

Stereo Matching for Unconstrained Face Recognition

Ph.D. Proposal

Carlos Castillo

Department of Computer Science
University of Maryland

April 10, 2009

1 Background

2 Method

- Epipolar Geometry
- Stereo Matching and Image Similarity
- Image Description

3 Results

- Pose Variation
- Pose+Illumination Variation

4 Research Plan

- Stereo Matching with Illumination Variation
- Learning for Unconstrained Face Recognition
- Face Recognition with Weight Variation

What is this talk about?

This talk is about comparing images of faces

Why is that interesting?

There can be variations:

- Pose
- Pose + Illumination
- Pose + Illumination + Expression
- Weight change



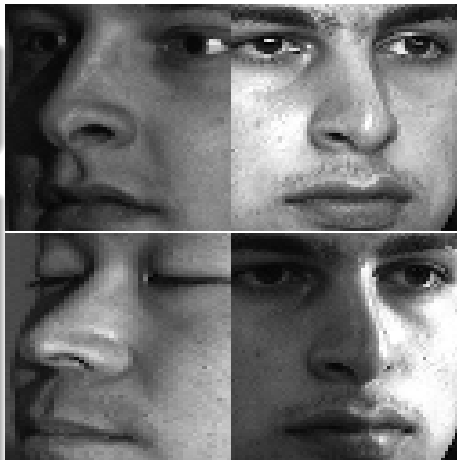
What is this talk about?

This talk is about comparing images of faces

Why is that interesting?

There can be variations:

- Pose
- Pose + Illumination
- Pose + Illumination + Expression
- Weight change



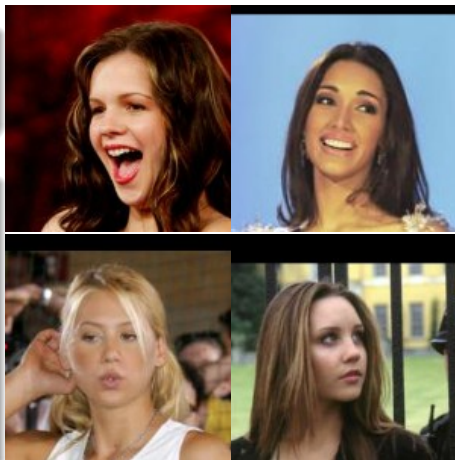
What is this talk about?

This talk is about comparing images of faces

Why is that interesting?

There can be variations:

- Pose
- Pose + Illumination
- Pose + Illumination + Expression
- Weight change



What is this talk about?

This talk is about comparing images of faces

Why is that interesting?

There can be variations:

- Pose
- Pose + Illumination
- Pose + Illumination + Expression
- Weight change



Existing Approaches

In this talk the following approaches will be mentioned:

- Eigen-lightfields (Gross et al.)
- LiST (Romdhani et al.)
- Extremely Randomized Clustering Forests (Nowak and Jurie)
- 3/4 patch Local Binary Patterns (Wolf et al.)

Existing Approaches

In this talk the following approaches will be mentioned:

- Eigen-lightfields (Gross et al.)
- LiST (Romdhani et al.)
- Extremely Randomized Clustering Forests (Nowak and Jurie)
- 3/4 patch Local Binary Patterns (Wolf et al.)

Existing Approaches

In this talk the following approaches will be mentioned:

- Eigen-lightfields (Gross et al.)
- LiST (Romdhani et al.)
- Extremely Randomized Clustering Forests (Nowak and Jurie)
- 3/4 patch Local Binary Patterns (Wolf et al.)

Existing Approaches

In this talk the following approaches will be mentioned:

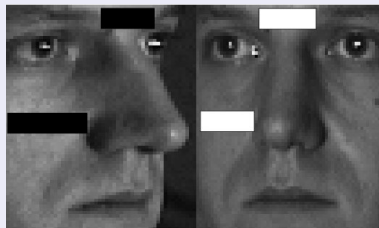
- Eigen-lightfields (Gross et al.)
- LiST (Romdhani et al.)
- Extremely Randomized Clustering Forests (Nowak and Jurie)
- 3/4 patch Local Binary Patterns (Wolf et al.)

Motivation

Correspondences for FR?

- Patches of the same area on the face will have different areas on the images.
- An affine transformation will maintain the ratios of images
- The greater generality afforded by stereo matching may be necessary for FR across pose.

Example



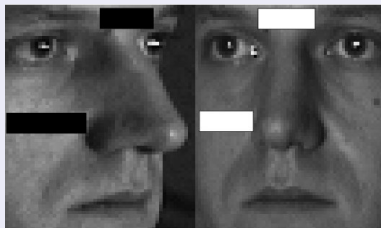
Example images from the CMU PIE dataset. Observe that no linear transformation can make corresponding boxes have equal size.

Motivation

Correspondences for FR?

- Patches of the same area on the face will have different areas on the images.
- An affine transformation will maintain the ratios of images
- The greater generality afforded by stereo matching may be necessary for FR across pose.

Example



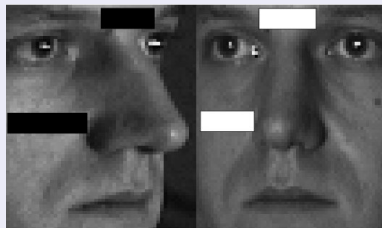
Example images from the CMU PIE dataset. Observe that no linear transformation can make corresponding boxes have equal size.

Motivation

Correspondences for FR?

- Patches of the same area on the face will have different areas on the images.
- An affine transformation will maintain the ratios of images
- The greater generality afforded by stereo matching may be necessary for FR across pose.

Example



Example images from the CMU PIE dataset. Observe that no linear transformation can make corresponding boxes have equal size.

Steps of our Method

- 1 Prior to recognition, build a gallery of 2D images of faces
- 2 Given a 2D probe image, compare the probe to each gallery image as follows:
 - Obtain the epipolar geometry and rectify the probe and gallery image.
 - Run a stereo algorithm on the image pair. Discard the correspondences and use the matching cost as a measure of image similarity.
- 3 Identify the probe with the gallery image that produces the lowest matching cost.

Epipolar Geometry

- Prior to stereo matching we need to obtain the epipolar geometry.
- Under scaled orthographic projection the epipolar geometry can be characterized by the tuple: (θ, γ, s, t) .
- We have two ways of obtaining the epipolar geometry:
 - ① Analytically from 4 hand-clicked points
 - ② Automatically using image information with off-the-shelf code to compute the egomotion.
- The analytic method solves a nonlinear system.
- The automatic method uses Probabilistic Egomotion of Domke and Aloimonos to compute the epipolar geometry.
- Computing the epipolar geometry this way is slow, and dominates the computation time of the similarity costs.

Epipolar Geometry

- Prior to stereo matching we need to obtain the epipolar geometry.
- Under scaled orthographic projection the epipolar geometry can be characterized by the tuple: (θ, γ, s, t) .
- We have two ways of obtaining the epipolar geometry:
 - 1 Analytically from 4 hand-clicked points
 - 2 Automatically using image information with off-the-shelf code to compute the egomotion.
- The analytic method solves a nonlinear system.
- The automatic method uses Probabilistic Egomotion of Domke and Aloimonos to compute the epipolar geometry.
- Computing the epipolar geometry this way is slow, and dominates the computation time of the similarity costs.

Epipolar Geometry

- Prior to stereo matching we need to obtain the epipolar geometry.
- Under scaled orthographic projection the epipolar geometry can be characterized by the tuple: (θ, γ, s, t) .
- We have two ways of obtaining the epipolar geometry:
 - ① Analytically from 4 hand-clicked points
 - ② Automatically using image information with off-the-shelf code to compute the egomotion.
- The analytic method solves a nonlinear system.
- The automatic method uses Probabilistic Egomotion of Domke and Aloimonos to compute the epipolar geometry.
- Computing the epipolar geometry this way is slow, and dominates the computation time of the similarity costs.

Epipolar Geometry

- Prior to stereo matching we need to obtain the epipolar geometry.
- Under scaled orthographic projection the epipolar geometry can be characterized by the tuple: (θ, γ, s, t) .
- We have two ways of obtaining the epipolar geometry:
 - ① Analytically from 4 hand-clicked points
 - ② Automatically using image information with off-the-shelf code to compute the egomotion.
- The analytic method solves a nonlinear system.
- The automatic method uses Probabilistic Egomotion of Domke and Aloimonos to compute the epipolar geometry.
- Computing the epipolar geometry this way is slow, and dominates the computation time of the similarity costs.

Epipolar Geometry

- Prior to stereo matching we need to obtain the epipolar geometry.
- Under scaled orthographic projection the epipolar geometry can be characterized by the tuple: (θ, γ, s, t) .
- We have two ways of obtaining the epipolar geometry:
 - 1 Analytically from 4 hand-clicked points
 - 2 Automatically using image information with off-the-shelf code to compute the egomotion.
- The analytic method solves a nonlinear system.
- The automatic method uses Probabilistic Egomotion of Domke and Aloimonos to compute the epipolar geometry.
- Computing the epipolar geometry this way is slow, and dominates the computation time of the similarity costs.

Epipolar Geometry

- Prior to stereo matching we need to obtain the epipolar geometry.
- Under scaled orthographic projection the epipolar geometry can be characterized by the tuple: (θ, γ, s, t) .
- We have two ways of obtaining the epipolar geometry:
 - 1 Analytically from 4 hand-clicked points
 - 2 Automatically using image information with off-the-shelf code to compute the egomotion.
- The analytic method solves a nonlinear system.
- The automatic method uses Probabilistic Egomotion of Domke and Aloimonos to compute the epipolar geometry.
- Computing the epipolar geometry this way is slow, and dominates the computation time of the similarity costs.

Stereo Matching

Why Stereo Matching?

Stereo matching allows for arbitrary, physically valid, continuous correspondences while maintaining an epipolar constraint. Correspondences are of fundamental importance for recognition.

Which Stereo Matching Algorithm?

- We require a fast stereo algorithm
- The algorithm has to be appropriate for wide baseline matching of faces

Stereo Matching

Why Stereo Matching?

Stereo matching allows for arbitrary, physically valid, continuous correspondences while maintaining an epipolar constraint. Correspondences are of fundamental importance for recognition.

Which Stereo Matching Algorithm?

- We require a fast stereo algorithm
- The algorithm has to be appropriate for wide baseline matching of faces

Stereo Matching

Performance Characteristics of Stereo for FR

- We are unaffected by difficulties that generate artifacts in stereo reconstruction
- We don't use the correspondences, we only use the matching cost

4-plane Stereo Matching Algorithm

- We use the 4-plane stereo matching algorithm of Criminisi et al.
- The method was developed for video conferencing applications and therefore seems fit for our needs.
- This method matches N pixels on one scan line to M pixels on the other

Stereo Matching

Performance Characteristics of Stereo for FR

- We are unaffected by difficulties that generate artifacts in stereo reconstruction
- We don't use the correspondences, we only use the matching cost

4-plane Stereo Matching Algorithm

- We use the 4-plane stereo matching algorithm of Criminisi et al.
- The method was developed for video conferencing applications and therefore seems fit for our needs.
- This method matches N pixels on one scan line to M pixels on the other

Stereo Matching

Performance Characteristics of Stereo for FR

- We are unaffected by difficulties that generate artifacts in stereo reconstruction
- We don't use the correspondences, we only use the matching cost

4-plane Stereo Matching Algorithm

- We use the 4-plane stereo matching algorithm of Criminisi et al.
- The method was developed for video conferencing applications and therefore seems fit for our needs.
- This method matches N pixels on one scan line to M pixels on the other

Stereo Matching

Performance Characteristics of Stereo for FR

- We are unaffected by difficulties that generate artifacts in stereo reconstruction
- We don't use the correspondences, we only use the matching cost

4-plane Stereo Matching Algorithm

- We use the 4-plane stereo matching algorithm of Criminisi et al.
- The method was developed for video conferencing applications and therefore seems fit for our needs.
- This method matches N pixels on one scan line to M pixels on the other

Stereo Matching

Performance Characteristics of Stereo for FR

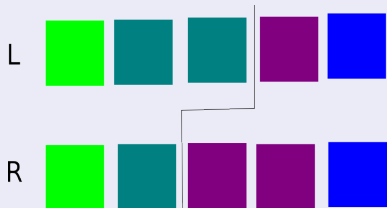
- We are unaffected by difficulties that generate artifacts in stereo reconstruction
- We don't use the correspondences, we only use the matching cost

4-plane Stereo Matching Algorithm

- We use the 4-plane stereo matching algorithm of Criminisi et al.
- The method was developed for video conferencing applications and therefore seems fit for our needs.
- This method matches N pixels on one scan line to M pixels on the other

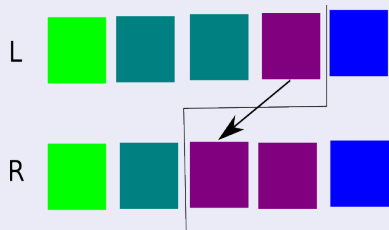
The Core of the 4-plane Stereo Algorithm

- Given two scan lines of length n and m respectively, the method searches the space of words of length $n + m$ in the alphabet $\Sigma = \{L_O, L_M, R_O, R_M\}$.
- At each step one of four things can be done:
 - Match on the left
 - Occlude on the left
 - Match on the right
 - Occlude on the right



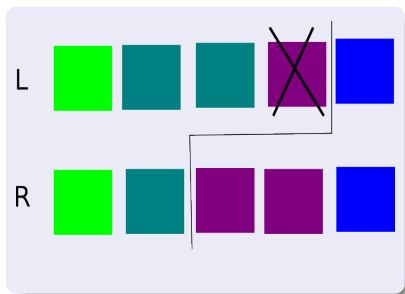
The Core of the 4-plane Stereo Algorithm

- Given two scan lines of length n and m respectively, the method searches the space of words of length $n + m$ in the alphabet $\Sigma = \{L_O, L_M, R_O, R_M\}$.
- At each step one of four things can be done:
 - Match on the left
 - Occlude on the left
 - Match on the right
 - Occlude on the right



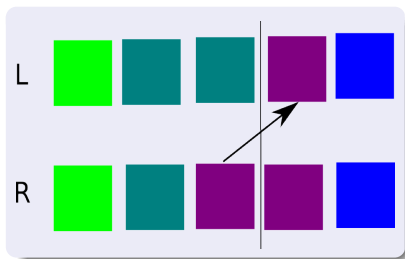
The Core of the 4-plane Stereo Algorithm

- Given two scan lines of length n and m respectively, the method searches the space of words of length $n + m$ in the alphabet $\Sigma = \{L_O, L_M, R_O, R_M\}$.
- At each step one of four things can be done:
 - Match on the left
 - Occlude on the left
 - Match on the right
 - Occlude on the right



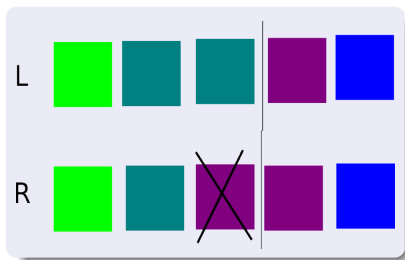
The Core of the 4-plane Stereo Algorithm

- Given two scan lines of length n and m respectively, the method searches the space of words of length $n + m$ in the alphabet $\Sigma = \{L_O, L_M, R_O, R_M\}$.
- At each step one of four things can be done:
 - Match on the left
 - Occlude on the left
 - Match on the right
 - Occlude on the right



The Core of the 4-plane Stereo Algorithm

- Given two scan lines of length n and m respectively, the method searches the space of words of length $n + m$ in the alphabet $\Sigma = \{L_O, L_M, R_O, R_M\}$.
- At each step one of four things can be done:
 - Match on the left
 - Occlude on the left
 - Match on the right
 - Occlude on the right

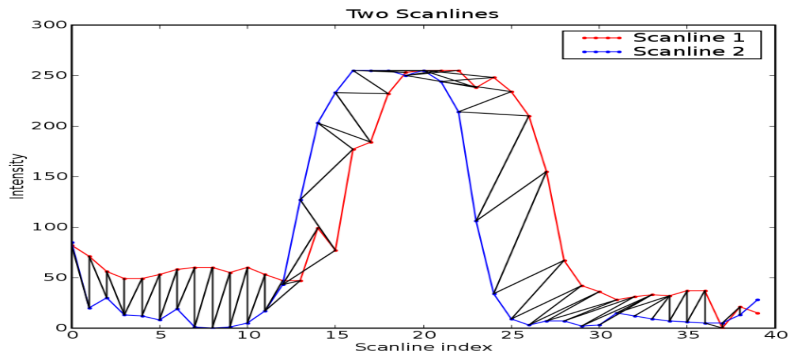


From Stereo Correspondences to Image Similarity

Image Comparison

- The stereo matching cost gives us a measure of how similar two scanlines are.

Correspondences



From Stereo Correspondences to Image Similarity

Image Comparison

- We can therefore add the similarity cost of every pair of corresponding scanlines:

$$\text{cost}(l_1, l_2) = \frac{\sum_{i=1}^n \text{cost}(l_{1,i}, l_{2,i})}{\sum_{i=1}^n |l_{1,i}| + |l_{2,i}|} \quad (1)$$

- Since the face is a vertically symmetric object we can define the similarity cost to be:

$$\text{similarity}(l_1, l_2) = \min \begin{cases} \text{cost}(\text{rectify}(l_1, l_2)) \\ \text{cost}(\text{rectify}(l_2, l_1)) \\ \text{cost}(\text{rectify}(\text{flip}(l_1), l_2)) \\ \text{cost}(\text{rectify}(l_2, \text{flip}(l_1))) \end{cases} \quad (2)$$

Image Description

Image Description in Stereo

- $M(l, r)$ indicates the similarity of a window centered (l, s) in one image and (r, s) on the other. Note that this can be any measure of similarity on any image representation.
- We have experimented with two: Normalized SSD and DHOG. We have found that:
 - In controlled conditions (but more pose variation) such as PIE, NSSD is a better image description.
 - In unconstrained conditions (but less pose variation) such as LFW, the SIFT-like DHOG descriptor is a better image description.














Image Description

Image Description in Stereo

- $M(l, r)$ indicates the similarity of a window centered (l, s) in one image and (r, s) on the other. Note that this can be any measure of similarity on any image representation.
- We have experimented with two: Normalized SSD and DHOG. We have found that:
 - In controlled conditions (but more pose variation) such as PIE, NSSD is a better image description.
 - In unconstrained conditions (but less pose variation) such as LFW, the SIFT-like DHOG descriptor is a better image description.

Pose Variation in PIE

Table: Accuracy for pose variation for 68 faces with 4ptSMD. The diagonals are not included in any average. The global average for this table is 82.4%

azimuth altitude Probe Pose	-66 3 c34 	-47 13 c31 	-46 2 c14 	-32 2 c11 	-17 2 c29 	0 15 c09 	0 2 c27 	0 1.9 c07 	16 2 c05 	31 2 c37 	44 2 c25 	44 13 c02 	62 3 c22 	avg
Gallery Pose														
c34	-	79	91	78	65	38	44	26	50	50	60	71	56	59
c31	91	-	99	96	94	78	65	50	62	65	84	72	60	76
c14	97	100	-	97	91	87	79	71	79	76	59	76	78	82
c11	94	97	99	-	100	97	94	94	88	94	79	87	65	90
c29	87	97	96	100	-	100	99	100	96	94	82	81	53	90
c09	54	91	84	99	100	-	100	97	94	94	85	90	65	87
c27	60	93	91	97	99	99	-	100	97	99	97	97	62	90
c07	40	62	79	97	100	96	100	-	100	99	88	97	32	82
c05	71	79	90	93	97	97	99	100	-	100	100	99	78	91
c37	66	74	85	94	90	91	97	99	100	-	100	100	91	90
c25	65	79	56	66	71	85	91	79	97	100	-	99	94	81
c02	81	71	74	81	69	93	90	85	93	100	99	-	99	86
c22	57	62	66	56	44	49	47	35	66	76	88	91	-	61

Summary of Results with Pose on PIE










34 Faces

Method	Accuracy
Eigenfaces	16.6%
Facelt	24.3%
Eigen light-fields (3-point norm.)	52.5%
Eigen light-fields (Multi-point norm.)	66.3%
4-point Stereo Matching Distance	86.8%

68 Faces

Method	Accuracy
LiST (Romdhani et al.)	74.3%
4-point Stereo Matching Distance	82.4%

Pose+Illumination Variation in PIE

light	F Gallery			S Gallery			P Gallery		
	F 	S 	P 	F 	S 	P 	F 	S 	P 
2	94/38	93/44	32/4	85/41	100/53	41/6	26/21	18/25	100/47
3	96/68	96/76	35/13	93/65	100/85	41/9	29/25	16/25	100/51
4	97/82	94/87	37/24	96/82	100/94	35/12	34/25	24/31	100/66
5	99/100	99/97	35/34	99/97	100/100	47/32	38/35	25/29	100/94
6	100/99	100/99	41/35	100/99	100/100	57/56	38/29	43/24	100/100
7	100/99	99/97	37/34	99/87	100/100	53/49	29/21	35/16	100/100
8	100/100	100/100	44/37	100/100	100/100	56/60	35/19	43/25	100/100
9	100/100	100/100	44/44	100/100	100/100	65/62	40/35	47/46	100/100
10	99/90	99/93	29/34	99/88	100/99	49/35	32/28	28/21	100/87
11	100/100	100/100	46/44	100/100	100/100	60/56	47/32	49/35	100/100
12	-/-	100/100	53/44	100/100	-/-	71/62	49/46	56/53	-/-
13	100/100	100/100	46/41	100/100	100/100	63/49	44/43	49/49	100/100
14	100/100	100/100	47/43	100/100	100/100	66/49	44/46	59/53	100/100
15	100/100	99/94	46/31	100/100	100/100	54/40	37/46	60/54	100/100
16	100/100	97/74	40/21	100/97	100/99	51/32	40/41	53/47	100/100
17	100/90	96/49	35/19	99/75	100/84	49/26	32/41	44/47	100/100
18	99/91	99/97	37/28	99/90	100/97	38/25	35/37	22/32	100/79
19	100/100	100/99	38/29	100/99	100/100	54/38	43/35	44/32	100/99
20	100/100	100/100	44/38	100/100	100/100	63/51	49/41	51/40	100/100
21	100/100	100/100	50/40	100/100	100/100	65/54	47/47	57/53	100/100
22	100/100	100/99	50/37	100/100	100/100	57/40	38/46	60/54	100/100
avg	99/92	98/90	41/32	98/91	100/95	54/40	38/35	42/37	100/91

Stereo Matching with Illumination Variation

Objective

Build a stereo matching method that is both locally contrast invariant and robust to viewpoint changes

Challenges

- We require dense correspondences. Every pixel in each image should be accounted for.
- We require effective handling for wide baselines. Wide baselines have important consequences on the treatment of slant.
- The images were not taken at the same instant of time. There is illumination variation and perhaps even some deformation.

Stereo Matching with Illumination Variation

Objective

Build a stereo matching method that is both locally contrast invariant and robust to viewpoint changes

Challenges

- We require dense correspondences. Every pixel in each image should be accounted for.
- We require effective handling for wide baselines. Wide baselines have important consequences on the treatment of slant.
- The images were not taken at the same instant of time. There is illumination variation and perhaps even some deformation.

Stereo Matching with Illumination Variation

Objective

Build a stereo matching method that is both locally contrast invariant and robust to viewpoint changes

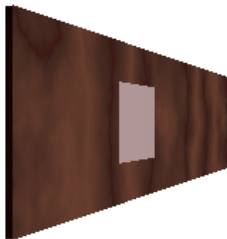
Challenges

- We require dense correspondences. Every pixel in each image should be accounted for.
- We require effective handling for wide baselines. Wide baselines have important consequences on the treatment of slant.
- The images were not taken at the same instant of time. There is illumination variation and perhaps even some deformation.

Illumination and Slant

Slant is Key!

- To provide contrast invariance it is necessary to use window based comparison. We therefore need to compensate for the effects of slant.
- There can be significant changes in slant in wide baseline stereo.
- Knowing the slant will allow us to adjust the matching window in each image and allow for less noisy local image comparisons.



Adjusting for Slant

Simple trigonometry gives us:

$$\cos(\alpha_1) = \frac{w_1}{w_f} \quad \text{and} \quad \cos(\alpha_2) = \frac{w_2}{w_f}$$

we know that relative slant the derivative of the disparity:

$$\tan(\alpha_1) - \tan(\alpha_2) \propto d'$$

which makes for a fairly constrained search problem.

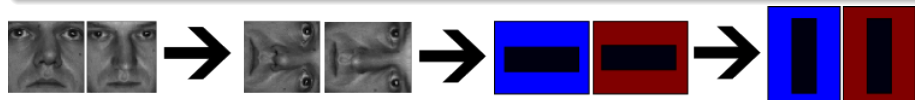
Image Representation, Correspondences and Learning for Unconstrained Face Recognition

Objective

Integrate spatial learning into our method to account for variations not explicitly handled by the pose+illumination model.

Challenges

- Requires a better understanding of image representation and correspondences.
- There are two types of alignment involved: one for image comparison and one for spatial learning.



Rectification

Description

Backprojection

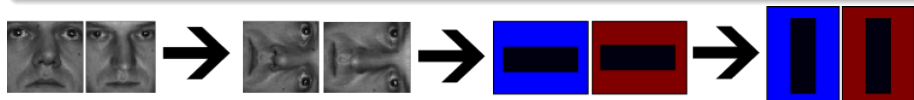
Image Representation, Correspondences and Learning for Unconstrained Face Recognition

Objective

Integrate spatial learning into our method to account for variations not explicitly handled by the pose+illumination model.

Challenges

- Requires a better understanding of image representation and correspondences.
- There are two types of alignment involved: one for image comparison and one for spatial learning.



Rectification

Description

Backprojection

Preliminary Results on LFW

Method	Accuracy
Eigenfaces	0.6002 \pm 0.0079
MERL	0.7052 \pm 0.0060
SMD + NSSD+ Probabilistic Egomotion, funneled	0.7092 \pm 0.0055
Nowak, original	0.7245 \pm 0.0040
SMD + DHOG + Probabilistic Egomotion, funneled	0.7251 \pm 0.0050
3x3 Multi-region histograms	0.7295 \pm 0.0055
Nowak, funneled	0.7393 \pm 0.0049
MERL+Nowak, funneled	0.7618 \pm 0.0058
SMD + difference descriptor + DHOG + Probabilistic Egomotion + Spatial learning, funneled	0.7690 \pm 0.0052
Hybrid descriptor-based, funneled	0.7847 \pm 0.0051

Face Recognition with Weight Variation

Motivation

- Many applications require that we recognize an individual using photos taken months or years apart.
- We have collected images with weight variations through different publicly available sources.



Figure: Facial changes as weight variation increases (images shown with permission of subject).

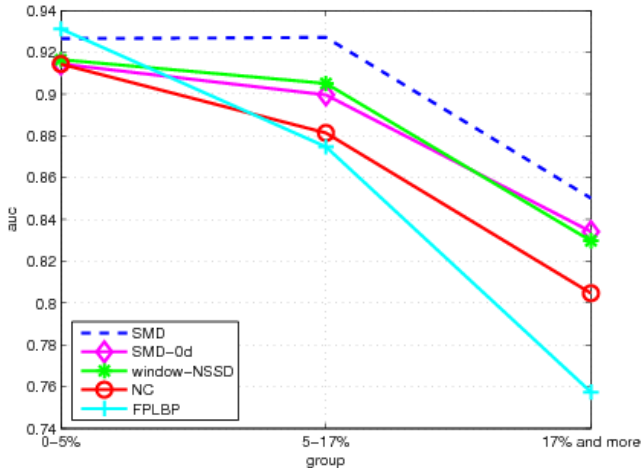


Figure: Performance of classifiers by groups of similar relative weight variation.

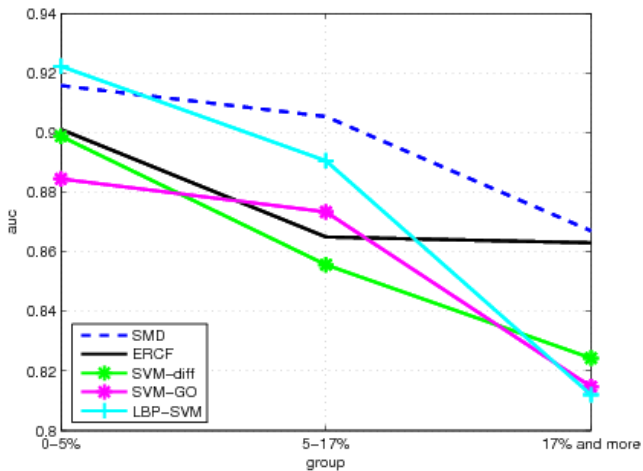


Figure: Performance of learning-based classifiers by groups of similar relative weight variation.

Closing Remarks

- Existing stereo algorithms produce correspondences that are useful for face recognition.
- Direct image comparison is very powerful. It is easier to compare two images than to reconstruct a scene.
- We plan to account for pose+illumination. We plan to handle other variations using spatial learning.

Closing Remarks

- Existing stereo algorithms produce correspondences that are useful for face recognition.
- Direct image comparison is very powerful. It is easier to compare two images than to reconstruct a scene.
- We plan to account for pose+illumination. We plan to handle other variations using spatial learning.

Closing Remarks

- Existing stereo algorithms produce correspondences that are useful for face recognition.
- Direct image comparison is very powerful. It is easier to compare two images than to reconstruct a scene.
- We plan to account for pose+illumination. We plan to handle other variations using spatial learning.