Face Recognition in Unconstrained Environments:
A Comparative Study

Rodrigo Verschae, Javier Ruiz-del-Solar and Mauricio Correa

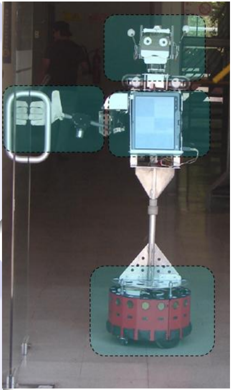
Department of Electrical Engineering, Universidad de Chile

ECCV'08 WS: Faces in Real-Life Images
October 17, 2008

Motivation

Applications

- Human Robot Interfaces (HRI)
- Face recognition in large databases
  - Personal data collections
  - Web segments
  - News databases



Requierements

Full online operation:

- Real world images: unconstrained environments.
- Incremental building of the database.
- Only one image per person in the database
- Fast processing (real time)

Analysis

We analyse/compare

- Variants of each methods
- Aligned (funneling) vs unaligned (output of the face detector)
- Amount of face/background
- Processing time

LFW

The analyzis is done using the image restricted setting of the LFW database

Methods

We consider four methods

Two local matching methods:

- Gabor Jet Descriptors + Borda Count [Zou et al. 2007]
- Local and global LBP histograms + distance between the histograms [Ahinene et al., 2004]

Two image matching methods:

- SIFT descriptors + matching and verification [Lowe 2004]
- ERCF of SIFT features + linear classifier (SVM) [Nowak et al. 2007]

Methods

We consider four methods

Two local matching methods:

- Gabor Jet Descriptors + Borda Count [Zou et al. 2007]
- Local and global LBP histograms + distance between the histograms [Ahinene et al., 2004]

Two image matching methods:

- SIFT descriptors + matching and verification [Lowe 2004]
- ERCF of SIFT features + linear classifier (SVM) [Nowak et al. 2007]

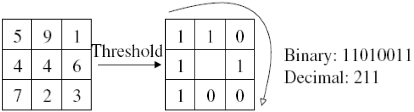
Genelized PCA wasn't considered

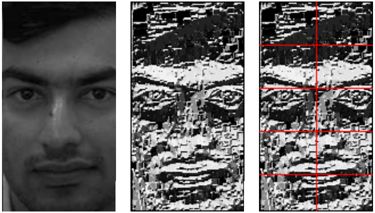
- Requieres very good alignment
- Low performance under occlusions and illumination changes
- Large processing time

[Ruiz-del-Solar et al. 20 08]

Local Binary Pattern (LBP) Histograms

LBP represents the local image structure and is invariant to local constrast changes.





Procedure

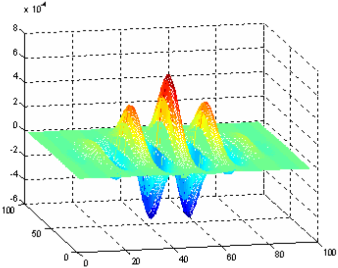
- Divided face area is into small regions:
  → 10 (2x5), 40 (4x10) and 80 (4x20) regions.
- Calculate the LBP histogram for each region, and for the full region.
- Concatenate the histograms to obtain a feature vector.
- Compare the two faces using:
  → euclidian distance (MSE) or chi square (XS).

Gabor Jet Descriptor (GJD) and Borda Count

GJD: Main Idea

A Gabor jet is the evaluation of a set gabor filters at a fixed scale (λ) and position (x, y), varying the orientation (ϕ).

- Gabor filters are applied different grid locations.
  - 8 orientations ($\phi = \frac{n\pi}{8}$ with $n = 0, .., 7$) .
- A grid is defined for each scale
  - 5 scales ($\lambda = 4, 4\sqrt{2}, 8, 8\sqrt{2}, 16$),



Gabor Jet Descriptor (GJD) and Borda Count

Borda Count: Main Idea

Voting system based on ranking

- Each voter ranks candidates in order preference.
- The accumulated inverse ranking is used to select the winner.

Procedure

- Each Gabor Jets assigns a vote (rank) to each image in the database.
- The ranking is done comparing the corresponding Gabor Jets using the normalized inner product.
- The final vote is obtained by adding the reverse rankings.

Gabor Jet Descriptor (GJD) and Borda Count

GJD and Borda Count for pairs of images

- BD is a voting system based on ranking
- It does not work on pairs of images
- Solution: use randomly selected reference set to build a similarity measure

Procedure to build a similarity measure $d(I_A, I_B)$

Take I_A and a reference set $S = \{I_1, \dots I_n\}$

- Rank $S \cup I_B$ using I_A
- Take position of I_B as a similarity measure d_A

Take I_B and a reference set $S = \{I_1, \dots I_n\}$

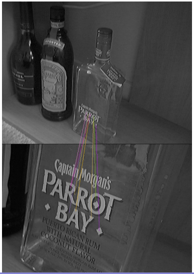
- Rank $S \cup I_A$ using I_B,
- Take position of I_A as a similarity measure d_B

return $d(I_A, I_B) = d_A + d_B$

SIFT

Main Idea

- Local interest points are extracted independently from both images.
- Characterized both images using invariant descriptors.
- Match the descriptors.
- Obtained a consistent transformation between the two images.
- Distance:
  - Number of matches (MATCHES)
  - Number of votes (SIMPLE)



ERCF

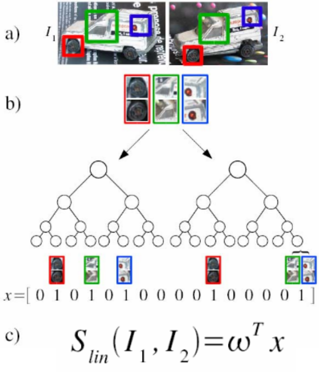
Main Idea

Learn similarity measure for pairs of images.

- Makes use of ERCF and SIFT descriptors.
- The learning is done for specific object classes (e.g frontal faces, car view).

Procedure

- Select pairs of similar patches using normalized cross-correlation.
- Code each pair of patches by means of an ERCF of SIFT descriptors.
- Obtain a similarity measure of the image pair using a SVM.



[Image from Nowak et al. CVPR'07]

Experiments: Cropped regions of size 100x185

Funneling

We compare two cases:

- Aligned (funneled) faces
- Unaligned faces

Cropping

- Regions are crop centered on the image
- Size: 100x185

Experiments: Cropped regions of size 100x185

	Without alignment		With alignment	
Method	MCA	SME	MCA	SME
H-MSE-10	0.6375	0.0049	**0.6585**	0.0046
H-XS-10	0.6500	0.0043	**0.6668**	0.0044
H-MSE-40	0.6217	0.0055	**0.6527**	0.0057
H-XS-40	0.6383	0.0064	**0.6650**	0.0059
H-MSE-80	0.6527	0.0047	**0.6725**	0.0032
H-XS-80	0.6532	0.0053	**0.6785**	0.0055
GJD-EU	0.6410	0.0084	0.6375	0.0071
GJD-BC-10	0.6777	0.0080	0.6753	0.0082
GJD-BC-50	0.6770	0.0075	0.6742	0.0061
GJD-BC-100	0.6798	0.0065	0.6762	0.0069
SD-MATCHES	0.6015	0.0049	**0.6215**	0.0036
SD-SIMPLE	0.6295	0.0071	0.6288	0.0051
ERCF (250x250)	0.7245	0.0040	**0.7333**	0.0060

MCA: Mean classification accuracy. SME: Standard error of the mean.

Experiments: region size

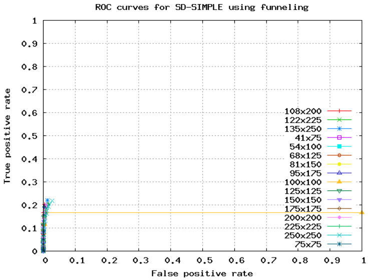
Sizes

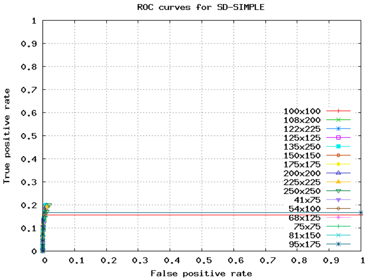
- Square image regions:
  $50x50, 75x75, 100x100, \ldots, 250x250$.
- Rectangular regions of ratio 1:1.85
  $41x75, \ldots, 135x250$.

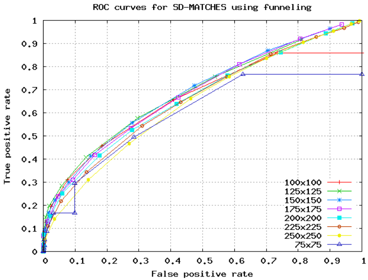
Evaluation:

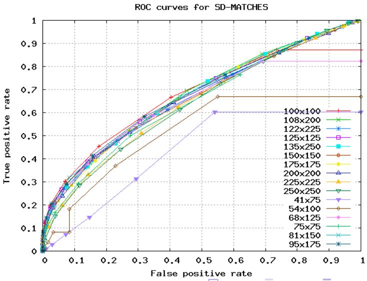
- Results are presented as ROC curves
  (True Positive rate vs False Positive rate).
- Processing time

Experiments: SD

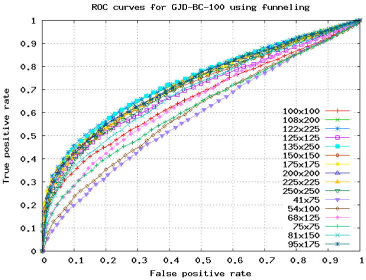


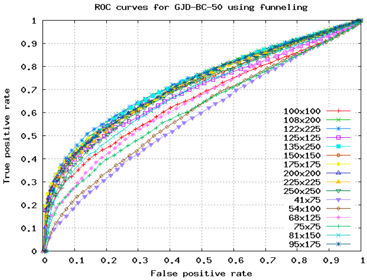


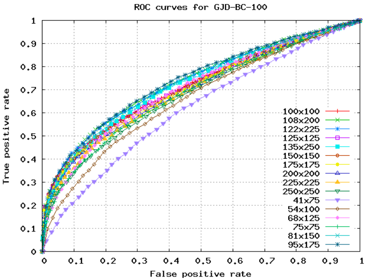


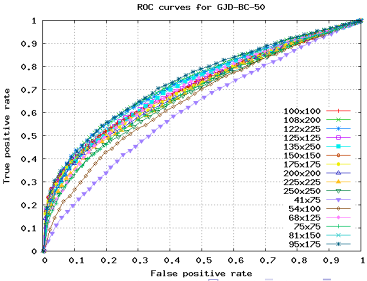


Experiments: GJD

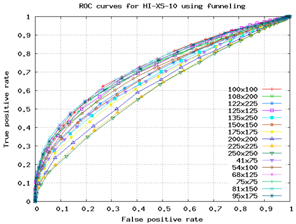


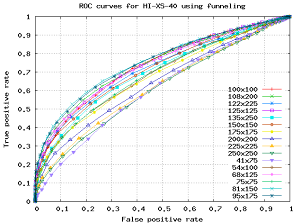


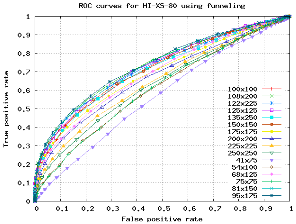


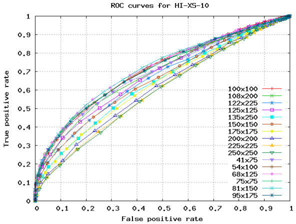


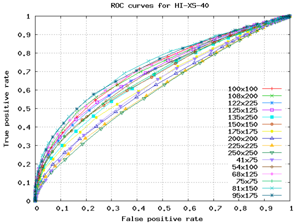
Experiments: H (LBP)

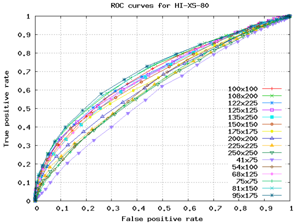




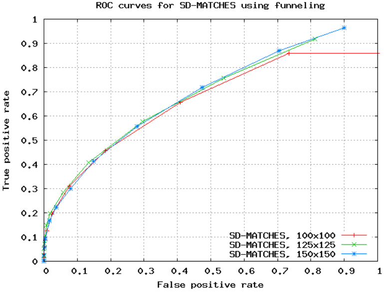


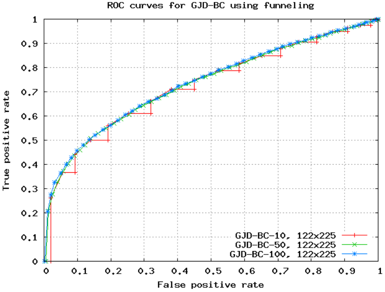


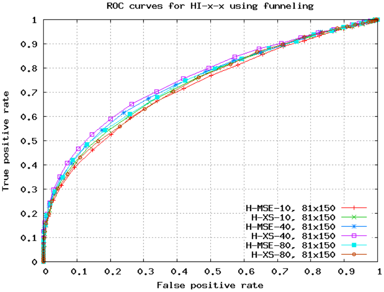




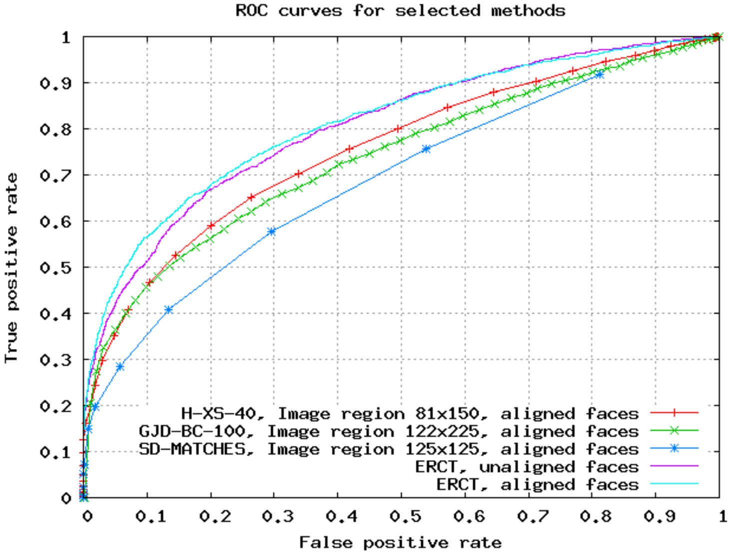
Experiments: Best for each method







Experiments



Experiments:Best size for each method

250x250	122x125	81x150	125x125
ERCF	GJD-BC	H-XS	SD

Discussion

Sizes

- In all cases (methods and parameters), small region sizes present the worst results, followed by the largest region sizes.
  Best results are obtained using medium-size regions.

Discussion

Sizes

- In all cases (methods and parameters), small region sizes present the worst results, followed by the largest region sizes.
  Best results are obtained using medium-size regions.

Sizes: GJD-BC-X

- Aligned images: best results for size 122x225
- Unaligned images: best results for size 95x175
- In both cases, best results are obtained with 100 reference images.
- Aligned faces: reference image set size (10, 50 or 100) gives **very similar results for the optimal face size**.
  (68.38%, 68.38% and 68.47% respectively)
- Unaligned faces: slightly larger difference
  (67.52%, 67.8% and 68.08% respectively).

Discussion

Sizes: H-X-X

- Aligned images: best results for size 81x150.
- Unaligned images: best results for size 95x175.
- In both cases,
  - **best: 40 divisions**
  - worst: 80 divisions
- For a fixed number of divisions, **Chi-Square works better** than the euclidean distance.

Discussion

Sizes: H-X-X

- Aligned images: best results for size 81x150.
- Unaligned images: best results for size 95x175.
- In both cases,
  - **best: 40 divisions**
  - worst: 80 divisions
- For a fixed number of divisions, **Chi-Square works better** than the euclidean distance.

Sizes: SD

- Aligned images: best results for size 125x125.
- Unaligned images: best results for size 100x100.
- **SD-Matches variant** gives best results in both cases.

Discussion: Processing time

Table: Accuracy

Method	H-XS-40	GJD-BC-100	SD-Matches	ERCF
Size	81x150	122x225	125x125	250x250
MCA	0.6945	0.6847	0.6410	0.7333
SME	0.0048	0.0065	0.0062	0.0060

Table: Average processing time [millisec]

Method	H-X			GJD-BC			SD	ERCF
Size	81x150			122x225			125x125	250x250
Params	10	40	80	10	50	100	-	Nowak'08
Time	3.8	5.0	6.4	200	320	480	65	2000

Best LBP-based method (H-XS-40) is almost 3.9% below ERCF, and about 1% over GJDs best method (GJD-BC-100). In terms of the processing speed of the methods, the best variant of the LBP-based methods (H-XS-40), is at least 400 times faster that ERCF, and 96 times faster that the best Gabor-best method (GJD-BC-100).

Conclusions

Dependence on the region size

- Large dependence of the methods to the amount of face and background information that is included
- Region size is as important as alignment
- When the optimal size is used, other parameters become less relevant.
- Masking might further help.

Conclusions

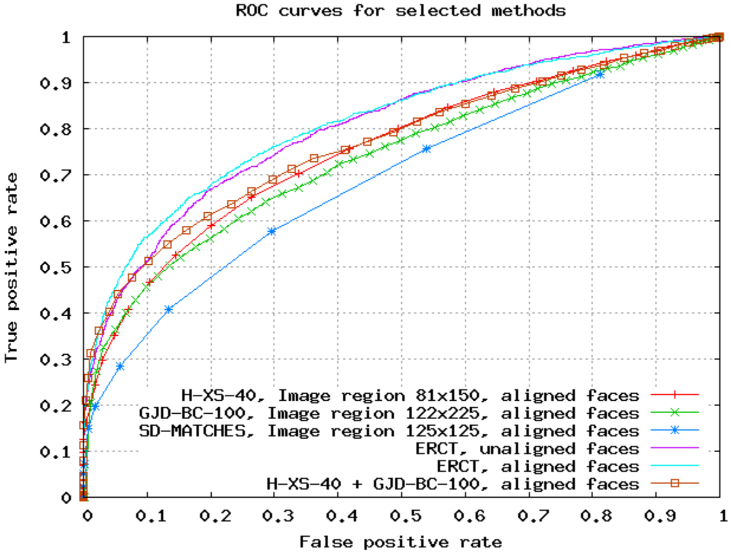
Dependence on the region size

- Large dependence of the methods to the amount of face and background information that is included
- Region size is as important as alignment
- When the optimal size is used, other parameters become less relevant.
- Masking might further help.

Summary

- LPB based method: Very fast, very simple and quite accurate.
- GJD based method: Relatively fast, simple and accurate.
- SIFT: not well fitted for face recognition.
- ERCF: Very slow, requires offline training, best performance on LFW. We have also evaluated ERCF in other databases: it seems to overfit and it has problems dealing with illumination changes.

Combining methods



Thank you for your attention

- http:\\rodrigo.verschae.org\about\ (presentation)
- http:\\vision.die.uchile.cl (UCH HRI Database)