CAN A SMILE REVEAL YOUR GENDER?

Antitza Dantcheva,
Piotr Bilinski,
Francois Bremond
INRIA Sophia Antipolis, France

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

OUTLINE

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]

[1] G. Mather and L. Murdoch. Gender discrimination in biological motion displays based on dynamic cues. In Biological Sciences B, 1994.
[2] X. Han, H. Ugail, and I. Palmer. Gender classification based on 3D face geometry features using SVM. In Proceedings of International Conference on CyberWorlds (CW), 2009.
[3] C. Chen and A. Ross. Evaluation of gender classification methods on thermal and near-infrared face images. IJCB 2011.

RELATED WORK 2: APPEARANCE BASED GENDER ESTIMATION [4]

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

[4] Dantcheva, A.; Elia, P.; Ross, A.: What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics. IEEE Transactions on Information Forensics and Security (TIFS), 2016.

RELATED WORK 3: INTUITION FOR APPROACH

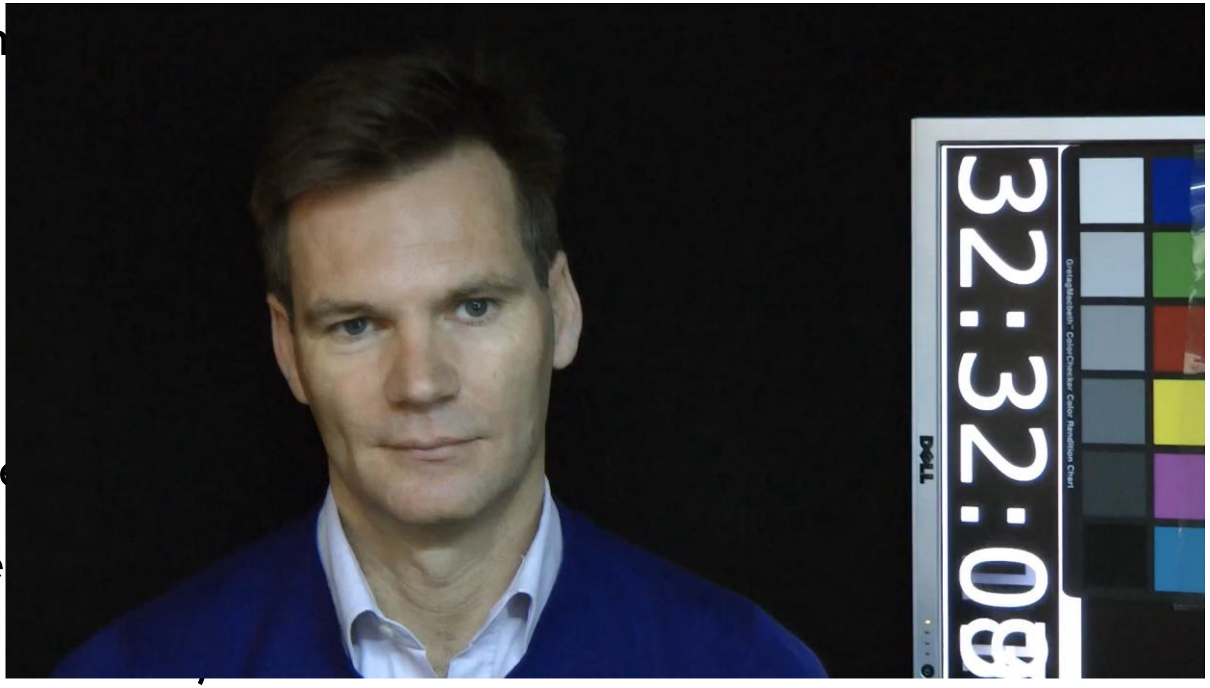
Male and female smile-dynamics differ in parameters such as intensity and duration [7].

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

[5] Cashdan, E.: Smiles, Speech, and Body Posture: How Women and Men Display Sociometric Status and Power. Journal of Nonverbal Behavior, 1998.
[6] Adams, R. B.; Hess, U.; Kleck, R. E.: The intersection of gender-related facial appearance and facial displays of emotion. Emotion Review, 2015.
[7] Dantcheva, A.; Bremond, F.: Gender estimation based on smile-dynamics. IEEE Transactions on Information Forensics and Security (TIFS), 2016.

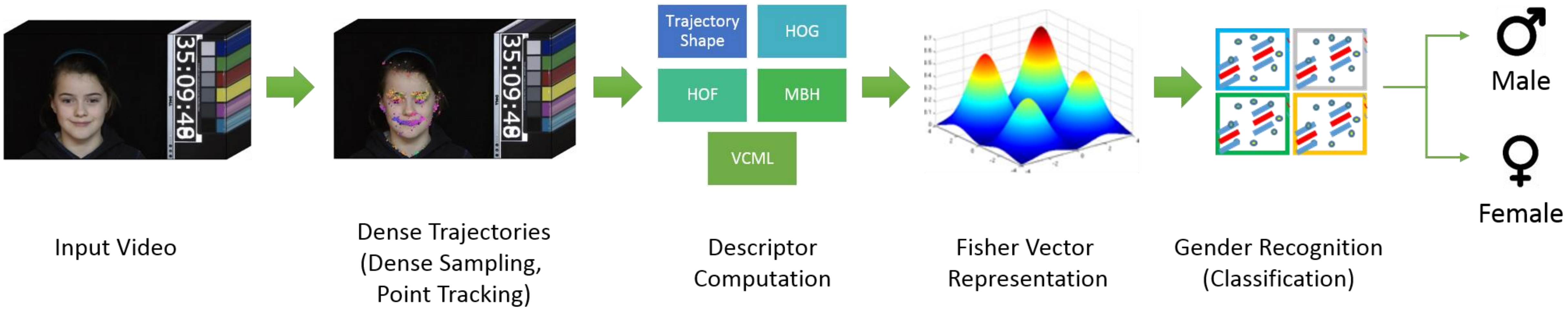
SCOPE OF THE WORK



- ➢ 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].



[8] Wang, J.; Li, J.; Yau, W.; Sung, E.: Boosting dense SIFT descriptors and shape contexts of face images for gender recognition. CVPRW, 2010.
[9] Perronnin, F.; Sanchez, J.; Mensink, T.: Improving the Fisher Kernel for large-scale image classification. ECCV, 2010.

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

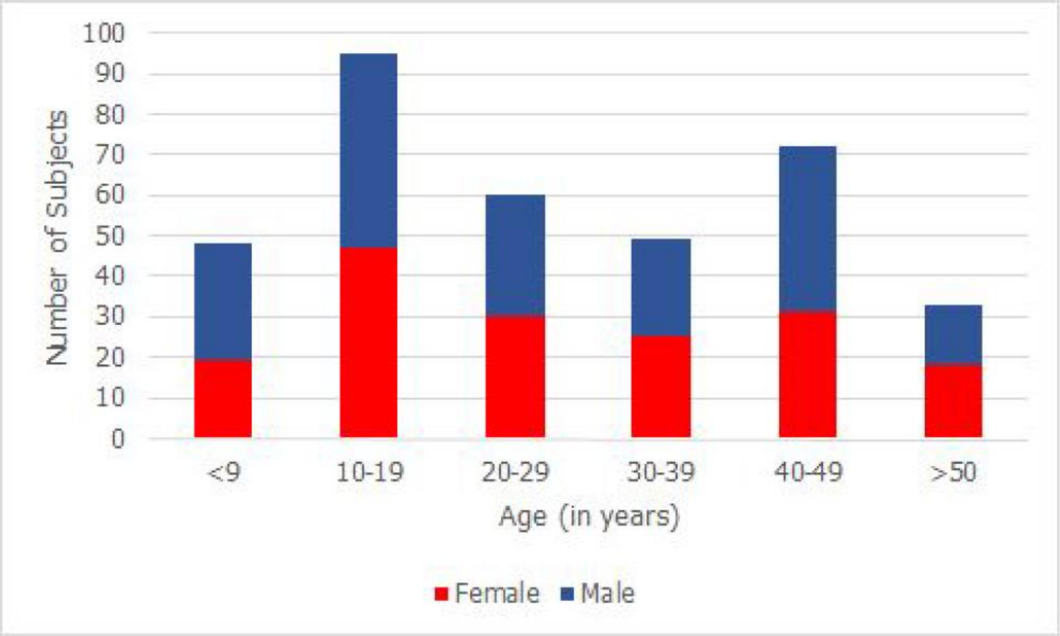
1-2 video sequences of 400 subjects

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

[10] Dibeklioglu, H.; Salah, A. A.; Gevers, T.: Are you really smiling at me? Spontaneous versus posed enjoyment smiles. ECCV, 2012.

[11] Dibeklioglu, H.; Gevers, T.; Salah, A. A.; Valenti, R.: A smile can reveal your age: Enabling facial dynamics in age estimation. ACM Multimedia, 2012.

BASELINE ALGORITHMS

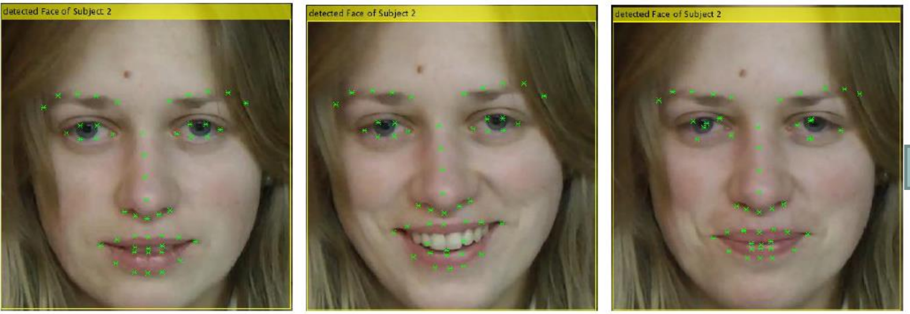
- OpenBR [12]: Publicly available toolkit for biometric recognition
  - Algorithm: Local binary pattern (LBP) - histograms -> scale-invariant feature transform (SIFT) - features on a dense grid of patches -> PCA projection -> Support Vector Machine (SVM) for classification
- how-old.net: underlying algorithm and the training dataset: not known

---

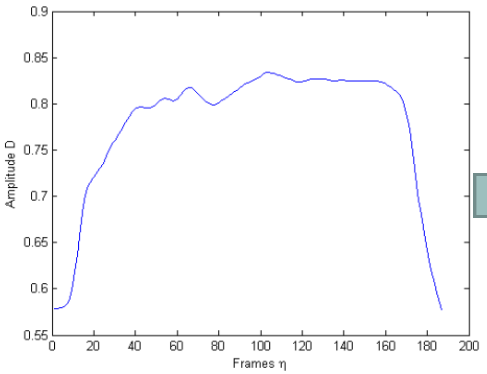
- Dynamics algorithm based on facial landmarks [13]
- Commercial Off-The-Shelf (COTS): underlying algorithm and training dataset: not known

[12] Klontz, J. C.; Klare, B. F.; Klum, S.; Jain, A. K.; Burge, M. J.: Open source biometric recognition. In: IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS). pp. 1–8, 2013.
[13] Dantcheva, A.; Bremond, F.: Gender estimation based on smile-dynamics. To appear in IEEE Transactions on Information Forensics and Security (TIFS), 2016.

DYNAMICS BASED ON FACIAL LANDMARKS



Facial landmark detection



Signal displacement of facial landmarks

Feature	Definition			
	General	Onset	Apex	Offset
Duration		$\frac{\eta(D^+)}{\omega}$	$\frac{\eta(D^a)}{\omega}$	$\frac{\eta(D^-)}{\omega}$
Duration Ratio		$\frac{\eta(D^+)}{\eta(D)}$		$\frac{\eta(D^-)}{\eta(D)}$
Maximal Amplitude	$max(D)$			
STD of Amplitude	$std(D)$			
Mean Amplitude		$mean(D^+)$	$mean(D^a)$	$mean(D^-)$
Total Amplitude		$\sum(D^+)$		$\sum(\lvert D^-\rvert)$
Net Amplitude	$\sum(D^+) - \sum(\lvert D^-\rvert)$			
Amplitude Ratio		$\frac{\sum(D^+)}{\sum(D^+)+\sum(D^-)}$		$\frac{\sum(D^-)}{\sum(D^+)+\sum(D^-)}$
Maximal Speed		$max(V^+)$		$max(V^-)$
Mean Speed		$mean(V^+)$		$mean(V^-)$
Maximum Acceleration		$max(A^+)$		$max(A^-)$
Mean Acceleration		$mean(A^+)$		$mean(A^-)$
Net Ampl., Duration Ratio	$\frac{(\sum(D^+)-\sum(\lvert D^-\rvert))\omega}{\eta(D)}$			

Statistics of signal displacement

Feature Selection
Min-Redundancy
Max-Relevance (mRMR)

Classification

Male

Female

RESULTS 1: TRUE GENDER CLASSIFICATION RATES

Age (Subject amount)	≤20 (148)	>20 (209)
OpenBR	52.3%	75.6%
how-old.net	55.5%	92%
COTS	76.9%	92.5%
Dynamics based on facial landmarks	59.4%	67.8%
COTS + Dynamics based on facial landmarks	76.9%	**93%**
Motion-based descriptors	77.7%	80.1%
Proposed Method	**86.3%**	91%

RESULTS 2: 5-CROSS-VALIDATION

True gender classification rates
for subjects ≤ 20 years old

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

True gender classification rates
for subjects > 20 years old

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

POSED SMILE (DYNAMICS BASED ON FACIAL LANDMARKS)

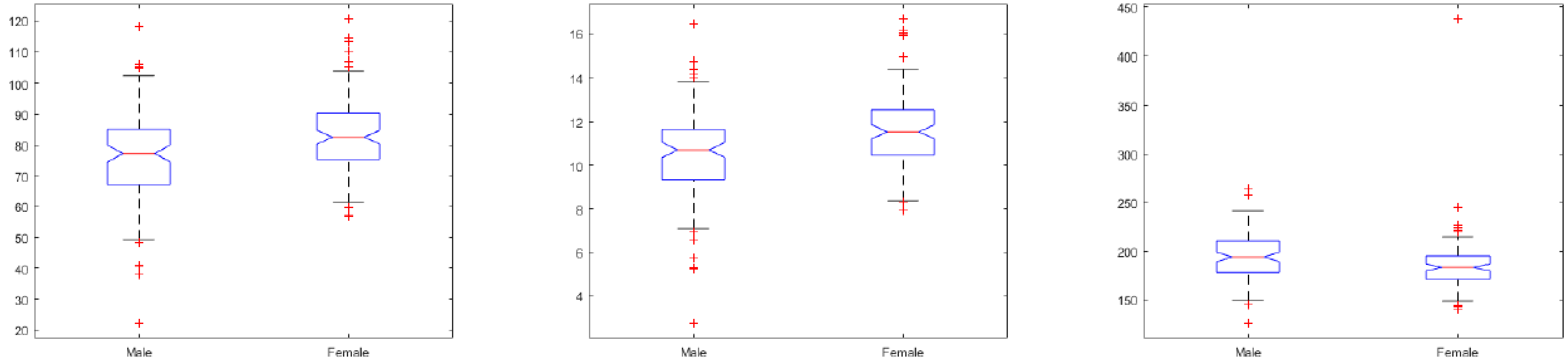
Age (Subject amount)	≤20 (143)	>20 (225)
how-old.net	51.1%	93.8%
COTS	76.9%	92%
Dynamics based on facial landmarks	59.4%	66.2%
COTS + Dynamics based on facial landmarks	76.9%	**92. 9%**

[13] Dantcheva, A.; Bremond, F.: Gender estimation based on smile-dynamics. To appear in IEEE Transactions on Information Forensics and Security (TIFS), 2016.

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

[13] Dantcheva, A.; Bremond, F.: Gender estimation based on smile-dynamics. To appear in IEEE TIFS, 2016.

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!

MOTION BOUNDARY HISTOGRAMS

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.

VIDEO COVARIANCE MATRIX LOGARITHM

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.