Detecting Sexually Provocative Images

Debashis Ganguly,
Mohammad H. Mofrad,
Adriana Kovashka

Department of Computer Science,
University of Pittsburgh
Real Life Challenges

- Overwhelming amount of visual data on the Internet
- Parents may want to restrict the visual contents which their children can see.
- Lots of manual effort is invested by digital content administrators to classify images in age restricted categories.
Limitations of Existing Approaches

- Existing approaches detect pornographic contents based on percentage of skin area exposed by the subjects in such images.

Duan et. al., “Adult image detection method based on skin color model and support vector machine”, Asian Conference on Computer Vision 2002
Lee et. al., “Naked image detection based on adaptive and extensible skin color model”, Pattern recognition 2007
Limitations of Existing Approaches (contd.)

- Current methods cannot differentiate between pornographic content, portrait or harmless body shot like below.
17 types of Attributes composed from:

- Posture and gesture
  - Posture, gesture with fingers, movement, head position, direction of body and face relative to camera, etc.

- Facial expression
  - Mouth open or closed, type of smile, biting lips, eyebrows, eyelids, looking direction

- Scene context
  - Outdoor scene, outdoor events, indoor scenes with props or with flat background

- Skin exposure
  - Fully clothed, bare bodied, private body parts exposed

5 types of Moods and Emotions:

- Defensive, suggestive, playful, relaxed, upset

3 Sexual Intents

- Yes, maybe, no
Hierarchical Framework

Sexual Intent
Y M N

Moods and Emotions
D S P R U

Posture and Gesture
P M B F H S E W G

Facial Expressions
L E S D M B

Image Background
S E P F

Skin Exposure
C B P

Automatically Extracted Features
Color/SIFT/HOG/FC6/FC7/FC8
Experiments: Dataset

- 1,146 celebrity images
  - 203 Hollywood celebrities from people.com
  - 892 and 254 images of female and male candidates respectively
  - 5.6 images per person ratio

- 19 questions per image for annotations
  - Amazon Mechanical Turk by majority voting of 3 annotators per image
  - 70.5% annotator consensus
Experiments: Baseline

- Automatically extracted features
  - Low level features: Color histogram, SIFT, HOG using VLFeat
  - CaffeNet Features: FC6, FC7, FC8 using Caffe

- Direct model
  - Single level of classification hierarchy trained from automatically extracted features to predict sexual intent

  - Subset of features mapped based on relevance to our problem domain
Results: Overview

F-MEASURE

<table>
<thead>
<tr>
<th>Method</th>
<th>Direct</th>
<th>Hierarchical</th>
</tr>
</thead>
<tbody>
<tr>
<td>JOO ET. AL.</td>
<td>52</td>
<td>36</td>
</tr>
<tr>
<td>COLOR HISTOGRAM</td>
<td>35</td>
<td>53</td>
</tr>
<tr>
<td>SIFT</td>
<td>31</td>
<td>50</td>
</tr>
<tr>
<td>HOG</td>
<td>34</td>
<td>47</td>
</tr>
<tr>
<td>FC6</td>
<td>56</td>
<td>57</td>
</tr>
<tr>
<td>FC7</td>
<td>53</td>
<td>58</td>
</tr>
<tr>
<td>FC8</td>
<td>49</td>
<td>58</td>
</tr>
</tbody>
</table>
## Results: Overview

### ACCURACY

<table>
<thead>
<tr>
<th>Method</th>
<th>Direct</th>
<th>Hierarchical</th>
</tr>
</thead>
<tbody>
<tr>
<td>JOO ET. AL.</td>
<td>35</td>
<td>37</td>
</tr>
<tr>
<td>COLOR HISTOGRAM</td>
<td>35</td>
<td>41</td>
</tr>
<tr>
<td>SIFT</td>
<td>36</td>
<td>39</td>
</tr>
<tr>
<td>HOG</td>
<td>36</td>
<td>37</td>
</tr>
<tr>
<td>FC6</td>
<td>54</td>
<td>50</td>
</tr>
<tr>
<td>FC7</td>
<td>52</td>
<td>51</td>
</tr>
<tr>
<td>FC8</td>
<td>51</td>
<td>50</td>
</tr>
</tbody>
</table>
Results: Overview

SENSITIVITY

<table>
<thead>
<tr>
<th>Method</th>
<th>JOO ET. AL.</th>
<th>COLOR HISTOGRAM</th>
<th>SIFT</th>
<th>HOG</th>
<th>FC6</th>
<th>FC7</th>
<th>FC8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>52</td>
<td>43</td>
<td>47</td>
<td>40</td>
<td>71</td>
<td>66</td>
<td>59</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>47</td>
<td>75</td>
<td>70</td>
<td>83</td>
<td>83</td>
<td>83</td>
<td>84</td>
</tr>
</tbody>
</table>
Results: Overview

SPECIFICITY

<table>
<thead>
<tr>
<th>Method</th>
<th>Direct</th>
<th>Hierarchical</th>
</tr>
</thead>
<tbody>
<tr>
<td>JOO ET. AL.</td>
<td>23</td>
<td>29</td>
</tr>
<tr>
<td>COLOR HISTOGRAM</td>
<td>30</td>
<td>13</td>
</tr>
<tr>
<td>SIFT</td>
<td>28</td>
<td>14</td>
</tr>
<tr>
<td>HOG</td>
<td>34</td>
<td>15</td>
</tr>
<tr>
<td>FC6</td>
<td>43</td>
<td>28</td>
</tr>
<tr>
<td>FC7</td>
<td>43</td>
<td>28</td>
</tr>
<tr>
<td>FC8</td>
<td>45</td>
<td>27</td>
</tr>
</tbody>
</table>
Conclusion

- Our method enables automated contents classification based on behaviors and intents of the portrayed subjects.
- It allows prompt intervention of human experts upon integrating the proposed methodology with mobile apps, social media websites, and media streaming websites.