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?
But ...
Are they safe?
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]
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

[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’s 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

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

- **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

Testing phase
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]

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

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

Pixel-based (evaluated)

Gradient-based (evaluated)
Framework

Training phase

1. Binary descriptors
2. BossaNova
3. SVM

Testing phase

1. Binary descriptors
2. BossaNova
3. Trained classifier
4. 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

1. Binary descriptors → BossaNova → SVM
2. Codebook
3. Training model

Testing phase

1. Binary descriptors → BossaNova → Trained classifier
2. Prediction
BossaNova [Avila et al., 2013]
BossaNova Video Descriptor
[Caetano et al., 2014]

\[ z = \begin{bmatrix} \bar{z}_{m,b} \\ N_m \end{bmatrix} \]

BossaNova (image)
BossaNova Video Descriptor
[Caetano et al., 2014]

\[ h(\{z^i\}) = [[\text{median}(z_{m,b}^i)], [\text{median}(N_m^i)]]^T \]
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

<table>
<thead>
<tr>
<th>Approach</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BossaNova &amp; HueSIFT</td>
<td>89.5 ± 1</td>
</tr>
<tr>
<td>BossaNova &amp; BRIEF</td>
<td>86.3 ± 3</td>
</tr>
<tr>
<td>BossaNova &amp; ORB</td>
<td>86.5 ± 3</td>
</tr>
<tr>
<td>BossaNova &amp; BRISK</td>
<td>88.6 ± 2</td>
</tr>
<tr>
<td>BossaNova &amp; FREAK</td>
<td>86.9 ± 3</td>
</tr>
</tbody>
</table>
Results: EUSIPCO'14

Binary descriptors + BossaNova Video Descriptor

<table>
<thead>
<tr>
<th>Approach</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BossaNova &amp; HueSIFT</td>
<td>89.5 ± 1</td>
</tr>
<tr>
<td>BossaNova VD &amp; BRIEF</td>
<td>89.0 ± 1</td>
</tr>
<tr>
<td>BossaNova VD &amp; ORB</td>
<td>89.0 ± 1</td>
</tr>
<tr>
<td>BossaNova VD &amp; BRISK</td>
<td>89.3 ± 1</td>
</tr>
<tr>
<td>BossaNova VD &amp; FREAK</td>
<td>89.7 ± 2</td>
</tr>
<tr>
<td><strong>BossaNova VD &amp; BinBoost</strong></td>
<td><strong>90.9 ± 1</strong></td>
</tr>
</tbody>
</table>
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


References


References


References


References
