Probabilistic Elastic Part Model for Face Processing

Haoxiang Li

Department of Computer Science
Stevens Institute of Technology

May 23, 2016

Outline

- **Introduction**
- Probabilistic elastic part (PEP) model
- Eigen-PEP
- Hierarchical-PEP
- Applications
  - Face Recognition
  - Face Detection
  - Face Tracking
- Conclusion

Face Processing - Face Recognition

?

Database

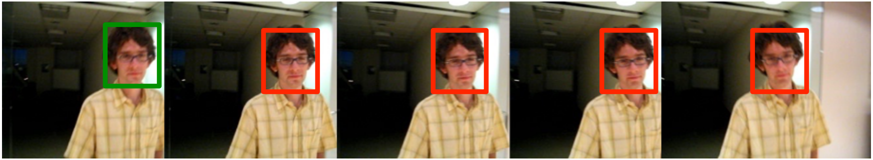
Trump

Clinton

Obama

Face Processing - Face Detection

Face Processing - Face Tracking



Face Processing - Existing Methods

Many methods were proposed for these tasks.

- Face Recognition
  [Turk, CVPR'91][Wolf, ECCV'08][Chen, ECCV'12]
  [Simonyan, BMVC'13][Hu, CVPR'14][Zhu, CVPR'15]
- Face Detection
  [Viola, IJCV'04][Zhu, CVPR'12][Shen, CVPR'13]
  [Li, CVPR'14][Chen, ECCV'14][Li, CVPR'15]
- Face Tracking
  [Wu, ICCV'01][Yu, CVPR'06][Kim, CVPR'08]
  [Zhou, CVPR'10][Zhang, CVPR'14][Liu, CVPR'15]

Motivation

Can we approach to these tasks with an unified model?

Face Processing - Face Recognition

Face identification

?

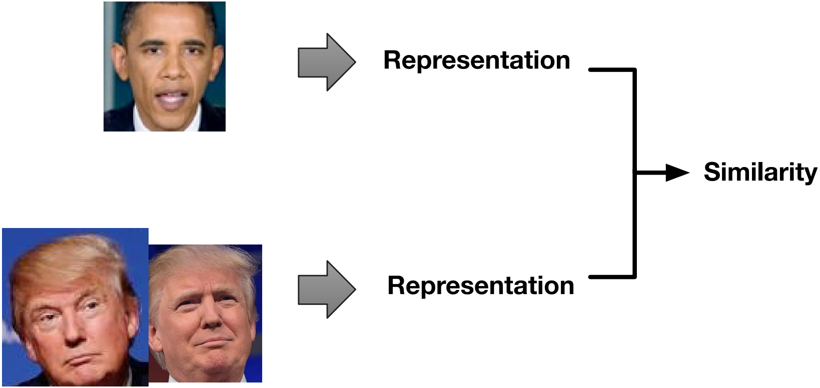
Database

Face verification

Same person ?

Same person ?

Face Processing - Face Recognition

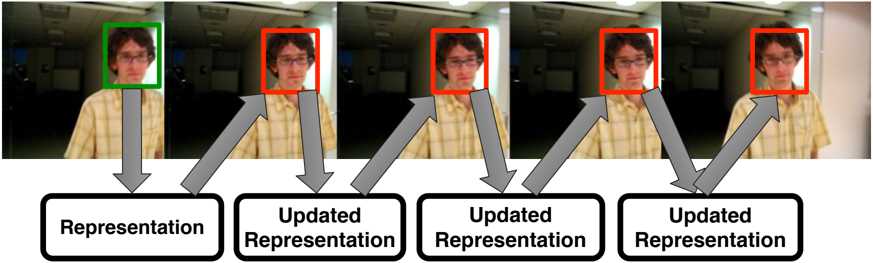


Face Processing - Face Detection

Background

Faces

Face Processing - Face Tracking



Motivation

We need an unified face representation/model for face image or image set.
The representation/model should

- contain identity information.
- be discriminative for face and background classification.
- be able to update incrementally.

Probabilistic Elastic Part (PEP) Model

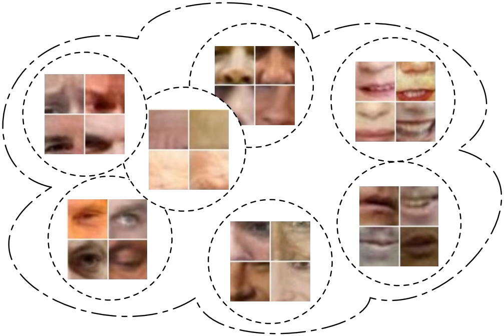
- PEP-model is a generative model for face.
- PEP-model constructs unified representation for face image and face video.
- PEP-model adapts to new data.
- The representation can be updated incrementally.

Outline

- Introduction
- **Probabilistic elastic part (PEP) model**
- Eigen-PEP
- Hierarchical-PEP
- Applications
  - Face Recognition
  - Face Detection
  - Face Tracking
- Conclusion

Probabilistic Elastic Part (PEP) Model

PEP-model is a mixture of part models. Each part model here is a generative model of a face part (appearance and location).

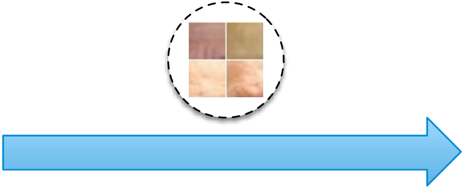


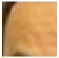
With this mixture model, we can build face representation.

Construct Representation

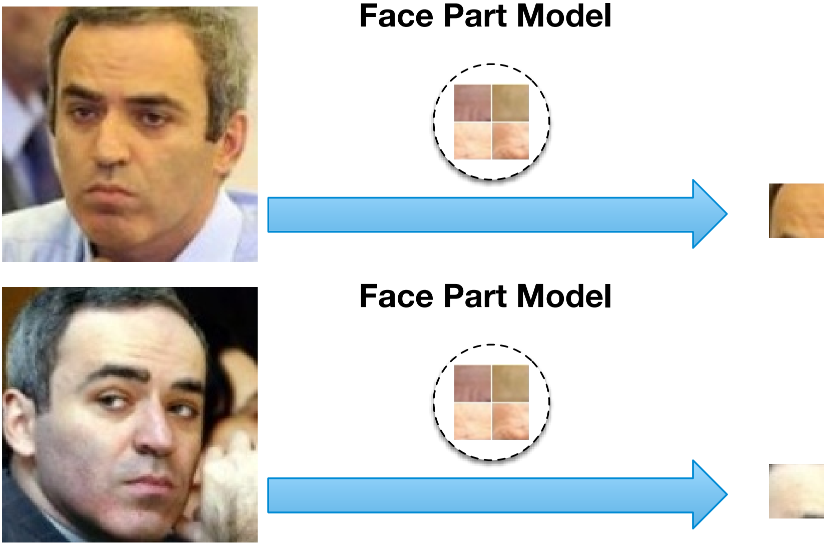
Construct Representation

Face Part Model

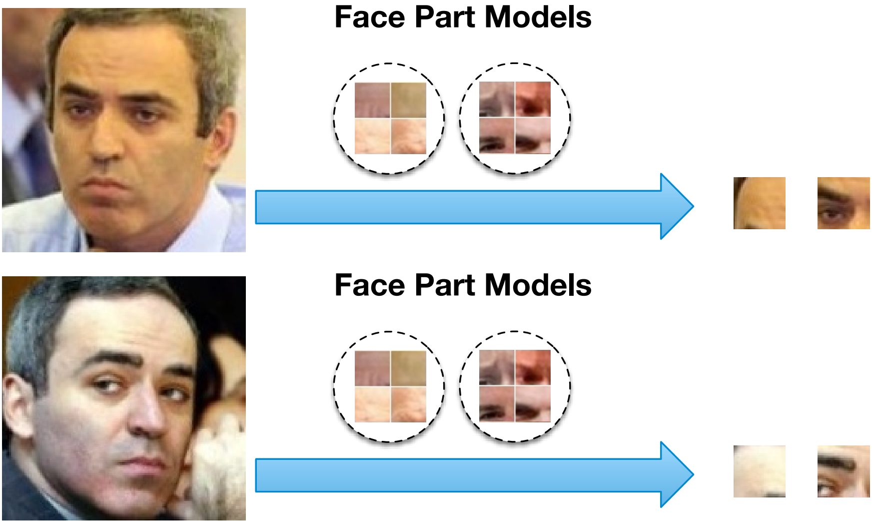




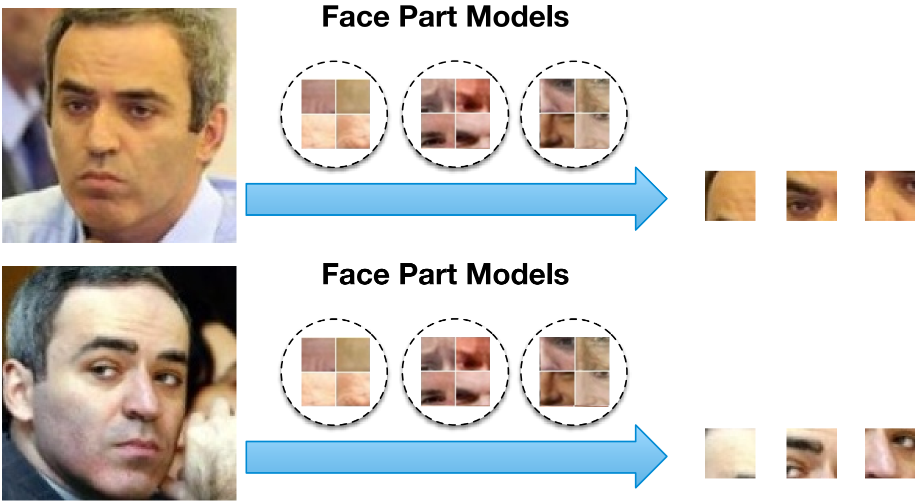
Construct Representation



Construct Representation



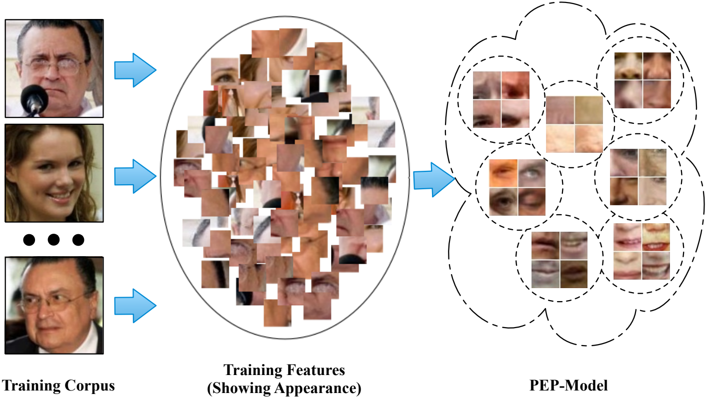
Construct Representation



PEP-representation

PEP-model can construct PEP-representation for face image/video.

Definition of PEP-model



Use Gaussian model as the part model. PEP-model is a Gaussian Mixture Model.

$$P(\mathbf{f}|\Theta) = \sum_{k=1}^{K} \omega_k \mathcal{G}(\mathbf{f}|\vec{\mu}_k, \sigma_k^2 \mathbf{I}),$$

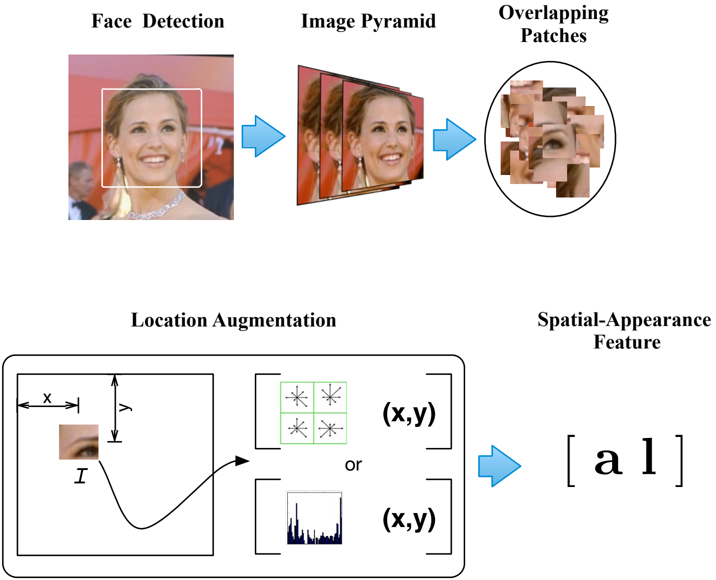
Learning the PEP-model

Expectation-Maximization (EM) algorithm obtains an estimate of GMM parameters Θ

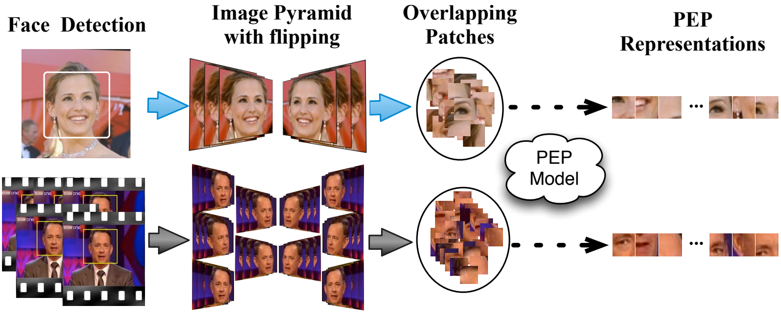
$$\Theta^* = \arg\max_{\Theta} \mathcal{L}(\chi|\Theta)$$

- $\Theta = (\omega_1, \vec{\mu}_1, \sigma_1, \ldots, \omega_K, \vec{\mu}_K, \sigma_K)$;
- K is the number of Gaussian mixture components;
- $\mathbf{I}$ is an identity matrix;
- ω_k is the mixture weight of the k-th Gaussian component;
- $\mathcal{G}(\mu_k, \sigma_k^2 \mathbf{I})$ is a spherical Gaussian (mean μ_k, variance $\sigma_k^2 \mathbf{I}$)
- $\mathbf{f}$ is image patch feature descriptor.

Image Patch Feature Descriptor

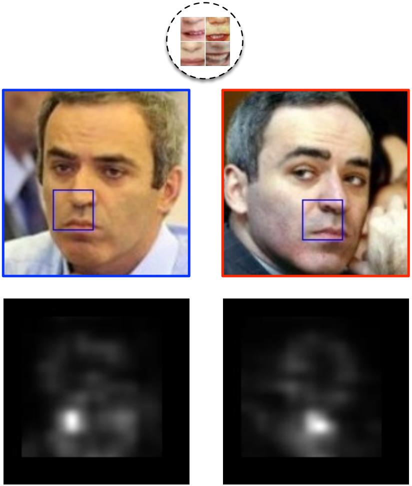


Unified Representation



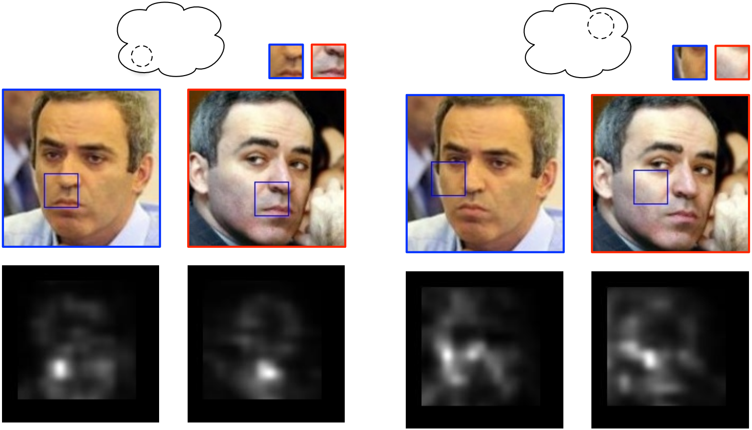
Face Part Model Selects Image Patches

A face part model is a Gaussian model.



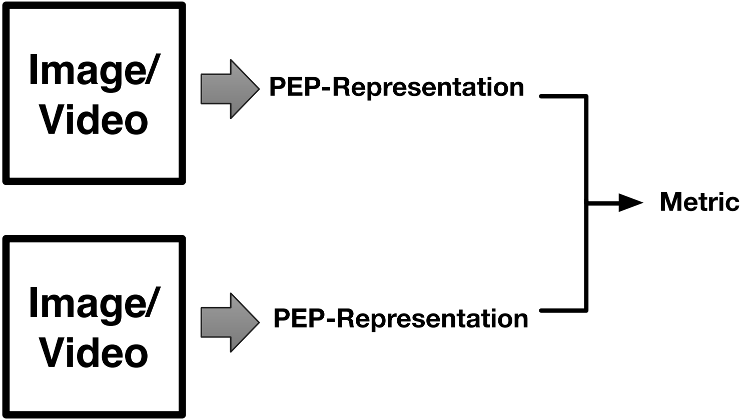
The Probabilistic Elastic Part-based Representation

- The Gaussian component is $(\omega_k, \mathcal{G}_k(\vec{\mu}_k, \sigma_k^2\mathcal{I}))$
- The face is firstly represented as a bag of descriptors $\mathbf{f}_\mathcal{F} = \{\mathbf{f}_i\}$
- The k-th Gaussian component select one descriptor $\mathbf{f}_{g_k(\mathcal{F})}$ from $\mathbf{f}_\mathcal{F}$, such that $g_k(\mathcal{F}) = \arg\max_i \omega_k \mathcal{G}(\mathbf{f}_i | \vec{\mu}_k, \sigma_k^2 \mathbf{I})$.

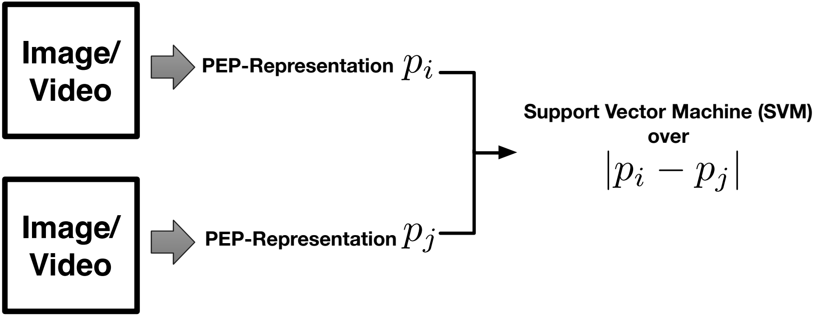


Evaluation of Similarity

One of the key problems we want to address is how to evaluate the similarity of two faces.



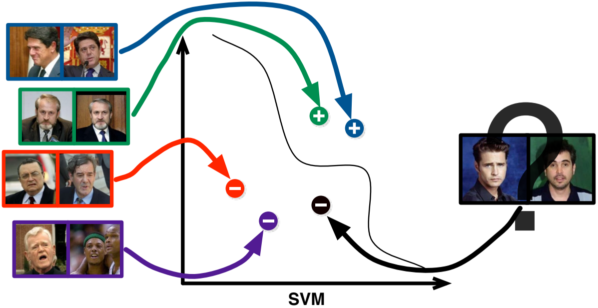
Probabilistic Elastic Matching



Haoxiang Li, Gang Hua, Zhe Lin, Jonathan Brandt, Jianchao Yang,
Probabilistic Elastic Matching for Pose Variant Face Verification, CVPR 2013

Classification Using SVM

- C pairs of faces are transformed into C difference vectors, $\{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_C\}$



- A kernel SVM classifier, i.e.,

$$f(\mathbf{d}) = \sum_{i=1}^{\mathcal{V}} \alpha_i k(\mathbf{d}_i, \mathbf{d}) + b,$$

is trained over training difference vectors with Gaussian Radial Basis Function (RBF) kernel

Limitations

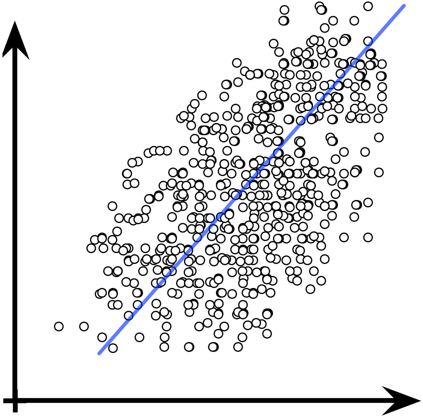
- High-dimensionality: a PEP-model with 1024 components using SIFT local descriptor generates $1024x128 = 131072$-d PEP-representations.
- Slow in testing: Kernel SVM is not efficient.

Outline

- Introduction
- Probabilistic elastic part (PEP) model
- **Eigen-PEP**
- Hierarchical-PEP
- Applications
  - Face Recognition
  - Face Detection
  - Face Tracking
- Conclusion

Eigen-PEP

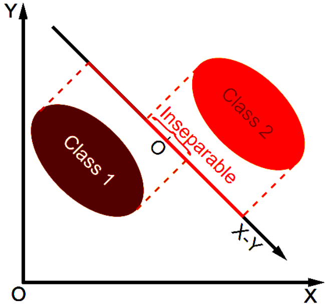
Reduce the dimensionality of PEP-representations by Principal Component Analysis (PCA) from $131,072$ to 100.



Joint Bayesian Classifier

$$r(x_1, x_2) = log \frac{P(x_1, x_2|H_I)}{P(x_1, x_2|H_E)}, \qquad X = [x_1 \quad x_2]$$

$$P(X|H_I) \sim \mathcal{N}(0, \Sigma_I), \qquad P(X|H_E) \sim \mathcal{N}(0, \Sigma_E)$$



[Chen *et al.* ECCV 2012]

Linear Discriminant Embedding (LDE)

Seek for a discriminative projection w

$$\bar{\mathbf{w}} = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{A} \mathbf{w}}{\mathbf{w}^T \mathbf{B} \mathbf{w}},$$

where

$$\begin{aligned} \mathbf{A} &= \sum_{l_{ij}=0} (\mathbf{p}_i - \mathbf{p}_j)(\mathbf{p}_i - \mathbf{p}_j)^T, \\ \mathbf{B} &= \sum_{l_{ij}=1} (\mathbf{p}_i - \mathbf{p}_j)(\mathbf{p}_i - \mathbf{p}_j)^T \end{aligned}$$

$\mathbf{p}_{i,j}$ are low-dimensional face representations.

Eigen-PEP Representation

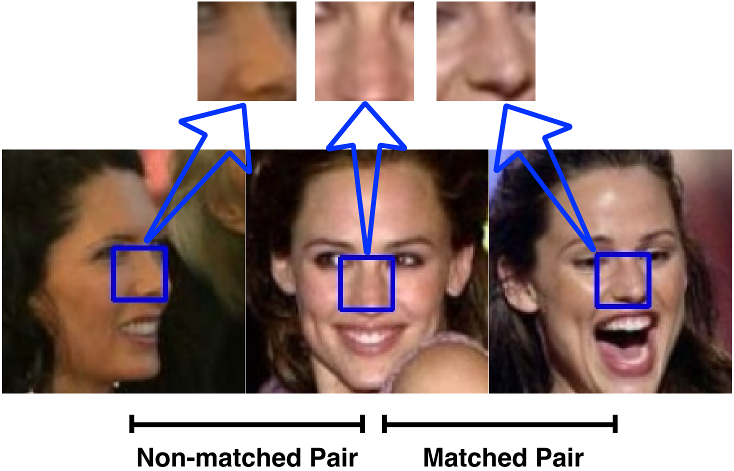
- Compact: Typical dimensionality of Eigen-PEP representation is 100.
- Discriminative: Performance is largely improved with Joint Bayesian Classifier or LDE.

Haoxiang Li, Gang Hua, Xiaohui Shen, Zhe Lin, Jonathan Brandt,
Eigen-PEP for Video Face Recognition, ACCV 2014

Outline

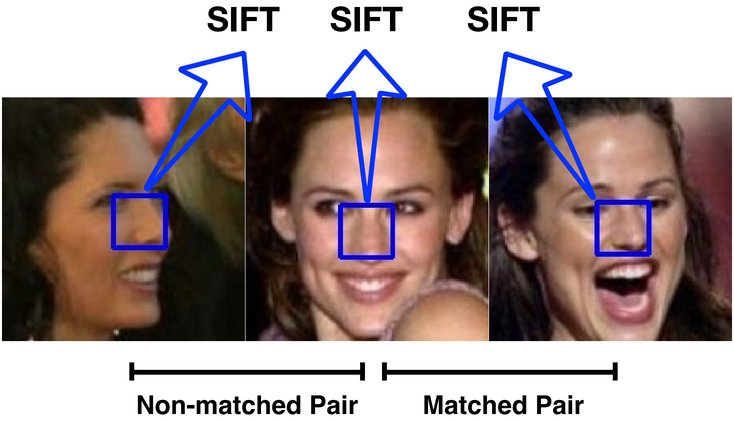
- Introduction
- Probabilistic elastic part (PEP) model
- Eigen-PEP
- **Hierarchical-PEP**
- Applications
  - Face Recognition
  - Face Detection
  - Face Tracking
- Conclusion

Local Variation



Local Variation

A local descriptor is not robust enough.



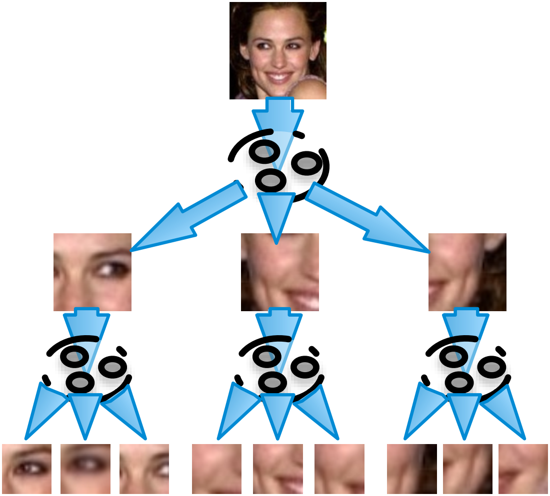
We need a better part representation.

PEP Model

PEP-model learned from face images can construct robust face representation.

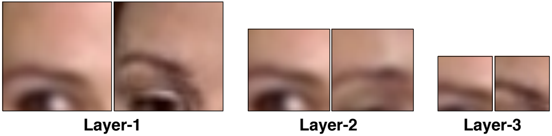
We can build PEP-representation for face part.

Hierarchically Apply PEP Model



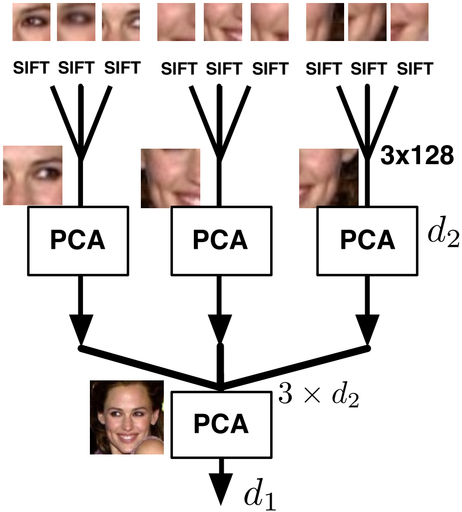
Hierarchically Apply PEP Model

The matches are better at lower layers.



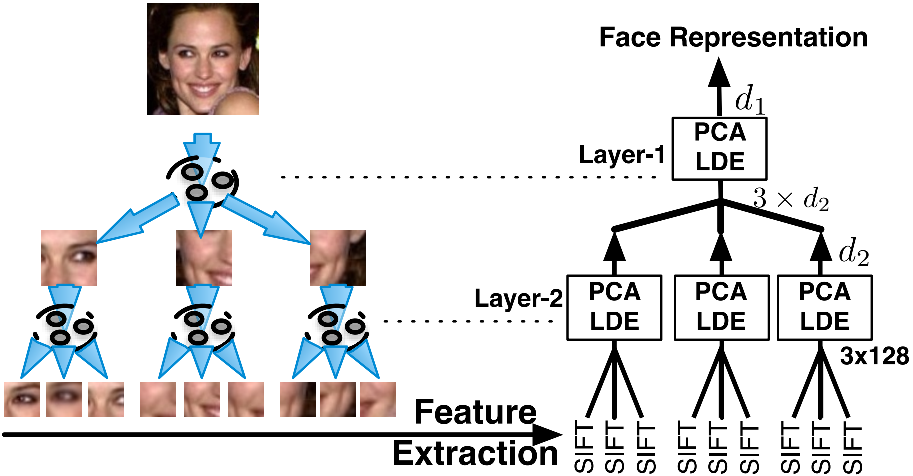
Face Representation

How to construct the face representation?

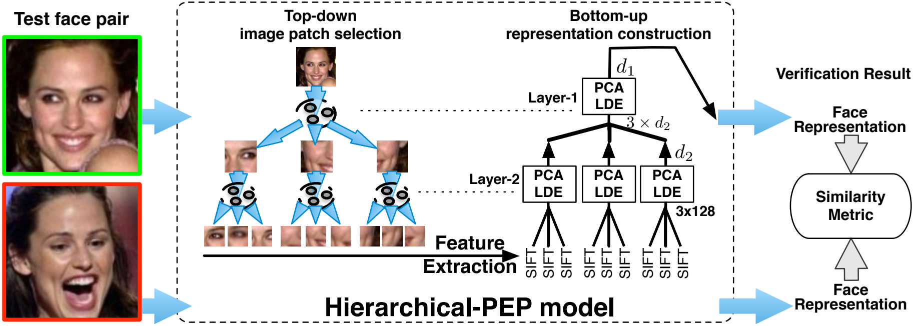


Supervised Information

LDE gives a discriminative projection.



Hierarchical-PEP Model



Haoxiang Li, Gang Hua,
Hierarchical-PEP Model for Real-world Face Recognition, CVPR 2015

Outline

- Introduction
- Probabilistic elastic part (PEP) model
- Eigen-PEP
- Hierarchical-PEP
- **Applications**
  - **Face Recognition**
  - Face Detection
  - Face Tracking
- Conclusion

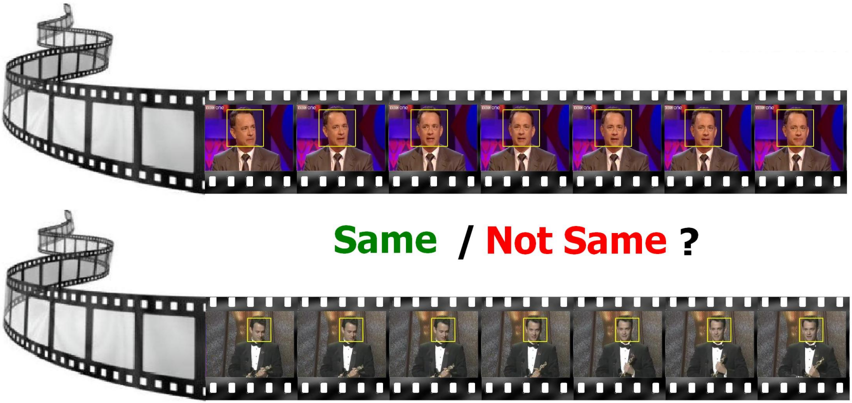
Applications - Face Recognition

Labeled Faces in the Wild (13,000 images from 5,749 people)
`http://vis-www.cs.umass.edu/lfw/`

Applications - Face Recognition

YouTube Faces Dataset (3,425 videos from 1,595 people)

`http://www.cs.tau.ac.il/~wolf/ytfaces/`



Applications - Face Recognition

Celebrity-1000 Dataset (159,726 video sequences from 1,000 people)
`http://www.lv-nus.org/facedb/`

Labeled Faces in the Wild (LFW)

Same person ?

Same person ?

LFW - Training Protocols:

- Unrestricted protocol: "personal name of each training image"
- Restricted protocol: "matched or non-matched"
  - least restricted protocol: outside training data for face alignment, feature extraction or building recognition model
  - less restricted protocol: outside training data for face alignment or feature extraction
  - most restricted protocol: without outside training data

Labeled Faces in the Wild (LFW)

LFW - Settings:

- crop image at center to 150x150
- down-sample image by factor 0.9 into 3-scales
- SIFT descriptor is extracted from a 8x8 sliding window with 2-pixel spacing
- PEP-model is of 1024 Gaussian components
- images are roughly aligned by funneling algorithm
- the dimensionality of Eigen-PEP is set to 100
- the Hierarchical-PEP has 3-layers for parts in 32x32, 24x24, and 16x16; number of components are 256, 4, and 4 respectively; $d_1 = 200$, $d_2 = 100$, $d_3 = 50$.

Labeled Faces in the Wild (LFW)

Table: Performance comparison on the most restricted LFW

Algorithm	Accuracy $\pm$ Error(%)
Hybrid descriptor-based[1]	78.47 ± 0.51
V1/MKL[2]	79.35 ± 0.55
MRF-MLBP[3]	79.08 ± 0.14
Fisher vector faces[4]	87.47 ± 1.49
Spartans[5]	87.55 ± 0.21
PEM (PEP + SVM)	81.38 ± 0.98
Eigen-PEP	88.97 ± 1.32
Hierarchical-PEP	91.10 ± 1.47

1. Lior Wolf et al. Descriptor Based Methods in the Wild, ECCV 2008
2. Nicolas Pinto et al. How far can you get with a modern face recognition test set using only simple features, CVPR 2009
3. Shervin Rahimzadeh Arashloo et al. Efficient Processing of MRFs for Unconstrained-Pose Face Recognition, Biometrics 2013
4. Karen Simonyan et al. Fisher Vector Faces in the Wild, BMVC 2013
5. Felix Juefei-Xu et al. Spartans: Single-Sample Periocular-Based Alignment-Robust Recognition Technique Applied to Non-Frontal Scenarios, TIP 2015

Labeled Faces in the Wild (LFW)

Yi et al. published a large-scale dataset CASIA-Webface contains 494,414 images of 10,575 subjects.

Table: Performance with outside training data.

Algorithm	Accuracy $\pm$ Error(%)
Eigen-PEP (LFW)	88.97 ± 1.32
Eigen-PEP (Webface)	93.77 ± 0.89

YouTube Faces (YTF)

Same person ?

YouTube Faces - Settings:

- crop image at center to 100x100
- down-sample image by factor 0.9 into 3-scales
- SIFT descriptor is extracted from a 8x8 sliding window with 2-pixel spacing
- PEP-model is of 1024 Gaussian components
- the dimensionality of Eigen-PEP is set to 100
- the Hierarchical-PEP has 2-layers for parts in 32x32 and 24x24; number of components are 256 and 16; $d_1 = 400$ and $d_2 = 200$.

Experimental Results

Table: Performance comparison over YouTube Faces

Algorithm	Accuracy $\pm$ Error(%)
MBGS[1]	76.4 ± 1.8
STFRD+PMML[2]	79.5 ± 2.5
VF^2[3]	84.7 ± 1.4
DDML (combined)[4]	82.3 ± 1.5
PEM (PEP + SVM)	77.52 ± 2.06
Eigen-PEP	84.80 ± 1.4
Hierarchical-PEP	87.00 ± 1.5

1. Lior Wolf et al. Face Recognition in Unconstrained Videos with Matched Background Similarity, CVPR 2011
2. Zhen Cui et al. Fusing Robust Face Region Descriptors via Multiple Metric Learning for Face Recognition in the Wild, CVPR 2013
3. Omkar M. Parkhi et al. A Compact and Discriminative Face Track Descriptor, CVPR 2014
4. Junlin Hu et al. Discriminative Deep Metric Learning for Face Verification in the Wild, CVPR 2014

Experimental Results

Table: Performance comparison over YouTube Faces

Algorithm	Accuracy $\pm$ Error(%)
Eigen-PEP (YTF)	84.80 ± 1.4
Eigen-PEP (Webface)	88.84 ± 1.36

Celebrity-1000 Dataset

Celebrity-1000 - Settings:

- images are of size 80x64
- down-sample image by factor 0.9 into 3-scales
- SIFT descriptor is extracted from a 8x8 sliding window with 2-pixel spacing
- PEP-model is of 200 Gaussian components
- For closed-set the dimensions of Eigen-PEP is from 100 to 400 (90% eigenvalues)
- For the open-set the dimensions of Eigen-PEP is set to 500

Celebrity-1000 Dataset

Table: Performance comparison on Celebrity-1000 dataset (closed-set): showing the rank-K accuracy.

Subjects		rank-1 (%)	rank-2 (%)	rank-5 (%)
100	Eigen-PEP	50.60	**59.76**	**68.92**
	MTJSR	50.60	55.78	66.53
200	Eigen-PEP	**45.02**	**52.49**	**65.33**
	MTJSR	40.80	48.47	55.56
500	Eigen-PEP	**39.97**	**48.21**	**57.85**
	MTJSR	35.46	40.05	46.35
1000	Eigen-PEP	**31.94**	**40.27**	**51.01**
	MTJSR	30.04	34.88	40.58

MTJSR: Luoqi Liu et al. Toward Large-Population Face Identification in Unconstrained Videos, CSVT 2014

Celebrity-1000 Dataset

Table: Performance comparison on Celebrity-1000 dataset (open-set): showing the rank-K accuracy.

Subjects		rank-1 (%)	rank-2 (%)	rank-5 (%)
100	Eigen-PEP	**51.55**	**61.63**	**68.22**
	MTJSR	46.12	55.04	62.02
200	Eigen-PEP	**46.15**	**55.03**	**66.07**
	MTJSR	39.84	46.55	54.64
400	Eigen-PEP	**42.33**	**49.57**	**61.23**
	MTJSR	37.51	42.91	48.41
800	Eigen-PEP	**35.90**	**44.27**	**54.60**
	MTJSR	33.50	37.71	42.41

MTJSR: Luoqi Liu et al. Toward Large-Population Face Identification in Unconstrained Videos, CSVT 2014

Outline

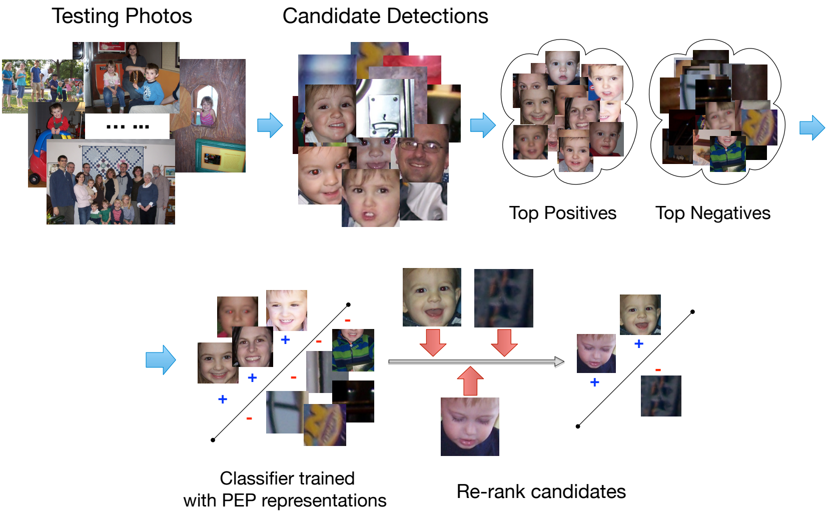
- Introduction
- Probabilistic elastic part (PEP) model
- Eigen-PEP
- Hierarchical-PEP
- **Applications**
  - Face Recognition
  - **Face Detection**
  - Face Tracking
- Conclusion

Applications - Face Detection

- Detector adaptation with PEP-model
- PEP face detector

Unsupervised Face Detector Adaptation

How to adapt a general face detector to a specific image collection?



Haoxiang Li, Gang Hua, Zhe Lin, Jonathan Brandt, Jianchao Yang,
Probabilistic Elastic Part Model for Unsupervised Face Detector Adaptation, ICCV 2013

Qualitative Results

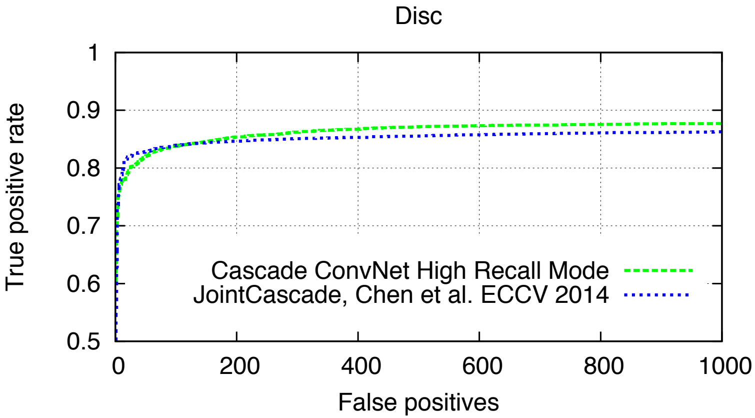
We evaluated this method on 3 datasets with 3 face detectors. The best detector in our experiments is the Convolutional Neural Networks (CNN) Cascade Detector.

Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, Gang Hua,
A Convolutional Neural Network Cascade for Face Detection, CVPR 2015

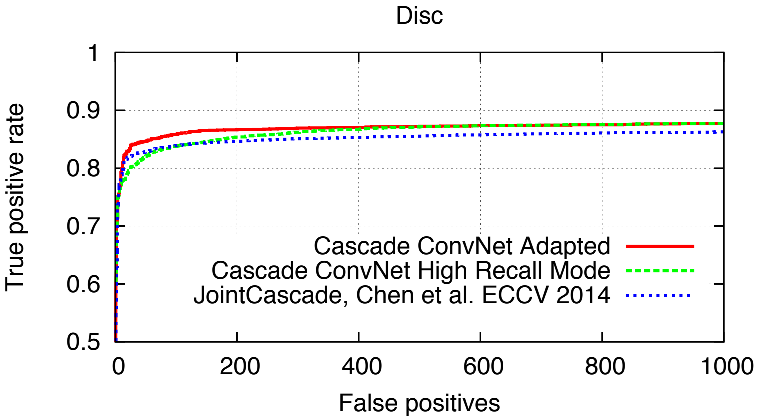
Face Detection Data Set and Benchmark (FDDB)

FDDB is published by Vidit Jain and Erik Learned-Miller in 2010.
It has 2,845 images with 5,171 faces.

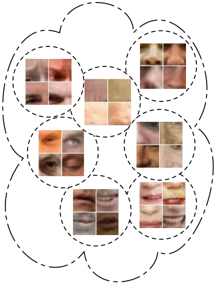
Face Detection Data Set and Benchmark (FDDB)



Face Detection Data Set and Benchmark (FDDB)



PEP Face Detector - Preliminary Result



PEP-model

testing image

PEP Face Detector - Preliminary Result

part model

face part casts votes with a Gaussian window

PEP Face Detector - Preliminary Result

part model

testing image

PEP Face Detector - Preliminary Result

part model

testing image

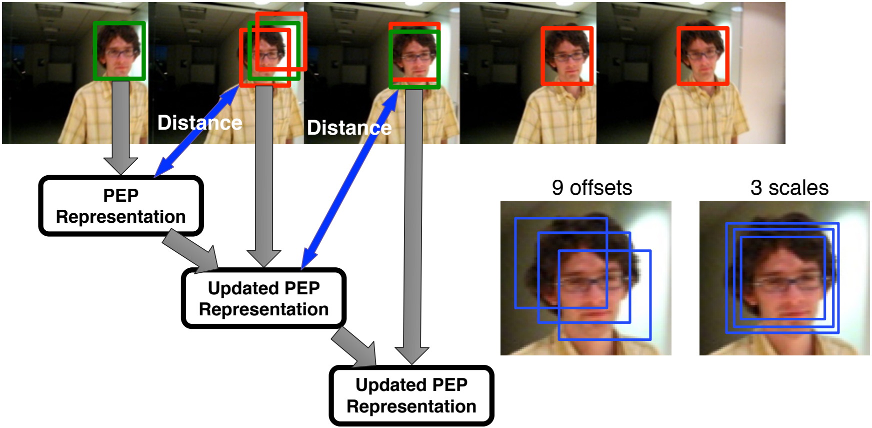
Outline

- Introduction
- Probabilistic elastic part (PEP) model
- Eigen-PEP
- Hierarchical-PEP
- **Applications**
  - Face Recognition
  - Face Detection
  - **Face Tracking**
- Conclusion

Applications - Face Tracking (Preliminary Result)

- Construct target PEP-representation from the initial face bounding box in frame 0
- Propogate and generate candidate bounding boxes in frame 1
- Construct PEP-representations for all candidates
- Select the nearest candidate to the target as the tracking result
- Update the target PEP-representation with the selected candidate PEP-representation

Applications - Face Tracking (Preliminary Result)



(See Demo video)

Applications - Robotic Wheelchair

We run Eigen-PEP for face recognition on the robotic wheelchair.
(See Demo video)

Conclusion

- We developed PEP-model as an unified framework for face processing.
- PEP-model constructs compact, flexible, and discriminative face representation.
- We demonstrated it can support accurate face recognition and detector adaptation.
- We presented preliminary results on face detection and tracking.

Relevant Publications

- Haoxiang Li, Jonathan Brandt, Zhe Lin, Xiaohui Shen, Gang Hua, *A Multi-Level Contextual Model For Person Recognition in Photo Albums*, CVPR 2016
- Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, Gang Hua, *A Convolutional Neural Network Cascade for Face Detection*, CVPR 2015
- Haoxiang Li, Gang Hua, *Hierarchical-PEP Model for Real-world Face Recognition*, CVPR 2015
- Haoxiang Li, Zhe Lin, Jonathan Brandt, Xiaohui Shen, Gang Hua, *Efficient Boosted Exemplar-based Face Detection*, CVPR 2014
- Haoxiang Li, Gang Hua, Xiaohui Shen, Zhe Lin, Jonathan Brandt, *Eigen-PEP for Video Face Recognition*, ACCV 2014
- Haoxiang Li, Gang Hua, Zhe Lin, Jonathan Brandt, Jianchao Yang, *Probabilistic Elastic Part Model for Unsupervised Face Detector Adaptation*, ICCV 2013
- Haoxiang Li, Gang Hua, Zhe Lin, Jonathan Brandt, Jianchao Yang, *Probabilistic Elastic Matching for Pose Variant Face Verification*, CVPR 2013

Thank You

Advisor

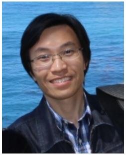
Gang Hua

Co-authors

Jon Brandt

Zhe Lin

Xiaohui Shen



Jianchao Yang

Thank the committee and everyone presented!