CAN A SMILE REVEAL YOUR GENDER?

JOURNEE de la BIOMETRIE 2017, Caen, 07/07/17

Antitza Dantcheva,
Piotr Bilinski,
Francois Bremond
INRIA Sophia Antipolis, France
1. Why Gender Estimation?
2. Related Work: appearance based gender estimation, other modalities
3. Scope of the Work
4. Proposed Method
5. Dense Trajectories
6. UVA Nemo Dataset
7. Baseline Algorithms
8. Results: true gender classification rates, posed smile, pertinent features
9. Conclusions
WHY GENDER ESTIMATION?

Gender can be used as a soft biometric, in order to

- index databases
- enhance the performance of e.g. face recognition

Or in applications such as:

- video surveillance
- human computer-interaction
- anonymous customized advertisement
- image retrieval
RELATED WORK 1: AUTOMATED GENDER ESTIMATION

Different biometric modalities: fingerprint, face, iris, voice, body shape, gait, signature, DNA, as well as clothing, hair, jewelry and even body temperature.

- **Dynamics - body-based** classification of gender
  - Cues include body sway, waist-hip ratio, and shoulder-hip ratio [1] for example
  - females have a distinct waist-to-hip ratio and swing their hips more, whereas males have broader shoulders and swing their shoulders more

- 3D facial images [2]
- Near-infrared (NIR) and thermal facial images [3]


CAN A SMILE REVEAL YOUR GENDER? A. DANTCHEVA

INRIA
## RELATED WORK 2: APPEARANCE BASED GENDER ESTIMATION [4]

<table>
<thead>
<tr>
<th>Work</th>
<th>Features</th>
<th>Classifier</th>
<th>Datasets used for evaluation</th>
<th>Performance numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gutta et al. (1998)</td>
<td>Raw pixels</td>
<td>Hybrid classifier</td>
<td>FERET, 3006 images</td>
<td>96.0%</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Nazhir et al. (2010)</td>
<td>DCT</td>
<td>KNN</td>
<td>SUMS, 400 images</td>
<td>99.3%</td>
</tr>
<tr>
<td>Ross and Chen (2011)</td>
<td>LBP</td>
<td>SVM</td>
<td>CBSR NIR, 3,200 images</td>
<td>93.59%</td>
</tr>
<tr>
<td>Cao et al. (2011)</td>
<td>Metrology</td>
<td>SVM</td>
<td>MUCT, 276 images</td>
<td>86.83%</td>
</tr>
<tr>
<td>Hu et al. (2011)</td>
<td>Filter banks</td>
<td>SVM</td>
<td>Flickr, 26,700 images</td>
<td>90.1%</td>
</tr>
<tr>
<td>Bekios-Calfa et al. (2011)</td>
<td>SDA</td>
<td>PCA</td>
<td>Multi-PIE, 337 images</td>
<td>88.04%</td>
</tr>
<tr>
<td>Shan (2012)</td>
<td>Boosted LBP</td>
<td>SVM</td>
<td>LFW, 7,443 images</td>
<td>94.81%</td>
</tr>
<tr>
<td>Ramon-Balmaseda (2012)</td>
<td>LBP</td>
<td>SVM</td>
<td>MORPH, LFW, Images of Groups, 17,814 images</td>
<td>75.10%</td>
</tr>
<tr>
<td>Jia and Cristianini (2005)</td>
<td>Multi-scale LBP</td>
<td>C-Pegasos</td>
<td>Private, 4 million images</td>
<td>96.86%</td>
</tr>
</tbody>
</table>

RELATED WORK 3: INTUITION FOR APPROACH

Male and female smile-dynamics differ in parameters such as intensity and duration [7].

Gender-dimorphism in the human expression, from cognitive-psychology [5, 6]

- females smile more frequently than males
- females are more accurate expressers of emotion
- men exhibiting restrictive emotionality
- gender-based difference in emotional expression starts as early as in 3 months old, shaped by how caregivers interact to males and females
- happiness and fear are assigned as female-gender-stereotypical expressions

SCOPE OF THE WORK

Goal of the work: Determine gender based on facial behavior in video sequences capturing smiling subjects.

Proposed approach can be instrumental in cases of:

- Omitted appearance information (e.g. low resolution)
- Gender spoofing (e.g. makeup-based face alteration)
- As well as to improve the performance of appearance-based algorithms - complementary
PROPOSED METHOD

Spatio-temporal features based on dense trajectories [8] represented by a set of descriptors encoded by Fisher Vectors [9].

DENSE TRAJECTORIES: VISUALIZATION
PROPOSED METHOD

We extract 5 features aligned with trajectories to characterize:

- **Shape** (sequence of displacement vectors normalized by the sum of displacement vector magnitudes)
- **Appearance** (Histogram of Oriented Gradients (HOG) and Video Covariance Matrix Logarithm (VCML))
- **Motion** (Histogram of Optical Flow (HOF) and Motion Boundary Histogram (MBH))
PROPOSED METHOD

Fisher Vectors:
- extension of bag-of-features approach
- local features -> deviation from a “universal” generative Gaussian Mixture Model (GMM)

Classification with Support Vector Machines
- five-folds cross-validation (balanced number of males and females)
- per split: mean person accuracy (i.e., mean accuracy from various video instances of a person, if applicable), average value from all subjects belonging to this split
- final accuracy (True Gender Classification Rate): averaging the accuracy from all splits
UVA NEMO DATASET

Website: http://www.uva-nemo.org

1-2 video sequences of 400 subjects
185 female, 215 male subjects

Trigger of smile: short funny video segment

Age-range: 8 - 76 years.

Facial-dynamics change significantly for adolescents

we present our results based on two age-categories

BASELINE ALGORITHMS

- OpenBR [12]: Publicly available toolkit for biometric recognition
  - Algorithm: Local binary pattern (LBP) - histograms -> scale-invariant feature transform (SIFT) - features on a dense grid of patches -> PCA projection -> Support Vector Machine (SVM) for classification

- how-old.net: underlying algorithm and the training dataset: not known

---

- Dynamics algorithm based on facial landmarks [13]

- Commercial Off-The-Shelf (COTS): underlying algorithm and training dataset: not known


DYNAMICS BASED ON FACIAL LANDMARKS

Facial landmark detection

Signal displacement of facial landmarks

Feature Selection
Min-Redundancy
Max-Relevance (mRMR)

Classification

Male
Female

CAN A SMILE REVEAL YOUR GENDER? A. DANTCHEVA
## RESULTS 1: TRUE GENDER CLASSIFICATION RATES

<table>
<thead>
<tr>
<th>Age</th>
<th>≤20 (148)</th>
<th>&gt;20 (209)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenBR</td>
<td>52.3%</td>
<td>75.6%</td>
</tr>
<tr>
<td>how-old.net</td>
<td>55.5%</td>
<td>92%</td>
</tr>
<tr>
<td>COTS</td>
<td>76.9%</td>
<td>92.5%</td>
</tr>
<tr>
<td>Dynamics based on facial landmarks</td>
<td>59.4%</td>
<td>67.8%</td>
</tr>
<tr>
<td>COTS + Dynamics based on facial landmarks</td>
<td>76.9%</td>
<td>93%</td>
</tr>
<tr>
<td>Motion-based descriptors</td>
<td>77.7%</td>
<td>80.1%</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>86.3%</td>
<td>91%</td>
</tr>
</tbody>
</table>
## RESULTS 2: 5-CROSS-VALIDATION

<table>
<thead>
<tr>
<th>Trial</th>
<th>Train</th>
<th>Test</th>
<th>OpenBR</th>
<th>How-old.net</th>
<th>Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>199(100/99)</td>
<td>48(24/24)</td>
<td>50%</td>
<td>56.2%</td>
<td>82.8%</td>
</tr>
<tr>
<td>2</td>
<td>197(99/98)</td>
<td>50(25/25)</td>
<td>52%</td>
<td>56%</td>
<td>87.9%</td>
</tr>
<tr>
<td>3</td>
<td>201(101/100)</td>
<td>46(23/23)</td>
<td>56.9%</td>
<td>54.9%</td>
<td>82.8%</td>
</tr>
<tr>
<td>4</td>
<td>196(98/98)</td>
<td>51(26/25)</td>
<td>49%</td>
<td>52.9%</td>
<td>91.4%</td>
</tr>
<tr>
<td>5</td>
<td>195(98/97)</td>
<td>52(26/26)</td>
<td>53.8%</td>
<td>57.7%</td>
<td>86.7%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td></td>
<td>52.3%</td>
<td>55.5%</td>
<td>86.3%</td>
</tr>
</tbody>
</table>

True gender classification rates for subjects ≤ 20 years old

<table>
<thead>
<tr>
<th>Trial</th>
<th>Train</th>
<th>Test</th>
<th>OpenBR</th>
<th>How-old.net</th>
<th>Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>278(142/136)</td>
<td>72(37/35)</td>
<td>80.6%</td>
<td>90.3%</td>
<td>95.4%</td>
</tr>
<tr>
<td>2</td>
<td>283(145/138)</td>
<td>67(34/33)</td>
<td>85.1%</td>
<td>89.5%</td>
<td>86.9%</td>
</tr>
<tr>
<td>3</td>
<td>278(142/136)</td>
<td>72(37/35)</td>
<td>77.8%</td>
<td>93.1%</td>
<td>91.7%</td>
</tr>
<tr>
<td>4</td>
<td>285(145/140)</td>
<td>65(33/32)</td>
<td>72.3%</td>
<td>93.8%</td>
<td>92.9%</td>
</tr>
<tr>
<td>5</td>
<td>276(141/135)</td>
<td>74(36/38)</td>
<td>62.2%</td>
<td>93.2%</td>
<td>88.4%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td></td>
<td>75.6%</td>
<td>92%</td>
<td>91%</td>
</tr>
</tbody>
</table>

True gender classification rates for subjects > 20 years old

CAN A SMILE REVEAL YOUR GENDER? A. DANTCHEVA
**POSED SMILE (DYNAMICS BASED ON FACIAL LANDMARKS)**

<table>
<thead>
<tr>
<th>Age (Subject amount)</th>
<th>≤20 (143)</th>
<th>&gt;20 (225)</th>
</tr>
</thead>
<tbody>
<tr>
<td>how-old.net</td>
<td>51.1%</td>
<td>93.8%</td>
</tr>
<tr>
<td>COTS</td>
<td>76.9%</td>
<td>92%</td>
</tr>
<tr>
<td>Dynamics based on facial landmarks</td>
<td>59.4%</td>
<td>66.2%</td>
</tr>
<tr>
<td>COTS + Dynamics based on facial landmarks</td>
<td>76.9%</td>
<td>92.9%</td>
</tr>
</tbody>
</table>

PERTINENT FEATURES (DYNAMICS BASED ON FACIAL LANDMARKS)

Adolescents: females show longer Duration Ratio (Offset) and Duration (Onset) on the right side of the mouth and a larger Amplitude Ratio (Onset) on the left side of the mouth, than males.

In adults, females show: a larger Mean Amplitude (Apex) of mouth opening, a higher Maximum Amplitude on the right side of the mouth, as well as a shorter Mean Speed Offset on the left side of the mouth, than males.

(a) $D_{11}$ Mean Amplitude Apex
(b) $D_{8}$ Maximum Amplitude
(c) $D_{9}$ Mean Speed Offset

CONCLUSIONS

- Proposed a novel gender estimation approach, based on smiling-behavior.
- Proposed algorithm utilizes dense trajectories represented by spatio-temporal facial features and Fisher Vector encoding.
- Benefits: being complementary to appearance-based approaches, and thus being robust to gender spoofing.
- Results suggest that for adolescents, our approach significantly outperforms the existing state-of-the-art appearance-based algorithms, while for adults, our approach is comparatively discriminative to the state-of-the-art algorithms.
- A smile can reveal your gender 😊

Future work will seek to estimate soft biometrics analyzing additional facial expressions.
THANK YOU FOR YOUR ATTENTION!
Optical flow represents the absolute motion between two frames, which contains motion from many sources, i.e., foreground object motion and background camera motion. If camera motion is considered as action motion, it may corrupt the action classification. Various types of camera motion can be observed in realistic videos, e.g., zooming, tilting, rotation, etc. In many cases, camera motion is locally translational and varies smoothly across the image plane.

Dalal et al. proposed the motion boundary histograms (MBH) descriptor for human detection by computing derivatives separately for the horizontal and vertical components of the optical flow. The descriptor encodes the relative motion between pixels, as shown in Figure 5 (right). Since MBH represents the gradient of the optical flow, locally constant camera motion is removed and information about changes in the flow field (i.e., motion boundaries) is kept. MBH is more robust to camera motion than optical flow, and thus more discriminative for action recognition.

In this work, we employ MBH as motion descriptor for trajectories. The MBH descriptor separates optical flow into its horizontal and vertical components. Spatial derivatives are computed for each of them and orientation information is quantized into histograms. The magnitude is used for weighting. We obtain a 8-bin histogram for each component (i.e., MBHx and MBHy). Both histogram vectors are normalized separately with their L2 norm. Compared to video stabilization [35] and motion compensation [34], this is a simpler way to discount for camera motion.
Based on a covariance matrix representation, and it models relationships between different low-level features, such as intensity and gradient.

-> Encode appearance information of local spatio-temporal video volumes, which are extracted by the Dense Trajectories.