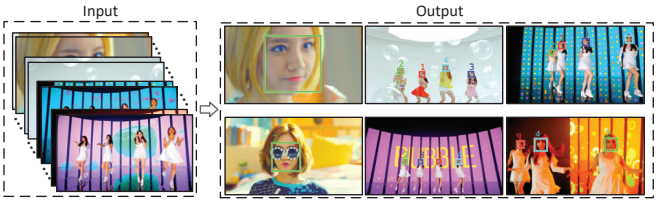
Tracking Persons-of-Interest via Adaptive Discriminative Features

Shun Zhang[1] Yihong Gong[1] Jia-Bin Huang[2] Jongwoo Lim[3] Jinjun Wang[1] Narendra Ahuja[2] Ming-Hsuan Yang[4]

[1]Xi'an Jiaotong University [2]University of Illinois, Urbana-Champaign [3]Hanyang University [4]University of California, Merced

Code and data available at http://bit.ly/multi-face-tracking-eccv2016

Problem

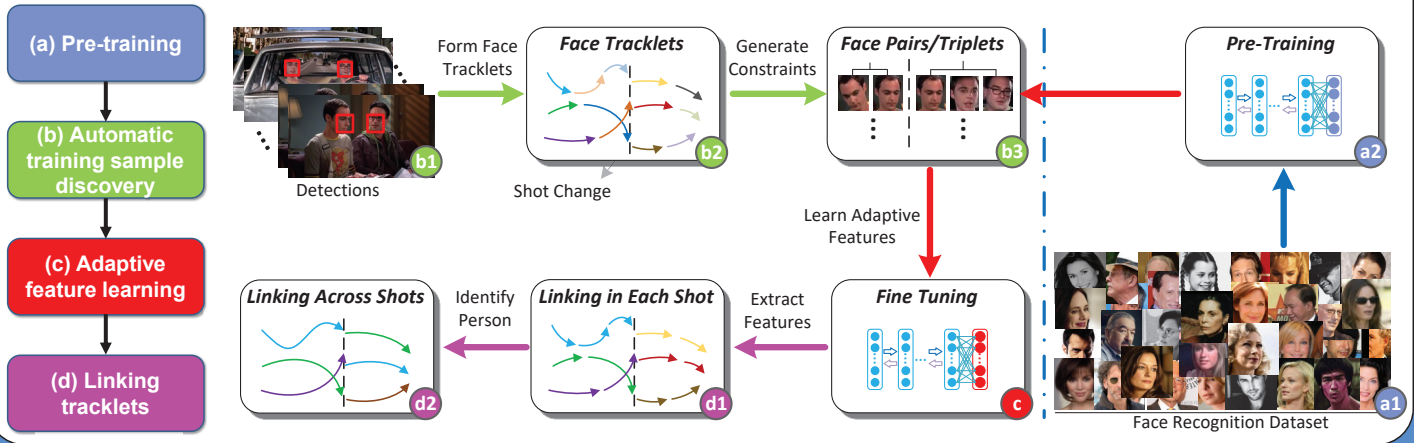


- Frequent shot changes
- Large face appearance variations
- Low resolution, motion blurring, occlusion and so on

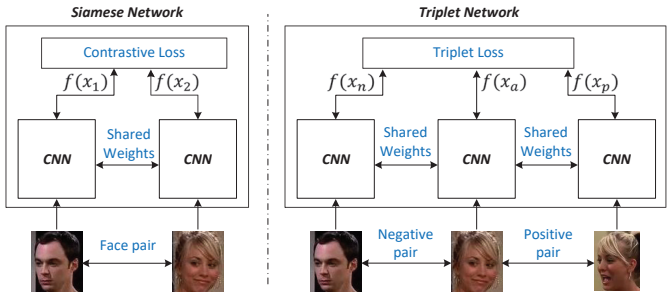
Contributions

- Learning video-specific features by adapting deep CNN based on contrastive and triplet loss
- An improved symmetric triplet loss function (SymTriplet)
- A fully automatic multi-face tracking algorithm (detection, tracking, clustering, and feature adaptation)
- A new dataset with 8 music videos from YouTube

Overview



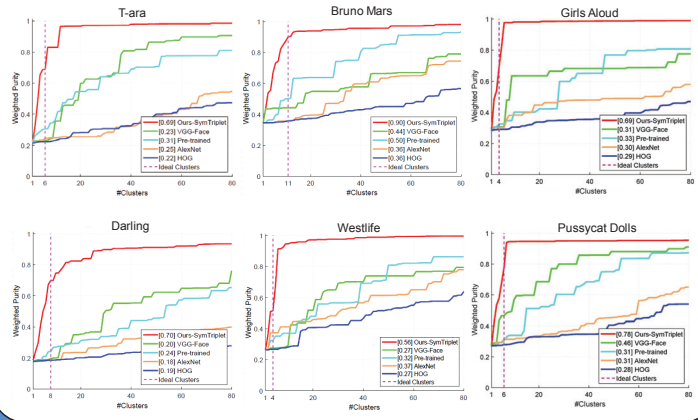
Adaptive Feature Learning



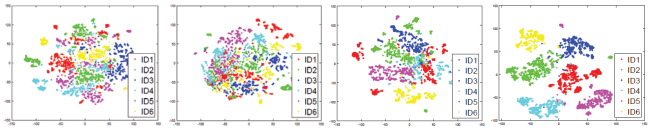
- SymTriplet loss:

$$L_s = \max\left\{0, d(x_a, x_p) - \frac{1}{2}\left(d(x_a, x_n) + d(x_p, x_n)\right) + \alpha\right\}$$

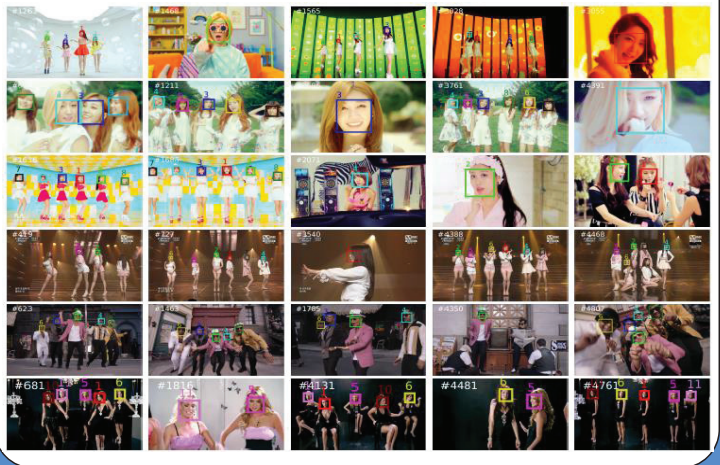
Weighted Purity vs. #Clusters



t-SNE visualization



Qualitative Results



Quantitative Evaluation

Music video dataset								
Method	Recall↑	Precision↑	F1↑	FAF↓	IDS↓	Frag↓	MOTA↑	MOTP↑
mTLD [Kalal et al. 2012]	9.7%	36.1%	15.3%	0.39	280	621	-7.7%	68.4%
ADMM [Ayazoglu et al. 2012]	75.5%	61.8%	68.0%	0.50	2382	2959	51.7%	63.7%
IHTLS [Dicle et al. 2013]	75.5%	68.0%	71.6%	0.41	2013	2880	56.2%	63.7%
Pre-trained	60.1%	88.8%	71.7%	0.17	931	2140	51.5%	79.5%
Ours-Siamese	71.5%	89.4%	79.5%	0.19	986	2512	62.3%	64.0%
Ours-Triplet	71.8%	88.8%	79.4%	0.20	902	2546	61.8%	64.2%
Ours-SymTriplet	71.8%	89.7%	79.8%	0.19	699	2563	62.8%	64.3%