FACE RECOGNITION BASED ON IMAGE SETS

Hakan Cevikalp
Eskisehir Osmangazi Unıversity
Electrical and Electronics Engineering
Eskisehir, Turkey

Bill Triggs
Laboratoire Jean Kuntzmann
Grenoble, France

OUTLINE

- Introduction and overview of the related methods
- Proposed Method
  - Affine and convex models for image sets
  - Finding distances between models
  - Extension to the nonlinear case (kernelization)
- Experiments
- Conclusion

FACE RECOGNITION BASED ON IMAGE SETS

- Each query is a set of images (multiple images) of an unknown individual, not a single image.

- The gallery contains a set of images for each known individual.
- The goal is to recover the individual whose gallery set is the best match for the given query set.
- The query and gallery sets may contain large variations in pose, illumination and scale.

FACE RECOGNITION BASED ON IMAGE SETS

Traditional face recognition methods using single images:

- Training images are taken under well-controlled conditions
- queries often are too, e.g., subject must stand in front of the camera
- System often fails in more challenging environments.

Image Sets:

- naturally incorporate information about the variability of the subject's appearance
- are the most natural form of the input (e.g., surveillance systems incorporating tracking).

RELATED METHODS (1)

Methods for recognition based on image sets have two key elements:

- the model used for approximating image sets,
- the distance metric used to evaluate similarity between models.

Related Methods:

- (Fitzgibbon & Zisserman, CVPR 2003) - Joint Manifold Distance metric over linear affine subspace representation.
- (Arandjelovic et al., CVPR 2005) - Kullback-Leibler divergence metric over parametric distribution representation.
- Mutual Subspace Method (MSM) (Yamaguchi et al., Int. Syp. Robotics Res. 1998) - Canonical Angles distance metric over linear subspace representation.

RELATED METHODS (2)

- Manifold Learning Methods
  - (A. Hadid and M. Pietikainen, Int. Conf. Aut. Face & Ges. Rec. 2004) - locally linear embedding + k-means clustering + cluster centers
  - (W. Fan & D. Y. Yeung, CVPR 2006) - hierarchical clustering + linear subspace representation
  - (Wang et al. CVPR 2008) - nearest neighbor clustering + linear subspace representation + cluster centers

We used spectral clustering in our experiments and tested affine subspace approximation as well.

PROPOSED METHOD – Convex Sets

Affine Hulls

- Each affine hull is the smallest affine (shifted linear) subspace containing affine combinations of a person's face descriptor vectors

$$H_c^{affine} = \left\{ \mathbf{x} \mid \sum_{k=1}^{n_c} \alpha_{ck} \mathbf{x}_{ck} \mid \sum_k \alpha_{ck} = 1 \right\},$$

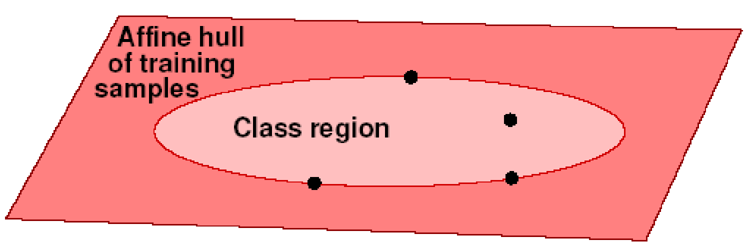
- Affine hulls are unbounded, and hence typically rather loose models for image sets.

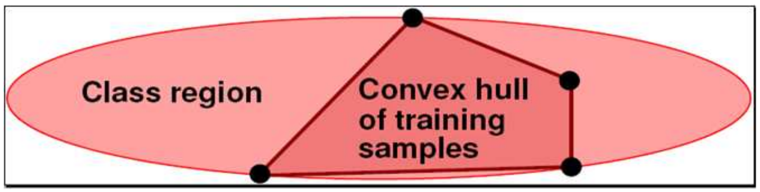
Convex Hulls

- Each convex hull is the smallest convex set containing convex combinations of a person's face descriptor vectors

$$H_c^{convex} = \left\{ \mathbf{x} \mid \sum_{k=1}^{n_c} \alpha_{ck} \mathbf{x}_{ck} \mid \sum_k \alpha_{ck} = 1, \alpha_{ck} \geq 0 \right\},$$

- Convex hulls are typically tight approximations for image sets.



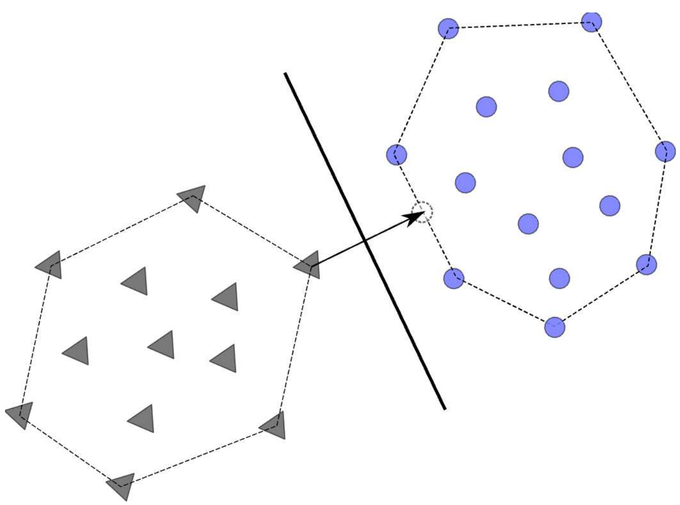


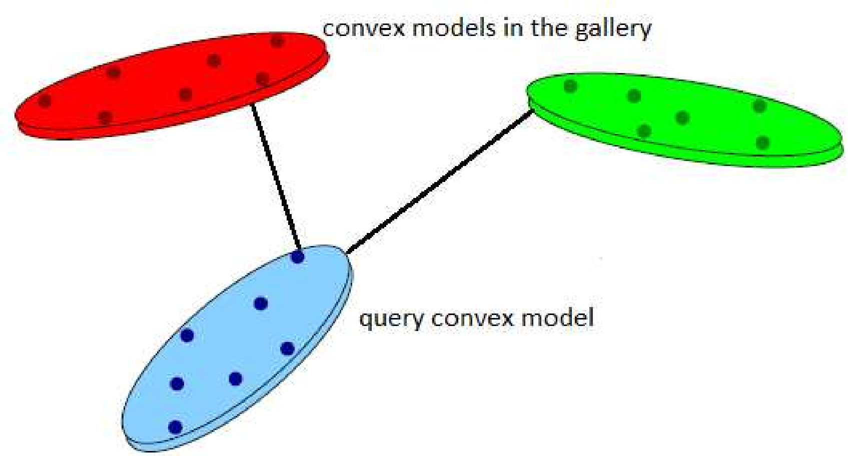
PROPOSED METHOD

- approximate each face image set with a convex model,
- use the geometric distances (distances of closest approach) in order to measure the similarity, i.e.,

$$D(H_i, H_j) = \min_{\mathbf{x}\in H_i, \mathbf{y}\in H_j} \| \mathbf{x} - \mathbf{y} \|$$

- assign the query to the gallery member whose convex model is closest to the query model.





Reduced Affine Hull Modeling

- In case of outliers (incorrect or very poor images) we need a more robust fitting. Thus we reduce affine hulls, i.e.,

$$H_c = \left\{ \mathbf{x} = \sum_{k=1}^{n_c} \alpha_{ck} \mathbf{x}_{ck} \mid \sum_{k=1}^{n_c} \alpha_{ck} = 1, L \le \alpha_{ck} \le U \right\}$$

- The minimum distance between two reduced affine hulls is computed by solving the following quadratic programming problem

$$(\boldsymbol{\alpha}_i^*, \boldsymbol{\alpha}_j^*) = \underset{\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j}{\arg\min} \| \mathbf{X}_i \boldsymbol{\alpha}_i - \mathbf{X}_j \boldsymbol{\alpha}_j \|^2$$

$$\text{such that} \sum_k \alpha_{ik} = \sum_{k'} \alpha_{jk'} = 1, L \le \alpha_{ik}, \alpha_{jk'} \le U.$$

$$\Rightarrow \quad D(H_i, H_j) = \| \mathbf{X}_i \boldsymbol{\alpha}_i^* - \mathbf{X}_j \boldsymbol{\alpha}_j^* \|$$

Convex Hull Modelling

- When the convex hulls are used for approximating image sets, the lower bound of the coefficients must set to zero in the quadratic problem, i.e.,

$$(\boldsymbol{\alpha}_i^*, \boldsymbol{\alpha}_j^*) = \underset{\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j}{\arg\min} \| \mathbf{X}_i \boldsymbol{\alpha}_i - \mathbf{X}_j \boldsymbol{\alpha}_j \|^2$$

$$\text{such that } \sum_k \alpha_{ik} = \sum_{k'} \alpha_{jk'} = 1, L \leq \alpha_{ik}, \alpha_{jk'} \leq U.$$

L = 0

- To handle outliers, we set U<1, resulting in an inner approximation to the image set.

EXTENSION TO THE NONLINEAR CASE – Kernel Trick

If linear models are not sufficient we can use the kernel trick to work over an implicit nonlinear feature space

- rewrite the problem in terms of inner products of image descriptors

$$(\boldsymbol{\alpha}_i^*, \boldsymbol{\alpha}_j^*) = \arg\min_{\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j} \| \mathbf{X}_i \boldsymbol{\alpha}_i - \mathbf{X}_j \boldsymbol{\alpha}_j \|^2 = \sum_{ij} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$

$$\text{such that } \sum_k \alpha_{ik} = \sum_{k'} \alpha_{jk'} = 1, L \le \alpha_{ik}, \alpha_{jk'} \le U.$$

- replace the inner products of image samples $\langle \mathbf{x}_{ck}, \mathbf{x}_{c'k'} \rangle$ with the kernel function $k(\mathbf{x}_{ck}, \mathbf{x}_{c'k'}) = \langle \phi(\mathbf{x}_{ck}), \phi(\mathbf{x}_{c'k'}) \rangle$

EXPERIMENTS (1)

Honda/UCSD Data Set:

- 59 video sequences of 300-500 frames, involving 20 individuals
  - we use 20 sequences for training, 39 for testing

To construct each image set we:

- detect faces in the sequence using Viola-Jones face detector
- resize the detected faces to 40x40 gray-scale images
- histogram equalize and take raw pixel features..

Some detected face images from videos of two subjects

EXPERIMENTS (2)

Table – Classification Rates (%) on the Honda/UCSD data set, respectively for the clean data, the data with noisy gallery sets but clean test ones, the data with clean gallery sets and noisy test ones, and the data with noise in both gallery and test sets.

Linear Methods	Clean	Noisy G.	Noisy T.	Noisy G+T.
Linear AHISD	97.4	**97.4**	**92.3**	87.2
Linear CHISD	94.9	92.3	**92.3**	82.1
MSM	97.4	**97.4**	87.2	76.9
Nonlinear Methods				
Kernel AHISD	97.4	**97.4**	**92.3**	**92.3**
Kernel CHISD	**100**	**97.4**	**92.3**	82.1
Spec Clus + Exemp.	94.9	89.7	84.6	79.5
Spec Clus + LS	97.4	**97.4**	89.7	79.5
Spec Clus + AH	97.4	94.9	**92.3**	82.1

AHISD – Proposed affine hull based method
CHISD - Proposed convex hull Based method
MSM - Mutual Subspace Method

EXPERIMENTS (3)

The MoBo (Motion of Body) Data Set:

- The MoBo dataset contains 4 videos each of 24 individuals walking on treadmill.
- Faces were detected, resized and histogram equalized as before.
- We also tested Local Binary Pattern descriptors over the normalized detections.
- One set from each four is used for the gallery, the rest for testing. The results are averages over 10 random repetitions.

EXPERIMENTS (4)

Average classification rates (%) for Gray Level and LBF Features on the MoBo dataset.

Linear Methods	**Gray Level**	**LBP**
Linear AHISD	$92.7^* \pm 3.3$	$94.6^* \pm 2.3$
Linear CHISD	94.2 ± 2.7	$\mathbf{98.1} \pm 0.9$
MSM	$92.0^* \pm 3.0$	$92.4^* \pm 1.9$
Nonlinear Methods		
Kernel AHISD	93.8 ± 2.8	97.6 ± 1.8
Kernel CHISD	$\mathbf{95.3} \pm 2.2$	$\mathbf{98.0} \pm 1.1$
Spec. Clus. + Exemplar	$85.5^* \pm 4.4$	$91.6^* \pm 3.0$
Spec. Clus. + LS	$88.2^* \pm 4.5$	$93.0^* \pm 2.8$
Spec. Clus. + AH	$89.5^* \pm 5.0$	$92.8^* \pm 2.2$

CONCLUSION

- Our convex model based methods are the best performers overall.
  - They are less sensitive to outliers than MSM and the other manifold learning methods.
- Among the manifold learning methods, linear models perform better than examplars (cluster centers).
- Kernelization improves the results at the expense of increased computation.

The proposed methods should be useful for other visual object classification problems too.

Thank you !

Software is available at

http://www2.ogu.edu.tr/~mlcv/softwares.html