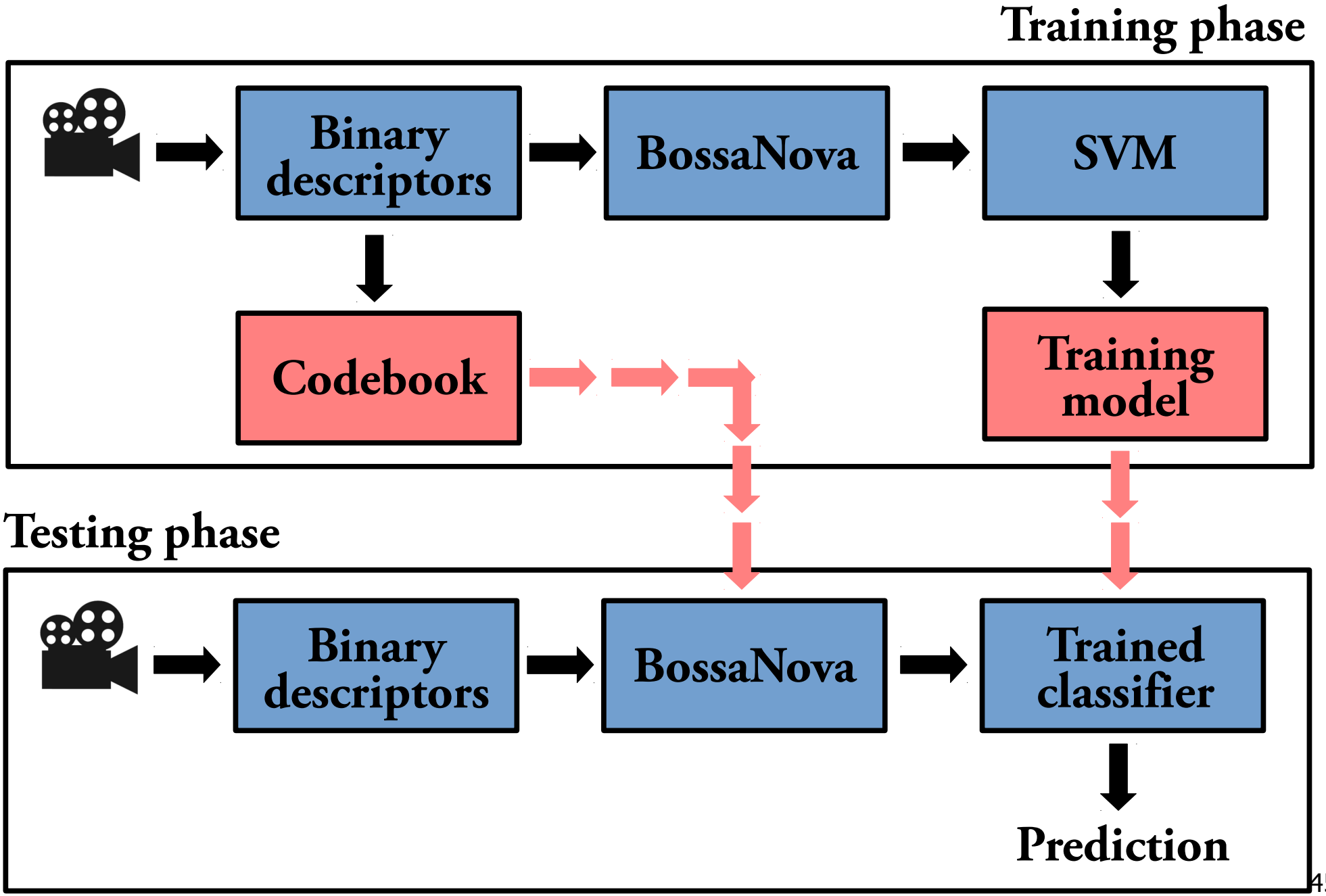
Pornography Detection using BossaNova Video Descriptor

Carlos Caetano, Sandra Avila, Silvio Guimarães, and Arnaldo Araújo

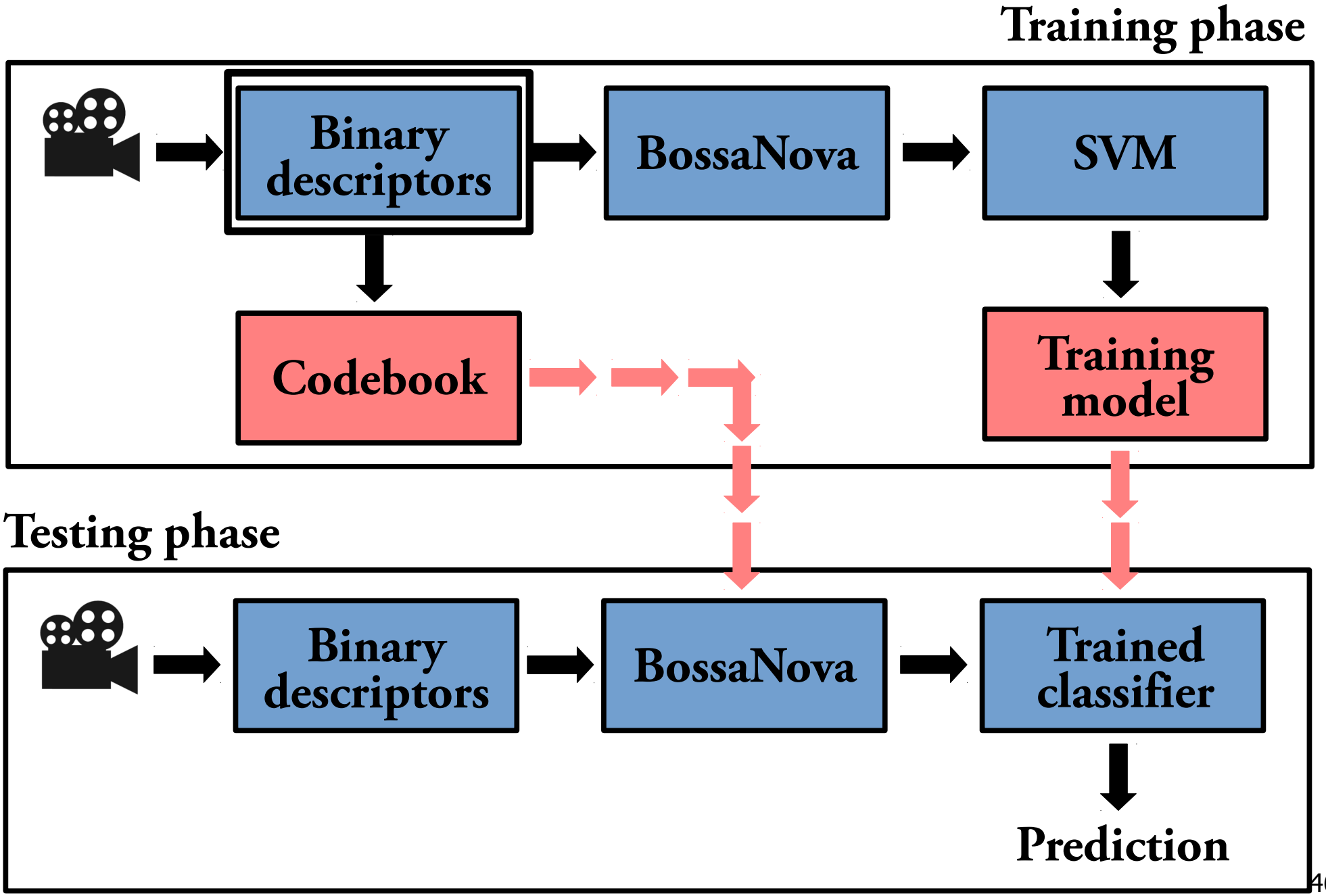
European Signal Processing Conference (EUSIPCO 2014)

Binary Descriptors + BossaNova Video Descriptor + Pornography

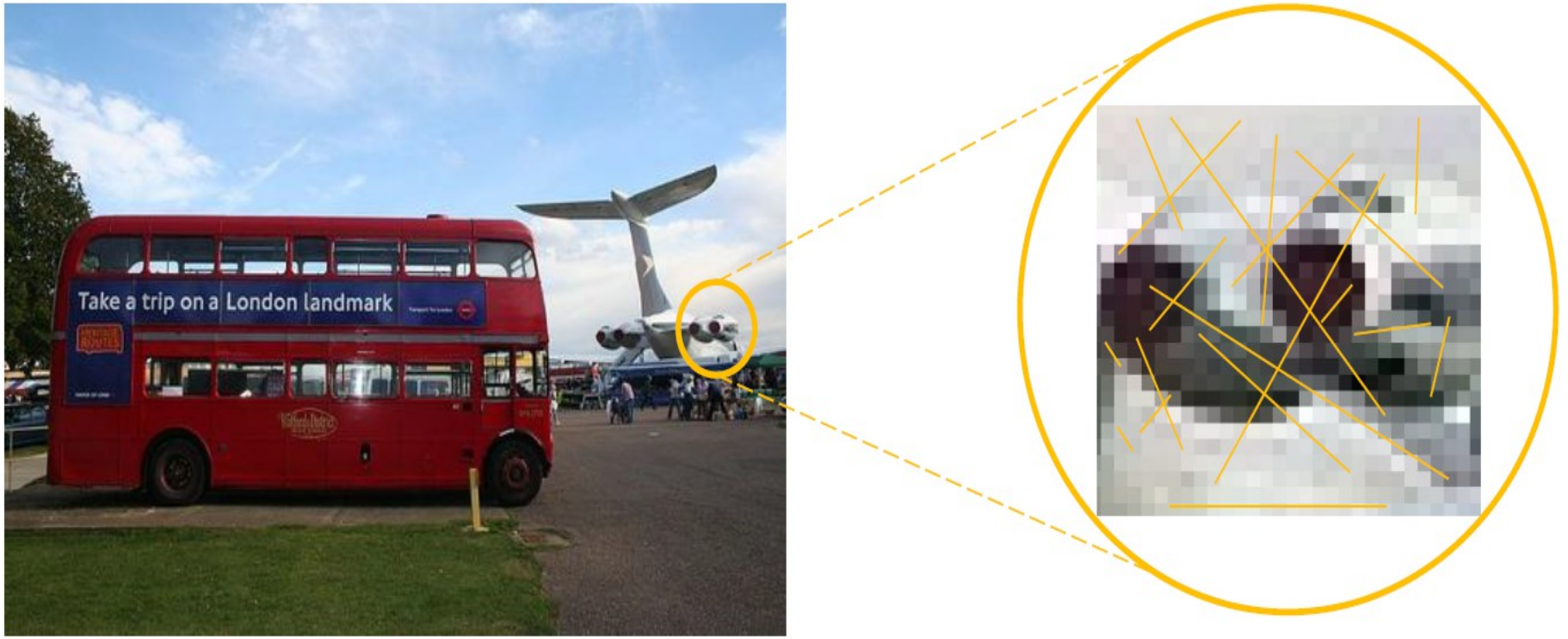
Framework


Training phase
Binary descriptors
BossaNova
SVM
Codebook
Training model
Testing phase
Binary descriptors
BossaNova
Trained classifier
Prediction

Framework


Training phase
Binary descriptors
BossaNova
SVM
Codebook
Training model
Testing phase
Binary descriptors
BossaNova
Trained classifier
Prediction

Binary descriptors: How it works



Pixel intensity comparisons!

Binary descriptors: Why?

Fast feature extraction!

No floating points!

Hamming distance!

Binary descriptors

BRIEF [Calonder et al., 2010]
D-BRIEF [Trzcinski et al., 2012]
ORB [Rublee et al., 2011]
BRISK [Leutenegger et al., 2011]
FREAK [Ortiz, 2012]
FRIF [Wang et al., 2013]
DRINK [Gadelha and Carvalho, 2014]

Pixel-based

BinBoost [Trzcinski et al., 2013]
BRIGHT [Iwamoto et al., 2013]
BiCE [Zitnick, 2010]
BGM [Trzcinski et al., 2012]

Gradient-based

Binary descriptors

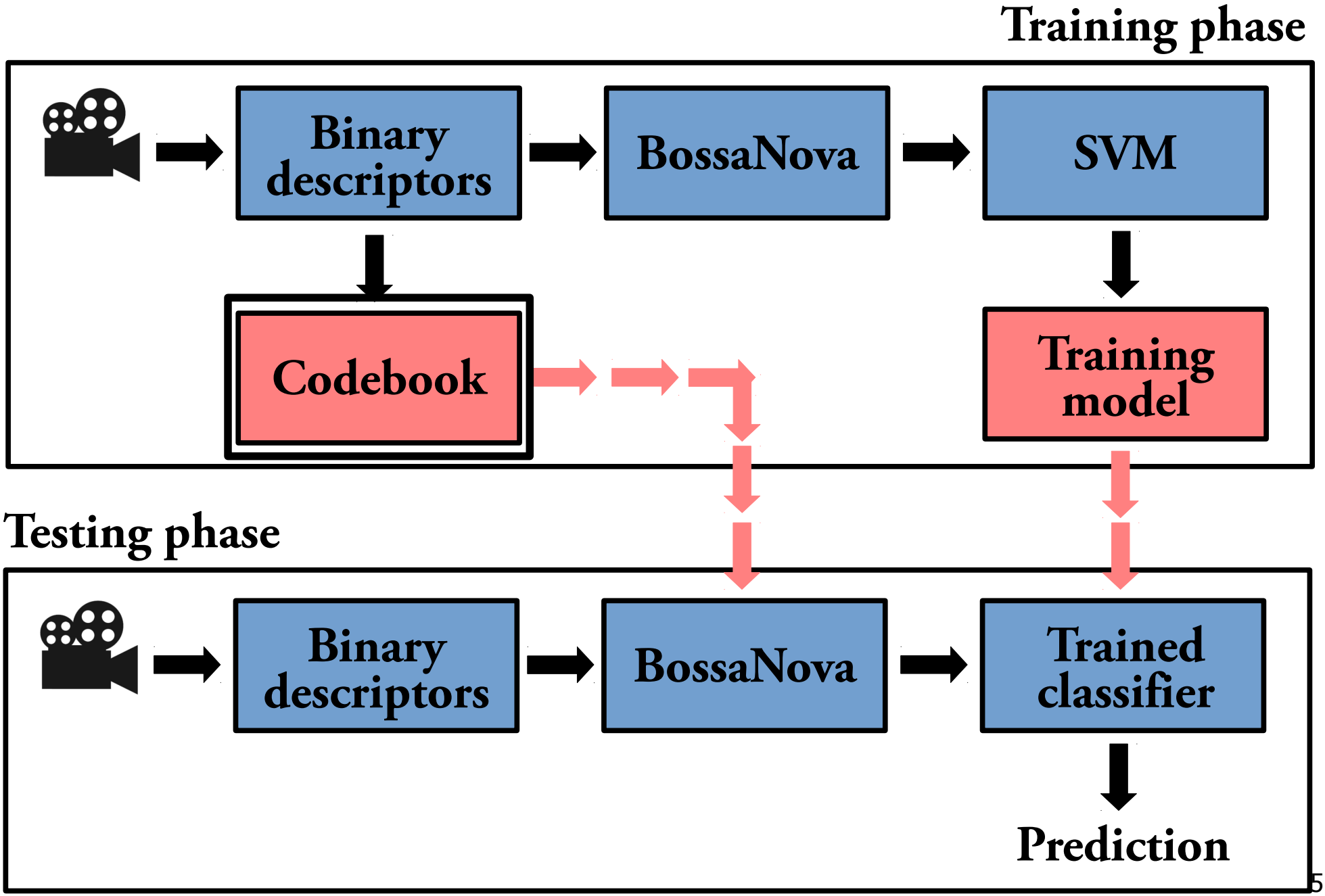
BRIEF [Calonder et al., 2010]
D-BRIEF [Trzcinski et al., 2012]
ORB [Rublee et al., 2011]
BRISK [Leutenegger et al., 2011]
FREAK [Ortiz, 2012]
FRIF [Wang et al., 2013]
DRINK [Gadelha and Carvalho, 2014]

Pixel-based (evaluated)

BinBoost [Trzcinski et al., 2013]
BRIGHT [Iwamoto et al., 2013]
BiCE [Zitnick, 2010]
BGM [Trzcinski et al., 2012]

Gradient-based (evaluated)

Framework


Training phase
Binary descriptors
BossaNova
SVM
Codebook
Training model
Testing phase
Binary descriptors
BossaNova
Trained classifier
Prediction

Codebook

k-medians:

Instead of calculating the mean for each cluster to determine its centroid, it calculates the median

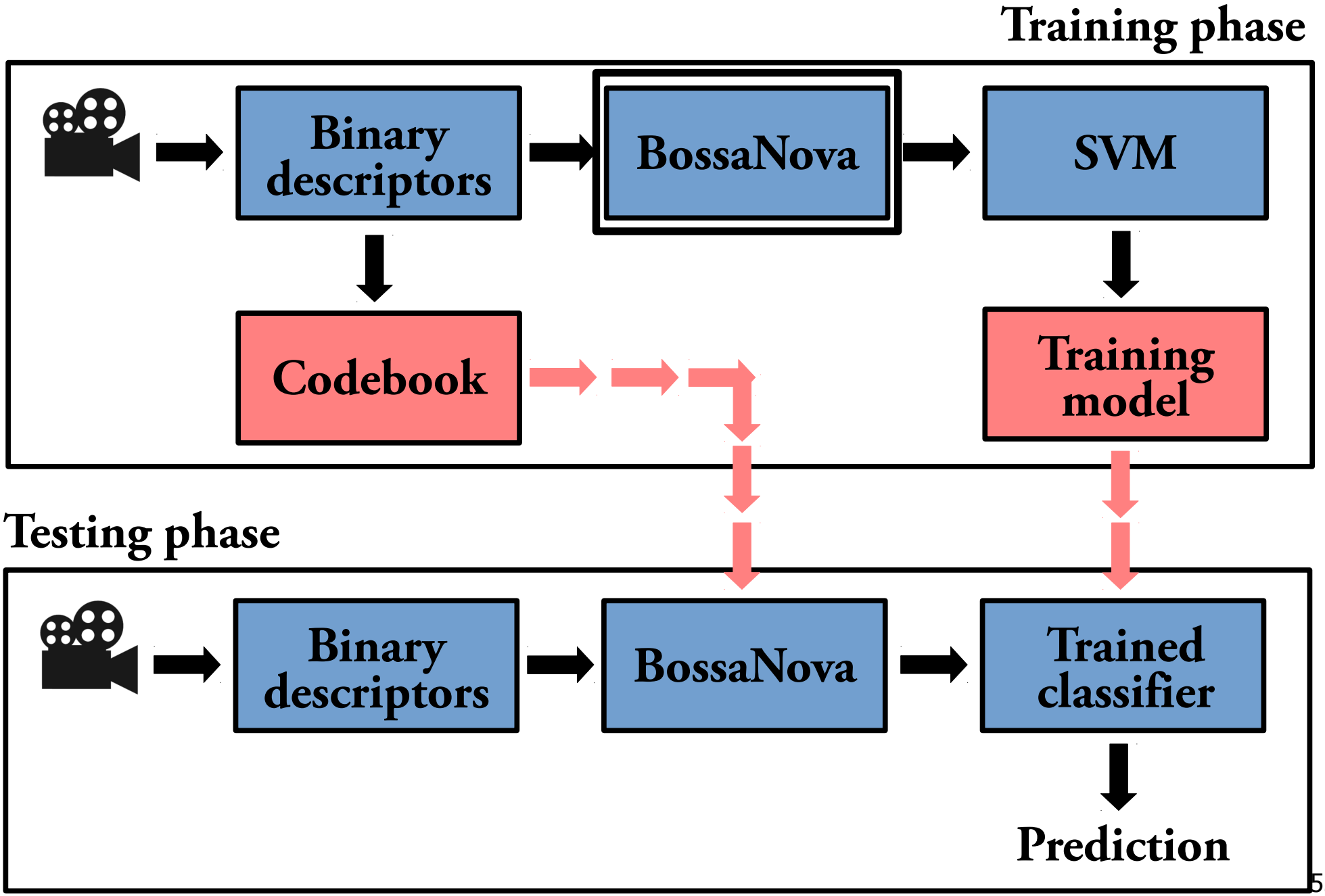
Hamming distance:

Hamming distance between:

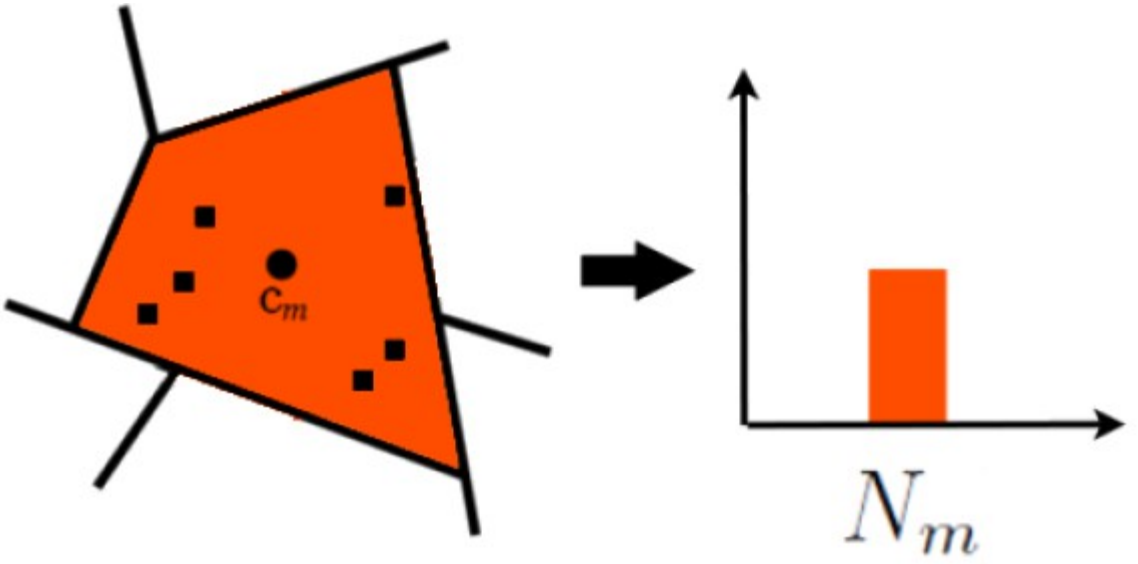
"Bed" and "Bad" is 1

1011001001 and 1001000011 is 3

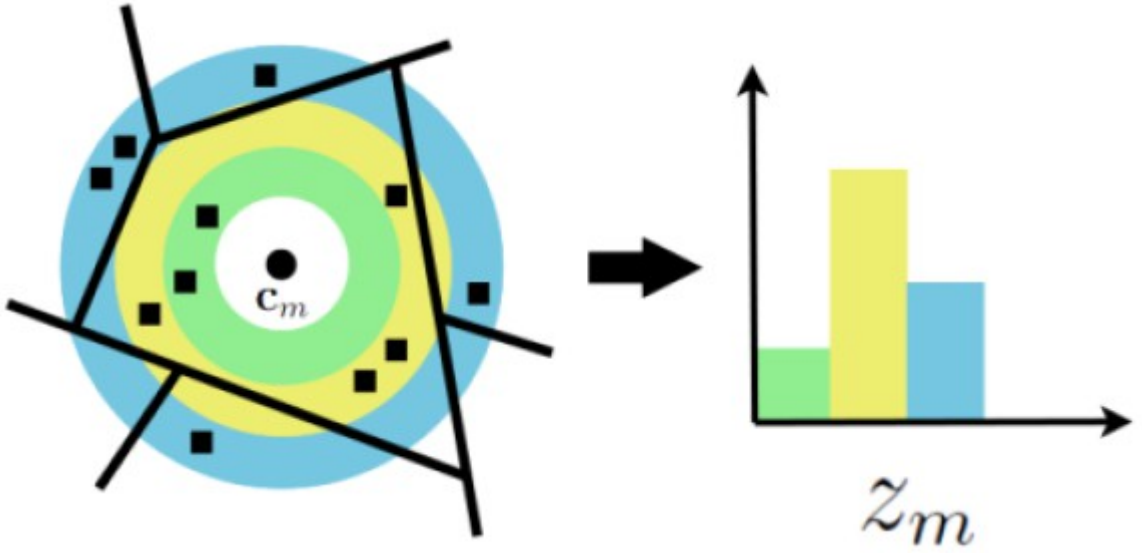
Framework


Training phase
Binary descriptors
BossaNova
SVM
Codebook
Training model
Testing phase
Binary descriptors
BossaNova
Trained classifier
Prediction

BossaNova [Avila et al., 2013]



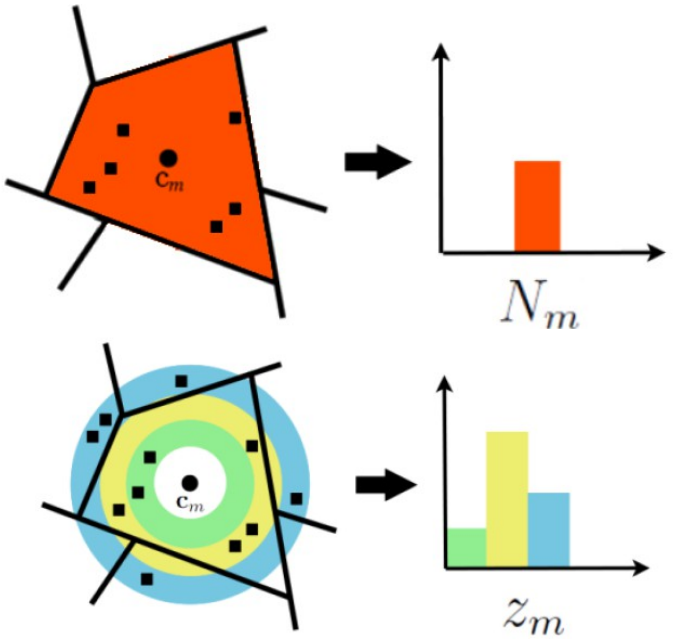
BoW Pooling



BossaNova Pooling

BossaNova Video Descriptor

[Caetano et al., 2014]

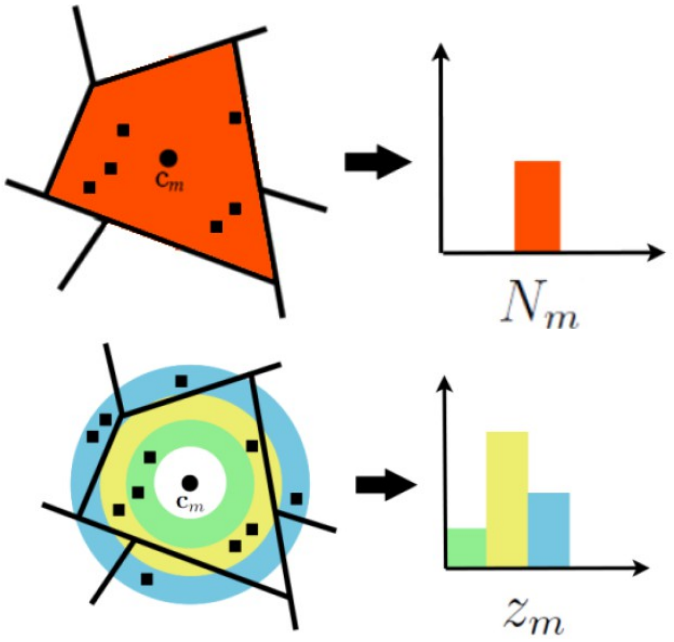


$$z = [[z_{m,b}], [N_m]]^T$$

BossaNova (image)

BossaNova Video Descriptor

[Caetano et al., 2014]



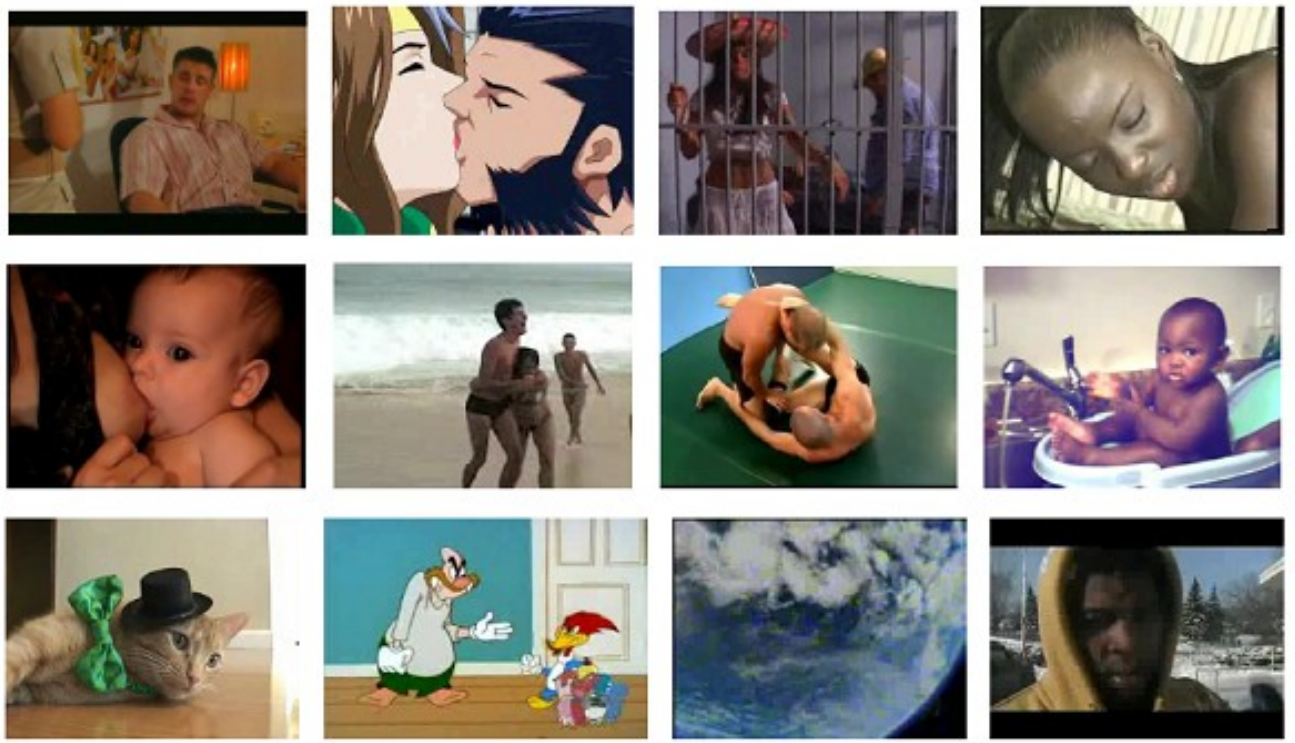
$$\mathbf{z} = [[z_{m,b}], [N_m]]^T$$

BossaNova (image)

BossaNova Video Descriptor:

$$\mathbf{h}(\{\mathbf{z}^i\}) = [[\text{median}(z^i_{m,b})], [\text{median}(N^i_m)]]^T$$

Pornography Dataset [Avila et al., 2013]



800 videos

~80 hours

2 seconds to 30 minutes

https://sites.google.com/site/pornographydatabase/

Results: SAC'14

Binary descriptors + BossaNova

Approach	Accuracy (%)
BossaNova & HueSIFT	89.5 ± 1
BossaNova & BRIEF	86.3 ± 3
BossaNova & ORB	86.5 ± 3
BossaNova & BRISK	88.6 ± 2
BossaNova & FREAK	86.9 ± 3

Results: EUSIPCO'14

Binary descriptors + BossaNova Video Descriptor

Approach	Accuracy (%)
BossaNova & HueSIFT	89.5 ± 1
BossaNova VD & BRIEF	89.0 ± 1
BossaNova VD & ORB	89.0 ± 1
BossaNova VD & BRISK	89.3 ± 1
BossaNova VD & FREAK	89.7 ± 2
BossaNova VD & BinBoost	**90.9 ± 1**

Conclusions

Binary descriptor can deal with such task!

- Comparable results, but **faster** and more **compact**
- BinBoost gives the best results

BossaNova Video Descriptor works!

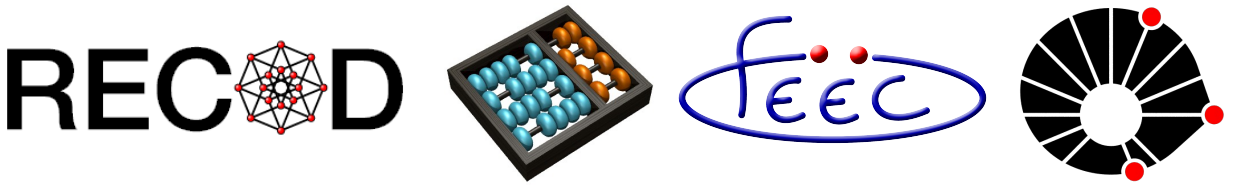
Thank You!

Sandra Avila

Postdoctoral Researcher

sandra@dca.fee.unicamp.br

https://sites.google.com/site/sandraefavila/



References

Fleck, M., Forsyth, D. A., and Bregler, C. (1996). Finding naked people. In *ECCV*, p. 593–602.

Forsyth, D. A. and Fleck, M. M. (1999). Automatic detection of human nudes. IJCV, 32(1):63–77.

Jones, M. J. and Rehg, J. M. (2002). Statistical color models with application to skin detection. IJCV, 46(1):81–96.

Rowley, H. A., Jing, Y., and Baluja, S. (2006). Large scale image-based adult-content filtering. In VISAPP, p. 290–296.

Lee, J.-S., Kuo, Y.-M., Chung, P.-C., and Chen, E.-L. (2007). Naked image detection based on adaptive and extensible skin color model. Pattern Recognition, 40(8):2261–2270.

Zuo, H., Hu, W., and Wu, O. (2010). Patch-based skin color detection and its application to pornography image filtering. In WWW, p. 1227–1228.

References

Hu, W., Zuo, H., Wu, O., Chen, Y., Zhang, Z., and Suter, D. (2011). Recognition of adult images, videos, and web page bags. TOMCCAP, 7S(1):28:1–28:24.

Ries, C. and Lienhart, R. (2012). A survey on visual adult image recognition. MTA, p. 1–28.

Kia, S., Rahmani, H., Mortezaei, R., Moghaddam, M., and Namaz, A. (2014). A novel scheme for intelligent recognition of pornographic images. Arxiv 1402.5792.

Rea, N., Lacey, G., Lambe, C., and Dahyot, R. (2006). Multimodal periodicity analysis for illicit content detection in videos. In CVMP, p. 106–114.

Liu, Y., Wang, X., Zhang, Y., and Tang, S. (2011). Fusing audio-words with visual features for pornographic video detection. In TRUSTCOM, p. 1488–1493.

Souza, F. and Valle, E. and Cámara-Chávez, G. and Araújo, A. (2012) An Evaluation on Color Invariant Based Local Spatiotemporal Features for Action Recognition. In SIBGRAPI.

References

Ulges, A., Schulze, C., Borth, D., and Stahl, A. (2012). Pornography detection in video benefits (a lot) from a multi-modal approach. In International Workshop on Audio and Multimedia Methods for Large-Scale Video Analysis, p. 21–26.

Tong, X., Duan, L., Xu, C., Tian, Q., Hanqing, L., Wang, J., , and Jin, J. (2005). Periodicity detection of local motion. In ICME, p. 650–653.

Endeshaw, T., Garcia, J., and Jakobsson, A. (2008). Fast classification of indecent video by low complexity repetitive motion detection. In Applied Imagery Pattern Recognition Workshop.

Jansohn, C., Ulges, A., and Breuel, T. M. (2009). Detecting pornographic video content by combining image features with motion information. In ACM International Conference on Multimedia.

Valle, E., Avila, S., da Luz Jr., A., Souza, F., Coelho, M., and Araújo, A. (2012). Content-based filtering for video sharing social networks. In SBSeg, p. 625-638.

References

Deselaers, T., Pimenidis, L., and Ney, H. (2008). Bag-of-visual-words models for adult image classification and filtering. In ICPR, p. 1–4.

Lopes, A., Avila, S., Peixoto, A., Oliveira, R., Coelho, M., and de A. Araújo, A. (2009a). Nude detection in video using bag-of-visual-features. In SIBGRAPI, p. 224–231.

Lopes, A., Avila, S., Peixoto, A., Oliveira, R., and de A. Araújo, A. (2009b). A bag-of-features approach based on hue-sift descriptor for nude detection. In EUSIPCO, p. 1552–1556.

Ulges, A. and Stahl, A. (2011). Automatic detection of child pornography using color visual words. In ICME, p. 1–6.

Steel, C. (2012). The mask-sift cascading classifier for pornography detection. In WorldCIS, p. 139–142.

Avila, S., Thome, N., Cord, M., Valle, E., and de A. Araújo, A. (2011). BOSSA: extended BoW formalism for image classification. In ICIP, p. 2909–2912.

References

Avila, S., Thome, N., Cord, M., Valle, E., and de A. Araújo, A. (2013). Pooling in image representation: the visual codeword point of view. CVIU, 117(5):453–465.

Caetano, C., Avila, S., Guimarães, S., and de A. Araújo. A. (2014). Representing local binary descriptors with BossaNova for visual recognition. In SAC, p. 49–54, 2014.

Caetano, C., Avila, S., Guimarães, S., and de A. Araújo. A. (2014). Pornography detection using BossaNova video descriptor. In EUSIPCO, 2014.

Sivic, J. and Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. In ICCV.

Csurka, G., Bray, C., Dance, C., and Fan, L. (2004). Visual categorization with bags of keypoints. In ECCV, pages 1–22.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. IJCV, 60:91–110.

References

Lazebnik, S., Schmid, C., and Ponce, J. (2006). Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In CVPR, pages 2169–2178

Bay, H., Tuytelaars, T., and Gool, L. V. (2006). SURF: Speeded up robust features. In ECCV, pages 404–417.

Perronnin, F. and Dance, C. (2007). Fisher kernels on visual vocabularies for image categorization. In CVPR.

Tuytelaars, T. and Mikolajczyk, K. (2008). Local invariant feature detectors: A survey. Foundations and Trends in Computer Graphics and Vision, 3(3):177–280.

Li, J. and Allinson, N. M. (2008). A comprehensive review of current local features for computer vision. Neurocomputing, 71(10-12):1771–1787.

Boureau, Y., Bach, F., LeCun, Y., and Ponce, J. (2010). Learning mid-level features for recognition. In CVPR, pages 2559–2566.

References

Perronnin, F., Sánchez, J., and Mensink, T. (2010c). Improving the Fisher Kernel for Large-Scale Image Classification. In ECCV, pages 143–156.

Zhou, X., Yu, K., Zhang, T., and Huang, T. (2010). Image classification using super-vector coding of local image descriptors. In ECCV, pages 141–154.

Jégou, H., Douze, M., Schmid, C., and Pérez, P. (2010). Aggregating local descriptors into a compact image representation. In CVPR, pages 3304–3311.

van Gemert, J., Veenman, C., Smeulders, A., and Geusebroek, J.-M. (2010). Visual word ambiguity. IEEE PAMI, 32:1271–1283.

Chatfield, K., Lempitsky, V., Vedaldi, A., and Zisserman, A. (2011). The devil is in the details: an evaluation of recent feature encoding methods. In BMVC.

Picard, D. and Gosselin, P. (2011). Improving image similarity with vectors of locally aggregated tensors. In ICIP, pages 669–672.

References

Koniusz, P., Yan, F., and Mikolajczyk, K. (2012). Comparison of mid-level feature coding approaches and pooling strategies in visual concept detection. CVIU.

Murray, N. and Perronnin F. (2014). Generalized Max Pooling. In CVPR.