Head pose recovery and shape estimation in still images

Marc Oliu Simón

Advisors: Sergio Escalera and Xavier Pérez-Sala

July 8, 2014

Motivations

Head pose recovery and *3D shape estimation* *have a wide range of applications for both image and behaviour analysis.*

Image quality assessment

- Pose is correct?
- Mouth open/eyes closed?
- Obstructions? (eyeglasses, scarf)

Driver attention evaluation

- Looking at the road?
- Yawning?
- Eyes closing?

Index

Outline

Supervised Descent Method

- Method overview
- Augmenting the training data
- Describing the landmarks
- Regressing the shape estimates

Supervised Descent Method: Overview

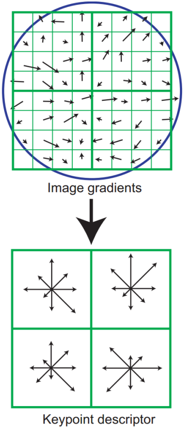
- Augment training data: multiple initializations
- Initialize validation estimates to mean shape and location
- While validation error descending
  - Extract descriptors at landmark estimates
  - Concatenate descriptors to single feature vector
  - Perform PCA to reduce features dimensionality
  - Train linear regressor for the shape
  - Update shape estimates using the regressor

Supervised Descent Method: Stochastic initializations

Multiple training instances are generated for each training image. These have different initial shape estimates in order to increase the variability of $\Phi_* - \Phi_0$.

- Affected parameters modelled as a normal distribution
- Monte-Carlo sampling for each training instance
- Affected parameters: Scale, rotation angle and offsets

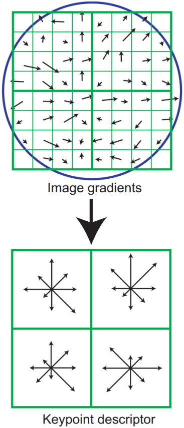
Supervised Descent Method: Simplified SIFT descriptor



The Scale Invariant Feature Transform descriptor describes a suqared region around a point as a grid of normalized gradient histograms.

- Smooth image with a gaussian filter ($\sigma = 1.6$)
- Find preferent gradient orientation
  - 36-bin orientations histogram
  - Gaussian weight to magnitudes ($\sigma = scale/2$)
- Rotate window to preferent orientation
- Extract histogram at each grid cell
  - Normalized 8-bin orientations histogram
  - Gaussian weight to magnitudes ($\sigma = scale/2$)

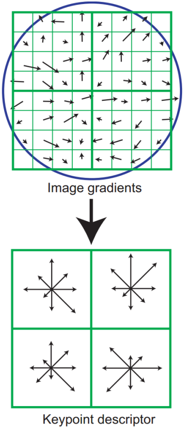
Supervised Descent Method: Simplified SIFT descriptor



The Scale Invariant Feature Transform descriptor describes a suqared region around a point as a grid of normalized gradient histograms.

- Smooth image with a gaussian filter ($\sigma = 1.6$)
- Find preferent gradient orientation
  - 36-bin orientations histogram
  - Gaussian weight to magnitudes ($\sigma = scale/2$)
- Rotate window to preferent orientation
- Extract histogram at each grid cell
  - Normalized 8-bin orientations histogram
  - Gaussian weight to magnitudes ($\sigma = scale/2$)

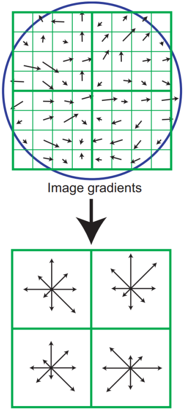
Supervised Descent Method: Simplified SIFT descriptor



The Scale Invariant Feature Transform descriptor describes a suqared region around a point as a grid of normalized gradient histograms.

- Smooth image with a gaussian filter ($\sigma = 1.6$)
- Find preferent gradient orientation
  - 36-bin orientations histogram
  - Gaussian weight to magnitudes ($\sigma = scale/2$)
- Rotate window to preferent orientation
- Extract histogram at each grid cell
  - Normalized 8-bin orientations histogram
  - Gaussian weight to magnitudes ($\sigma = scale/2$)

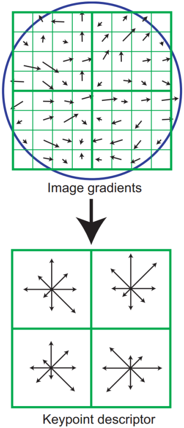
Supervised Descent Method: Simplified SIFT descriptor



The Scale Invariant Feature Transform descriptor describes a suqared region around a point as a grid of normalized gradient histograms.

- Smooth image with a gaussian filter ($\sigma = 1.6$)
- Find preferent gradient orientation
  - 36-bin orientations histogram
  - Gaussian weight to magnitudes ($\sigma = scale/2$)
- Rotate window to preferent orientation
- Extract histogram at each grid cell
  - Normalized 8-bin orientations histogram
  - Gaussian weight to magnitudes ($\sigma = scale/2$)

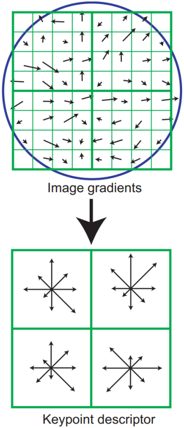
Supervised Descent Method: Simplified SIFT descriptor



The Scale Invariant Feature Transform descriptor describes a suqared region around a point as a grid of normalized gradient histograms.

- Smooth image with a gaussian filter ($\sigma = 1.6$)
- Find preferent gradient orientation
  - 36-bin orientations histogram
  - Gaussian weight to magnitudes ($\sigma = scale/2$)
- Rotate window to preferent orientation
- Extract histogram at each grid cell
  - Normalized 8-bin orientations histogram
  - Gaussian weight to magnitudes ($\sigma = scale/2$)

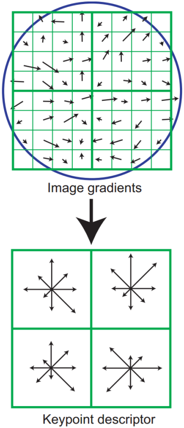
Supervised Descent Method: Simplified SIFT descriptor



The Scale Invariant Feature Transform descriptor describes a suqared region around a point as a grid of normalized gradient histograms.

- Smooth image with a gaussian filter ($\sigma = 1.6$)
- Find preferent gradient orientation
  - 36-bin orientations histogram
  - Gaussian weight to magnitudes ($\sigma = scale/2$)
- Rotate window to preferent orientation
- Extract histogram at each grid cell
  - Normalized 8-bin orientations histogram
  - Gaussian weight to magnitudes ($\sigma = scale/2$)

Supervised Descent Method: Simplified SIFT descriptor

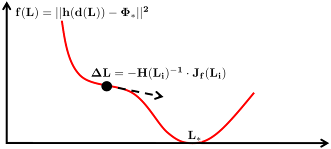


The Scale Invariant Feature Transform descriptor describes a suqared region around a point as a grid of normalized gradient histograms.

- Smooth image with a gaussian filter ($\sigma = 1.6$)
- Find preferent gradient orientation
  - 36-bin orientations histogram
  - Gaussian weight to magnitudes ($\sigma = scale/2$)
- Rotate window to preferent orientation
- Extract histogram at each grid cell
  - Normalized 8-bin orientations histogram
  - Gaussian weight to magnitudes ($\sigma = scale/2$)

Supervised Descent Method: Linear regressors

Linear regressors are used as a data-driven approach to gradient descent, minimizing the difference between the target feature vector Φ_* and the current one, Φ_i



Newton method

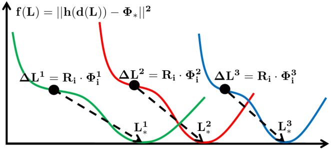
Prone to local minima

Descent depends on slope

- Start with Newton's descent method
  $f(L_i + \Delta L_{i+1}) = ||h(d(L_i + \Delta L_{i+1})) - \Phi_*||_2^2$
- Second order taylor expansion
  $f(L_i + \Delta L_{i+1}) \approx$
  $f(L_i) + J_f(L_i)^T \Delta L_i + \frac{1}{2} \Delta L^T H_f(L_i) \Delta L$
- Simplification
  $\Delta L_1 = R_i \cdot \Phi_i - R_i \cdot \Phi_* = R_i \cdot \Phi_i - b_i$
  where $R_i = -2H_f^{-1} J_f^T$

Supervised Descent Method: Linear regressors

Linear regressors are used as a data-driven approach to gradient descent, minimizing the difference between the target feature vector Φ_* and the current one, Φ_i



SDM method

Avoids local minima

Direct descent to target

- Start with Newton's descent method
  $f(L_i + \Delta L_{i+1}) = ||h(d(L_i + \Delta L_{i+1})) - \Phi_*||_2^2$
- Second order taylor expansion
  $f(L_i + \Delta L_{i+1}) \approx$
  $f(L_i) + J_f(L_i)^T \Delta L_i + \frac{1}{2} \Delta L^T H_f(L_i) \Delta L$
- Simplification
  $\Delta L_1 = R_i \cdot \Phi_i - R_i \cdot \Phi_* = R_i \cdot \Phi_i - b_i$
  where $R_i = -2 H_f^{-1} J_f^T$

Outline

Parametric approach
- Active Shape Models
- Parametric representation
- Method improvements

ASM: Shapes alignment

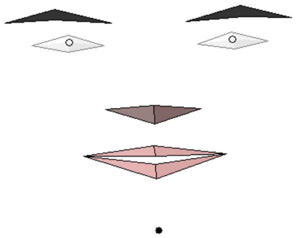
Shapes are aligned through generalized procrustes analysis, an iterative process aligning all the shapes to their mean in the canonical form.

- Initialize each shape L_i^{2D} transform tfm_i to the identity matrix
- While not converged
  - Calculate mean shape
    $M = \left[(L_1^{2D} \; .. \; L_n^{2D}) \cdot (tfm_1^T \; .. \; tfm_n^T)^T\right] / n$
  - Bring mean shape to the canonical form
    $M = M \cdot (M_{i_C}^{\dagger} \cdot C)$
  - Find transforms aligning shapes to the mean
    $tfm_i = (L_i^{2D})^{\dagger} \cdot M$

ASM: Principal Component Analysis

Find a transform for the feature space giving a set of uncorrelated dimensions following the directions of maximum variance of the data.

- Subtract mean from aligned shapes
- Obtain correlation matrix
- Diagonalize matrix
  - Matrix of eigenvectors
  - Eigenvalues diagonal matrix
- Keep 95% of the variance



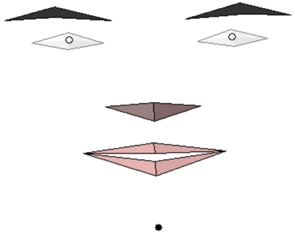
PCA components describe main modes of shape deformation.

- 1st eigenvector: Face yaw
- 2nd eigenvector: Face pitch

ASM: Principal Component Analysis

Find a transform for the feature space giving a set of uncorrelated dimensions following the directions of maximum variance of the data.

- Subtract mean from aligned shapes
- Obtain correlation matrix
- Diagonalize matrix
  - Matrix of eigenvectors
  - Eigenvalues diagonal matrix
- Keep 95% of the variance



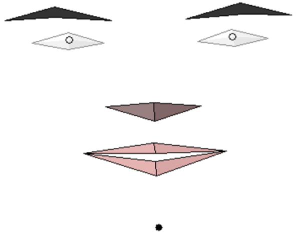
PCA components describe main modes of shape deformation.

- 1st eigenvector: Face yaw
- 2nd eigenvector: Face pitch

ASM: Principal Component Analysis

Find a transform for the feature space giving a set of uncorrelated dimensions following the directions of maximum variance of the data.

- Subtract mean from aligned shapes
- Obtain correlation matrix
- Diagonalize matrix
  - Matrix of eigenvectors
  - Eigenvalues diagonal matrix
- Keep 95% of the variance



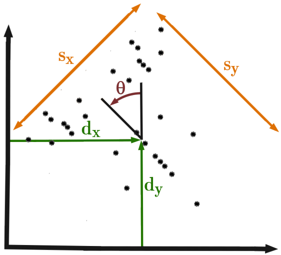
PCA components describe main modes of shape deformation.

- 1st eigenvector: Face yaw
- 2nd eigenvector: Face pitch

Parametric representation: Model location parameters

$$p = \langle b, s_x, s_y, d_x, d_y, \theta \rangle$$

The ASM weights only describe the aligned shape deformation. Extra parameters are required to align the shape to the image.



- ASM weights
- Scaling factors
- Translation parameters
- Rotation

Parametric representation: Reduced feature vectors

- Fixed number of location parameters
- Less ASM weights than landmark coordinates
  - AFLW: 42 landmarks, 21+5 parameters
  - LFPW: 58 landmarks, 23+5 parameters

The number of weights at each linear regressor is proportional to the number of regressed parameters. With a smaller weights matrix the algorithm can generalize better.

Adaptive SIFT window sizes

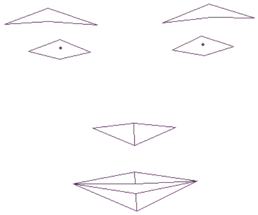
$$SIFT_d = SIFT_{d'} \cdot \frac{s_x + s_y}{2}$$

The parametric approach directly provides a scaling factor for the face at each cascade step. This can be used to adapt the SIFT window size, keeping it proportional to the face scale.

- Provides invariance to scale
- SIFT descriptors consistent across instances

Centroid fit selection

Selects the fitting where most initializations converge, ignoring those stuck in a local minima.



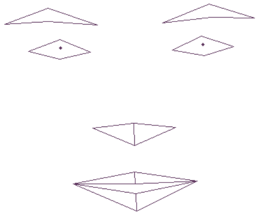
- Multiple initializations at different rotation angles
- Fit all initializations to the image
- Calculate euclidean distance between each pair of fits

$$d(L^i, L^j) = \sum_{p=1}^{n} \sqrt{(x_p^j - x_p^i)^2 + (y_p^j - y_p^i)^2}$$

- Select fit minimizing the sum of distances

Centroid fit selection

Selects the fitting where most initializations converge, ignoring those stuck in a local minima.



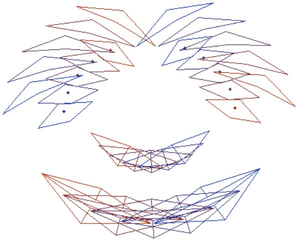
- Multiple initializations at different rotation angles
- Fit all initializations to the image
- Calculate euclidean distance between each pair of fits

$$d(L^i, L^j) = \sum_{p=1}^{n} \sqrt{(x_p^j - x_p^i)^2 + (y_p^j - y_p^i)^2}$$

- Select fit minimizing the sum of distances

Centroid fit selection

Selects the fitting where most initializations converge, ignoring those stuck in a local minima.



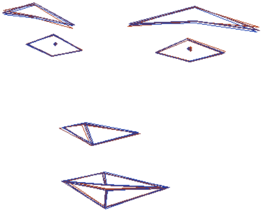
- Multiple initializations at different rotation angles
- Fit all initializations to the image
- Calculate euclidean distance between each pair of fits

$$d(L^i, L^j) = \sum_{p=1}^{n} \sqrt{(x_p^j - x_p^i)^2 + (y_p^j - y_p^i)^2}$$

- Select fit minimizing the sum of distances

Centroid fit selection

Selects the fitting where most initializations converge, ignoring those stuck in a local minima.



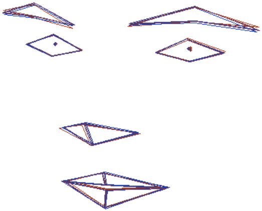
- Multiple initializations at different rotation angles
- Fit all initializations to the image
- Calculate euclidean distance between each pair of fits

$$d(L^i, L^j) = \sum_{p=1}^{n} \sqrt{(x_p^j - x_p^i)^2 + (y_p^j - y_p^i)^2}$$

- Select fit minimizing the sum of distances

Centroid fit selection

Selects the fitting where most initializations converge, ignoring those stuck in a local minima.



- Multiple initializations at different rotation angles
- Fit all initializations to the image
- Calculate euclidean distance between each pair of fits

$$d(L^i, L^j) = \sum_{p=1}^{n} \sqrt{(x_p^j - x_p^i)^2 + (y_p^j - y_p^i)^2}$$

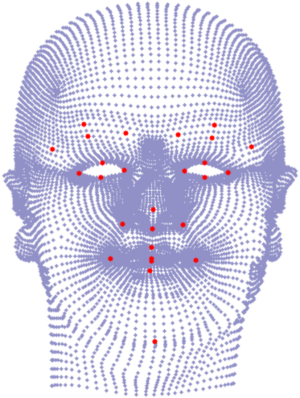
- Select fit minimizing the sum of distances

Outline

3D models

- 3D alignment
- 3D regression

3D Active Shape Model



Facewarehouse dataset:

- 3D face range scans
- 150 individuals, 20 facial expressions
- Semi-automated landmark selection

ASM features:

- 3 coordinates per landmark
- Small increase of PCA bases
  - AFLW: from 21 to 23
  - LFPW: from 23 to 26

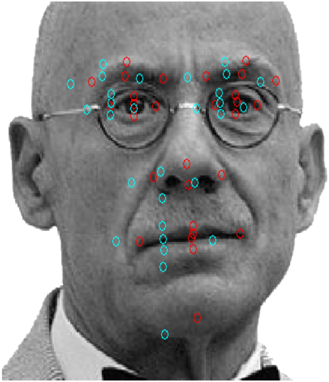
Restricted camera model

Find the best 3D rotation, scaling and translation adjusting a a 3D shape to a 2D one through an ortographic projection.

- Single scaling factor
- Three rotation angles: roll, pitch, yaw
- Closed-form solution
  - Find projection matrix (least-squares)
  - Extrapolate third dimension (cross-product)
  - Force dimensions orthogonality and equal scaling at each dimension (QR decomposition)

Iterative alignment

Iteratively adjust restricted camera model parameters and 3D ASM weights until convergence.



- Initialize to mean 3D shape
- Align to 2D shape
- While not converged
  - Find best ASM weights
  - Re-align shape to 2D landmarks

Iterative alignment

Iteratively adjust restricted camera model parameters and 3D ASM weights until convergence.

- Initialize to mean 3D shape
- Align to 2D shape
- While not converged
  - Find best ASM weights
  - Re-align shape to 2D landmarks

Iterative alignment

Iteratively adjust restricted camera model parameters and 3D ASM weights until convergence.

- Initialize to mean 3D shape
- Align to 2D shape
- While not converged
  - Find best ASM weights
  - Re-align shape to 2D landmarks

Iterative alignment

Iteratively adjust restricted camera model parameters and 3D ASM weights until convergence.

- Initialize to mean 3D shape
- Align to 2D shape
- While not converged
  - Find best ASM weights
  - Re-align shape to 2D landmarks

Extended parametric representation

$$p = \langle b, s, d_x, d_y, \theta, \gamma, \eta \rangle$$

Extend the parametric representation to include the restricted camera model parameters and 3D ASM weights.

$$L^{2D} = [P \cdot (s \cdot R_\eta \cdot R_\gamma \cdot R_\theta)] \cdot L^{3D'} + T = R \cdot L^{3D'} + T$$

Outline

Datasets

- Face alignment datasets
- Head pose recovery dataset

AFLW and LFPW datasets

AFLW

2359 images, 21 landmarks
Mostly frontal poses

Two in-the-wild datasets with images at different resolutions, qualities, poses and for different ages, genders and ethnicities.

LFPW

836 images, 29 landmarks
Wide range of poses

Pointing '04 dataset

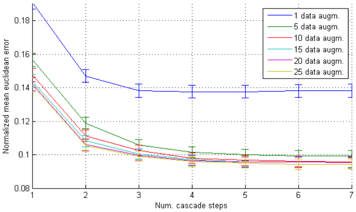
Dataset created in a controlled environment, labelled with the pitch and yaw pose angles, but not with geometric information.

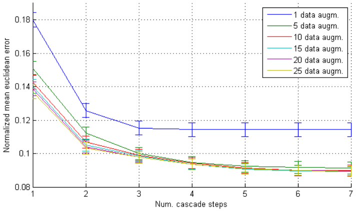
- 15 individuals
- 2 series per person
- 93 poses per serie
- Discrete angles, 15 degrees apart

Outline

2D methods: Parameter selection

- Lower tendency to overfit with more data augmentations
- SDM converges faster in both datasets





AFLW

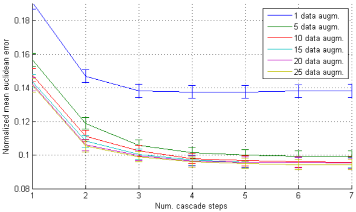
Smaller effect of data augmentation

Converges faster

Bigger average NMED error

2D methods: Parameter selection

- Lower tendency to overfit with more data augmentations
- SDM converges faster in both datasets

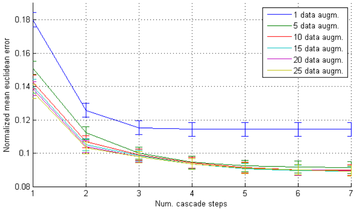


AFLW

Smaller effect of data augmentation

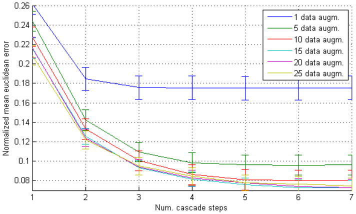
Converges faster

Bigger average NMED error



2D methods: Parameter selection

- Lower tendency to overfit with more data augmentations
- SDM converges faster in both datasets

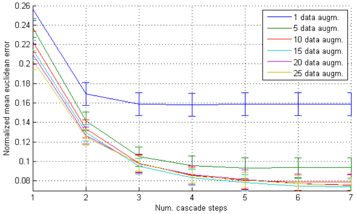


LFPW

Bigger effect of data augmentation

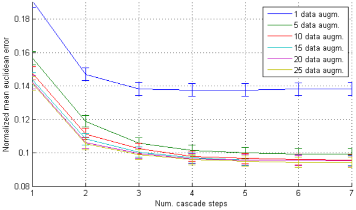
Converges slowlier

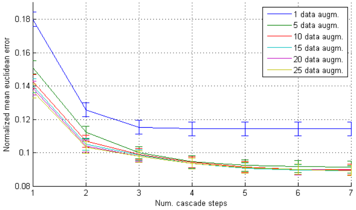
Smaller average NMED error



2D methods: Parameter selection

- Lower tendency to overfit with more data augmentations
- SDM converges faster in both datasets





AFLW

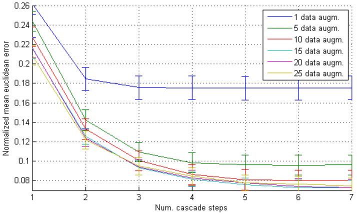
Smaller effect of data augmentation

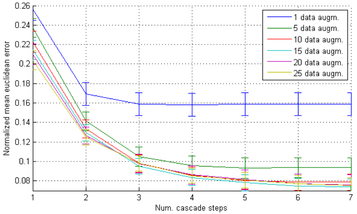
Converges faster

Bigger average NMED error

2D methods: Parameter selection

- Lower tendency to overfit with more data augmentations
- SDM converges faster in both datasets

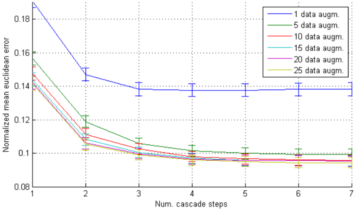




LFPW
Bigger effect of data augmentation
Converges slowlier
Smaller average NMED error

2D methods: Parameter selection

- Lower tendency to overfit with more data augmentations
- SDM converges faster in both datasets

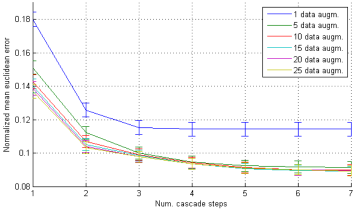


AFLW

Smaller effect of data augmentation

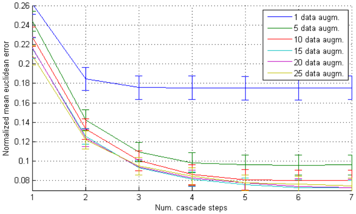
Converges faster

Bigger average NMED error



2D methods: Parameter selection

- Lower tendency to overfit with more data augmentations
- SDM converges faster in both datasets

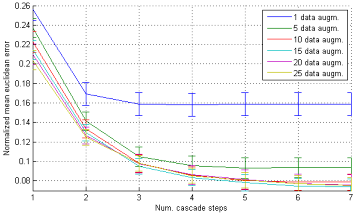


LFPW

Bigger effect of data augmentation

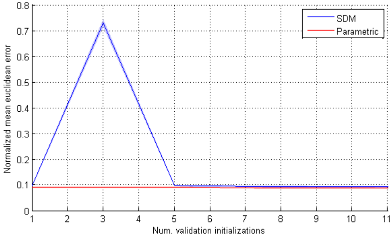
Converges slowlier

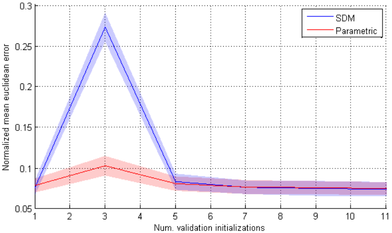
Smaller average NMED error



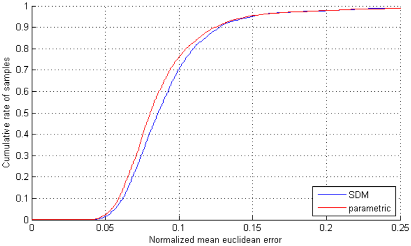
2D methods: Parameter selection

	AFLW		LFPW	
	SDM	Parametric	SDM	Parametric
Data augmentations	6	6	6	6
Cascade steps	25	15	15	15
Initializations	1	1	1	1





2D methods: Shape alignment accuracies



AFLW

Minimum NMED error of 0.05

Error below 0.15 for 95% of cases

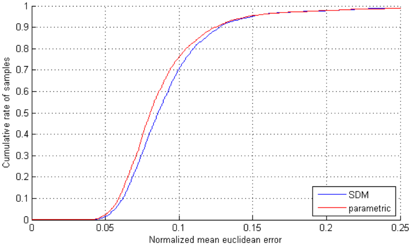
Better accuracy for parametric

SDM runtime: 13ms

Parametric runtime: 13ms

	SDM	Parametric	p-value	Independence
AFLW	0.0973 ± 0.0032	0.0928 ± 0.0031	$1.7 \cdot 10^{-10}$	yes
LFPW	0.0763 ± 0.0090	0.0774 ± 0.0085	0.2671	no

2D methods: Shape alignment accuracies



AFLW

Minimum NMED error of 0.05

Error below 0.15 for 95% of cases

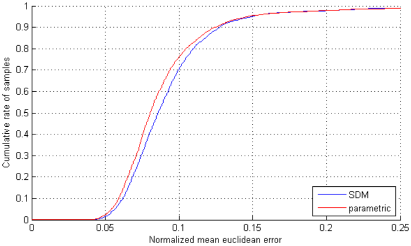
Better accuracy for parametric

SDM runtime: 13ms

Parametric runtime: 13ms

	SDM	Parametric	p-value	Independence
AFLW	0.0973 ± 0.0032	0.0928 ± 0.0031	$1.7 \cdot 10^{-10}$	yes
LFPW	0.0763 ± 0.0090	0.0774 ± 0.0085	0.2671	no

2D methods: Shape alignment accuracies



AFLW

Minimum NMED error of 0.05

Error below 0.15 for 95% of cases

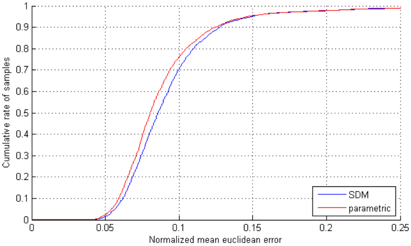
Better accuracy for parametric

SDM runtime: 13ms

Parametric runtime: 13ms

	SDM	Parametric	p-value	Independence
AFLW	0.0973 ± 0.0032	0.0928 ± 0.0031	$1.7 \cdot 10^{-10}$	yes
LFPW	0.0763 ± 0.0090	0.0774 ± 0.0085	0.2671	no

2D methods: Shape alignment accuracies



AFLW

Minimum NMED error of 0.05

Error below 0.15 for 95% of cases

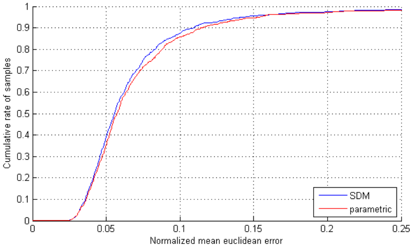
Better accuracy for parametric

SDM runtime: 13ms

Parametric runtime: 13ms

	SDM	Parametric	p-value	Independence
AFLW	0.0973 ± 0.0032	0.0928 ± 0.0031	$1.7 \cdot 10^{-10}$	yes
LFPW	0.0763 ± 0.0090	0.0774 ± 0.0085	0.2671	no

2D methods: Shape alignment accuracies



LFPW

Minimum NMED error of 0.03

Error below 0.15 for 95% of cases

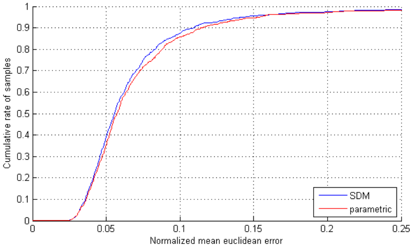
Better accuracy for SDM

SDM runtime: 20ms

Parametric runtime: 20ms

	SDM	Parametric	p-value	Independence
AFLW	0.0973 ± 0.0032	0.0928 ± 0.0031	$1.7 \cdot 10^{-10}$	yes
LFPW	0.0763 ± 0.0090	0.0774 ± 0.0085	0.2671	no

2D methods: Shape alignment accuracies



LFPW

Minimum NMED error of 0.03

Error below 0.15 for 95% of cases

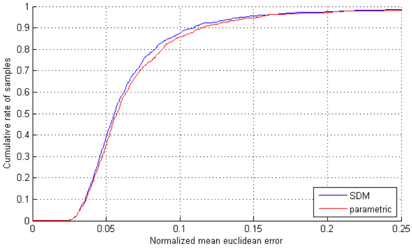
Better accuracy for SDM

SDM runtime: 20ms

Parametric runtime: 20ms

	SDM	Parametric	p-value	Independence
AFLW	0.0973 ± 0.0032	0.0928 ± 0.0031	$1.7 \cdot 10^{-10}$	yes
LFPW	0.0763 ± 0.0090	0.0774 ± 0.0085	0.2671	no

2D methods: Shape alignment accuracies



LFPW

Minimum NMED error of 0.03

Error below 0.15 for 95% of cases

Better accuracy for SDM

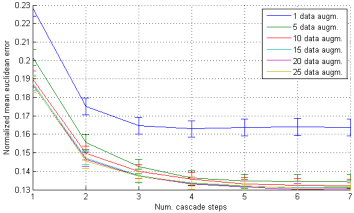
SDM runtime: 20ms

Parametric runtime: 20ms

	SDM	Parametric	p-value	Independence
AFLW	0.0973 ± 0.0032	0.0928 ± 0.0031	$1.7 \cdot 10^{-10}$	yes
LFPW	0.0763 ± 0.0090	0.0774 ± 0.0085	0.2671	no

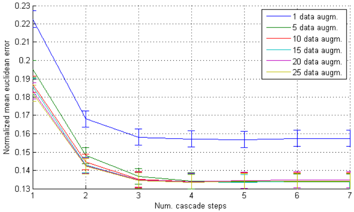
3D methods: Parameter selection

- 3D regression always converges faster
- No effect of data augmentation after 5 augments



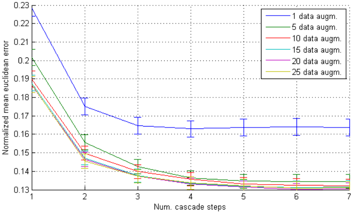
AFLW

Smaller effect of data augmentation
Converges faster
Bigger average NMED error



3D methods: Parameter selection

- 3D regression always converges faster
- No effect of data augmentation after 5 augments

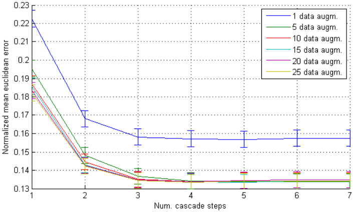


AFLW

Smaller effect of data augmentation

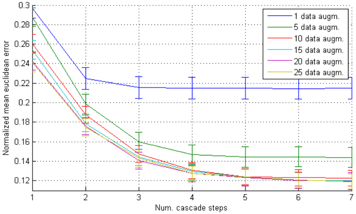
Converges faster

Bigger average NMED error



3D methods: Parameter selection

- 3D regression always converges faster
- No effect of data augmentation after 5 augments

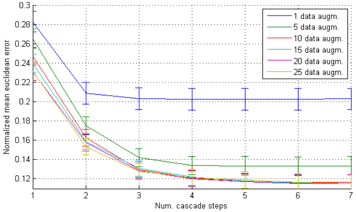


LFPW

Bigger effect of data augmentation

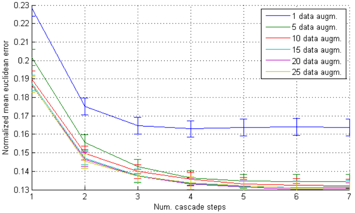
Converges slowlier

Smaller average NMED error



3D methods: Parameter selection

- 3D regression always converges faster
- No effect of data augmentation after 5 augments

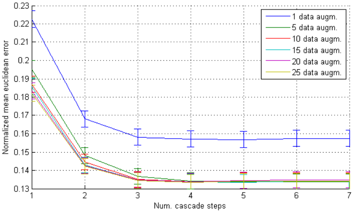


AFLW

Smaller effect of data augmentation

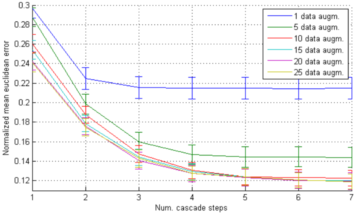
Converges faster

Bigger average NMED error



3D methods: Parameter selection

- 3D regression always converges faster
- No effect of data augmentation after 5 augments

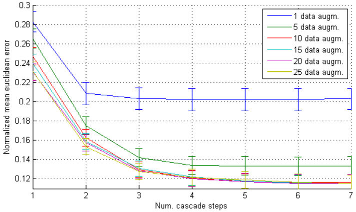


LFPW

Bigger effect of data augmentation

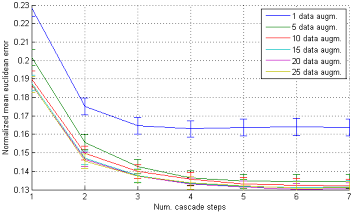
Converges slowlier

Smaller average NMED error



3D methods: Parameter selection

- 3D regression always converges faster
- No effect of data augmentation after 5 augments

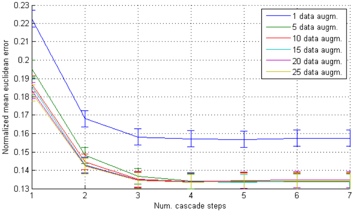


AFLW

Smaller effect of data augmentation

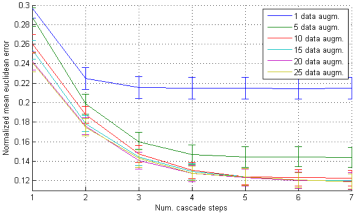
Converges faster

Bigger average NMED error



3D methods: Parameter selection

- 3D regression always converges faster
- No effect of data augmentation after 5 augments

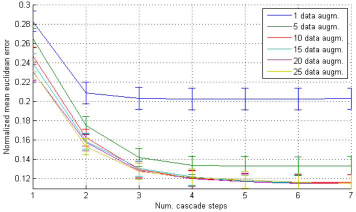


LFPW

Bigger effect of data augmentation

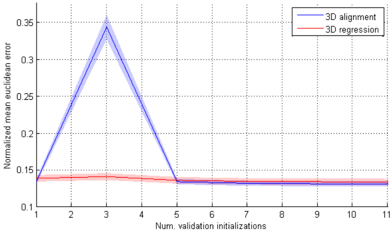
Converges slowlier

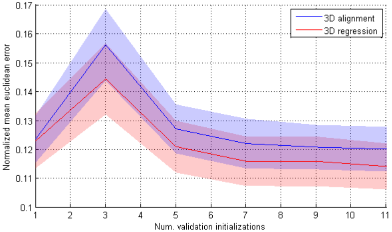
Smaller average NMED error



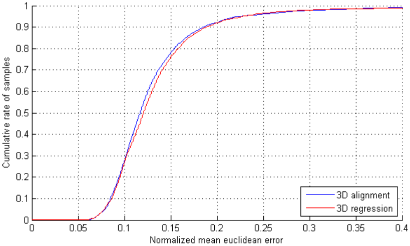
3D methods: Parameter selection

	AFLW		LFPW	
	3D alig.	3D regr.	3D alig.	3D regr.
Data augmentations	6	5	6	6
Cascade steps	20	15	20	20
Initializations	1	1	1	1





3D methods: Shape alignment accuracies



AFLW

Minimum NMED error of 0.06

Error below 0.23 for 95% of cases

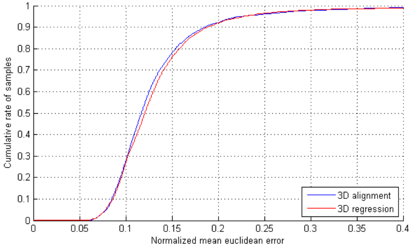
Better accuracy for 3D alignment

3D alignment runtime: 27ms

3D regression runtime: 11ms

	3D alignment	3D regression	p-value	Independence
AFLW	0.1340 ± 0.0036	0.1387 ± 0.0045	0.0134	yes
LFPW	0.1235 ± 0.0081	0.1229 ± 0.0094	0.3412	no

3D methods: Shape alignment accuracies



AFLW

Minimum NMED error of 0.06

Error below 0.23 for 95% of cases

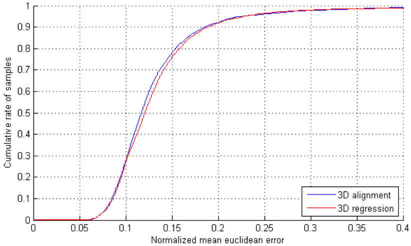
Better accuracy for 3D alignment

3D alignment runtime: 27ms

3D regression runtime: 11ms

	3D alignment	3D regression	p-value	Independence
AFLW	0.1340 ± 0.0036	0.1387 ± 0.0045	0.0134	yes
LFPW	0.1235 ± 0.0081	0.1229 ± 0.0094	0.3412	no

3D methods: Shape alignment accuracies



AFLW

Minimum NMED error of 0.06

Error below 0.23 for 95% of cases

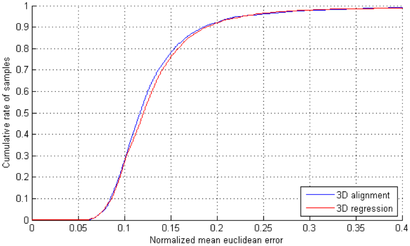
Better accuracy for 3D alignment

3D alignment runtime: 27ms

3D regression runtime: 11ms

	3D alignment	3D regression	p-value	Independence
AFLW	0.1340 ± 0.0036	0.1387 ± 0.0045	0.0134	yes
LFPW	0.1235 ± 0.0081	0.1229 ± 0.0094	0.3412	no

3D methods: Shape alignment accuracies



AFLW

Minimum NMED error of 0.06

Error below 0.23 for 95% of cases

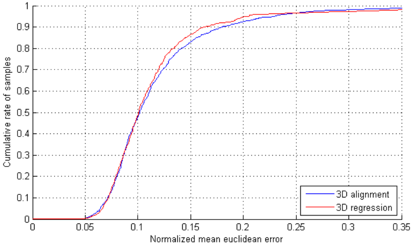
Better accuracy for 3D alignment

3D alignment runtime: 27ms

3D regression runtime: 11ms

	3D alignment	3D regression	p-value	Independence
AFLW	0.1340 ± 0.0036	0.1387 ± 0.0045	0.0134	yes
LFPW	0.1235 ± 0.0081	0.1229 ± 0.0094	0.3412	no

3D methods: Shape alignment accuracies



LFPW

Minimum NMED error of 0.05

Error below 0.23 for 95% of cases

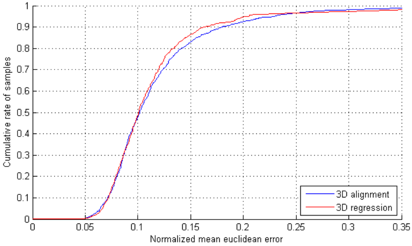
Better accuracy for 3D regression

3D alignment runtime: 34ms

3D regression runtime: 16ms

	3D alignment	3D regression	p-value	Independence
AFLW	0.1340 ± 0.0036	0.1387 ± 0.0045	0.0134	yes
LFPW	0.1235 ± 0.0081	0.1229 ± 0.0094	0.3412	no

3D methods: Shape alignment accuracies



LFPW

Minimum NMED error of 0.05

Error below 0.23 for 95% of cases

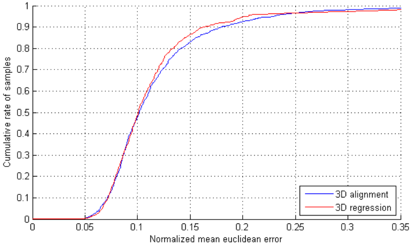
Better accuracy for 3D regression

3D alignment runtime: 34ms

3D regression runtime: 16ms

	3D alignment	3D regression	p-value	Independence
AFLW	0.1340 ± 0.0036	0.1387 ± 0.0045	0.0134	yes
LFPW	0.1235 ± 0.0081	0.1229 ± 0.0094	0.3412	no

3D methods: Shape alignment accuracies



LFPW

Minimum NMED error of 0.05

Error below 0.23 for 95% of cases

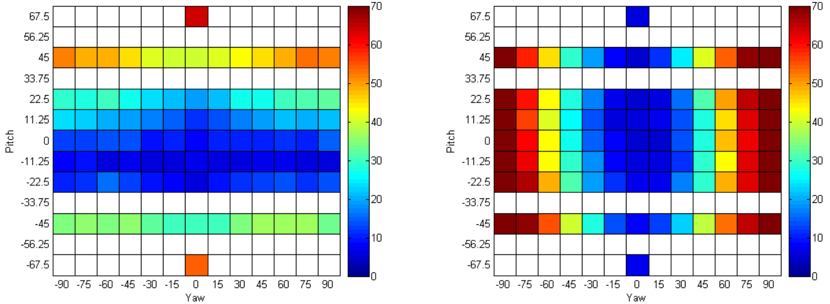
Better accuracy for 3D regression

3D alignment runtime: 34ms

3D regression runtime: 16ms

	3D alignment	3D regression	p-value	Independence
AFLW	0.1340 ± 0.0036	0.1387 ± 0.0045	0.0134	yes
LFPW	0.1235 ± 0.0081	0.1229 ± 0.0094	0.3412	no

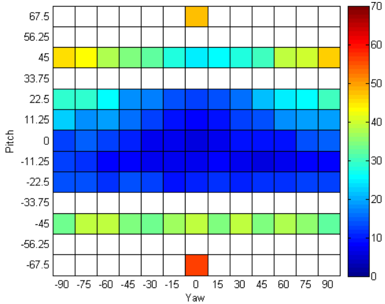
3D methods: Pose recovery error distribution

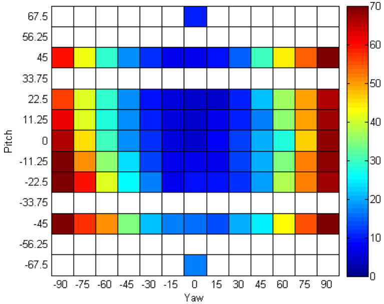


Pitch 3D alignment Yaw

Both 3D methods are good only at predicting small variations from the frontal pose, with 3D regression being able to predict a wider range of poses.

3D methods: Pose recovery error distribution





Pitch 3D regression Yaw

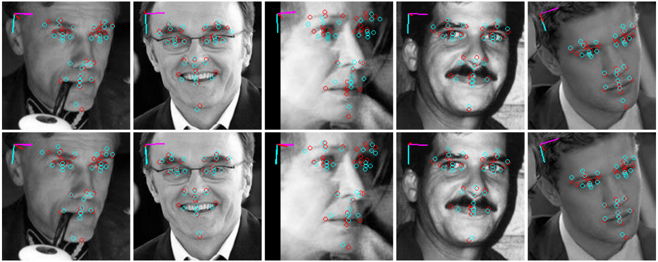
Both 3D methods are good only at predicting small variations from the frontal pose, with 3D regression being able to predict a wider range of poses.

Qualitative evaluation

2D methods (over AFLW)

Parametric approach has better accuracy for low quality images, as well as better locating the face countour.

Qualitative evaluation



3D methods (over LFPW)

3D regression obtains a better pose estimate. 3D alignment fails to accurately predict the landmarks giving a visual cue for the pose in exchange of increasing the overall shape alignment accuracy.

Outline

Conclusions

Parametric approach

- Generalizes better when enough data is available
- More robust to local minima
- ASM model may restrict valid shape deformations

3D regression

- Better prediction for both pitch and yaw
- Much faster than 3D alignment
- Slightly worse shape alignment accuracy

Questions?