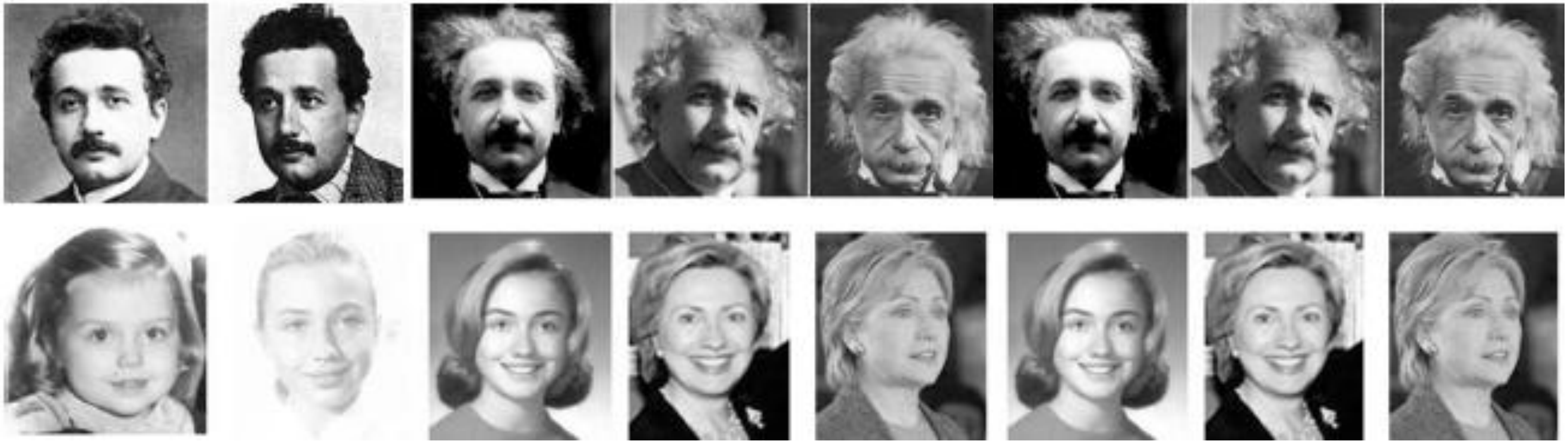
Karlsruhe Institute of Technology

Age Estimation

Narine Kokhlikyan | January 19, 2012

Human Age Estimation Based On Changes Of Facial Appearance

FACIAL IMAGE PROCESSING AND ANALYSIS, INSTITUTE FOR ANTHROPOMATICS, FACULTY OF INFORMATICS



Overview

- Introduction
- Problems
- Approaches
  - AGES
  - LLD
- Experimental Evaluation
- Conclusion and Disscussion

Human Aging Process …

... leads to remarkable changes of human facial appearance

Why facial age estimation ?

- Directly inferred from facial appearance
- Real-world applications
  - Age Specific HCI
  - Children Protection
  - Security Control and Surveillance Monitoring
  - Multi-cue identification

How old is this man on the picture in the middle ?

Age ?

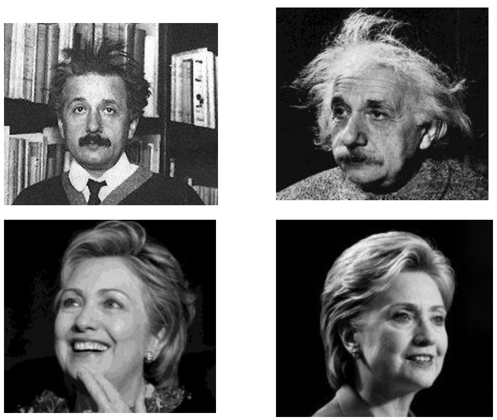
How old is this man on the picture in the middle ?

51

Problems

- Age estimation is even difficult for human
- Different people age differently

40 61



- Limited number of aging images

Related Work

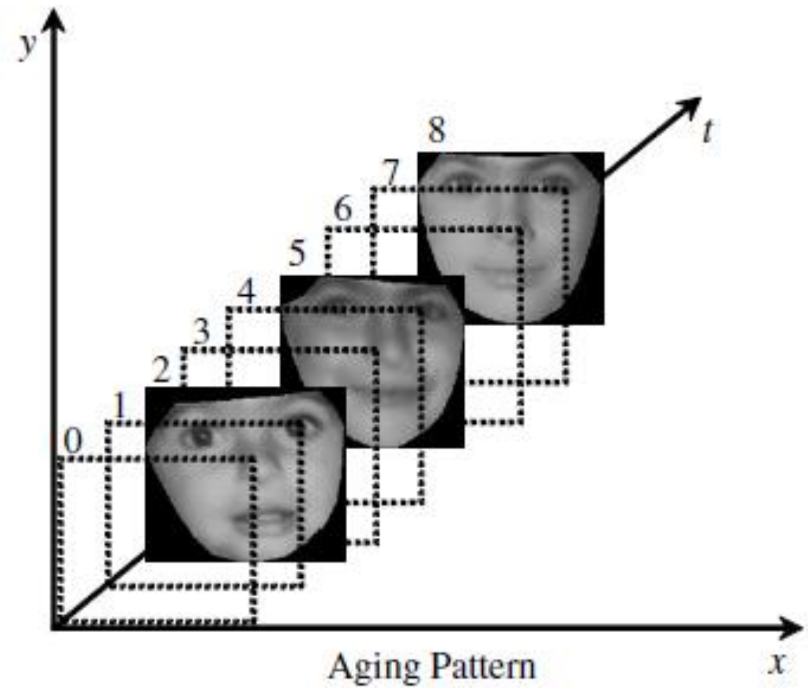
- Anthropometric approach
  - Based on measured sizes and proportions on human faces
  - Considers only the facial geometry but not texture
  - Face image classification in 3 groups (babies, young adults and seniors)

- Aging function
  - Considers both shape and structure
  - Deals with any age
  - Does not consider personal and temporal characteristics of aging

Aging Pattern

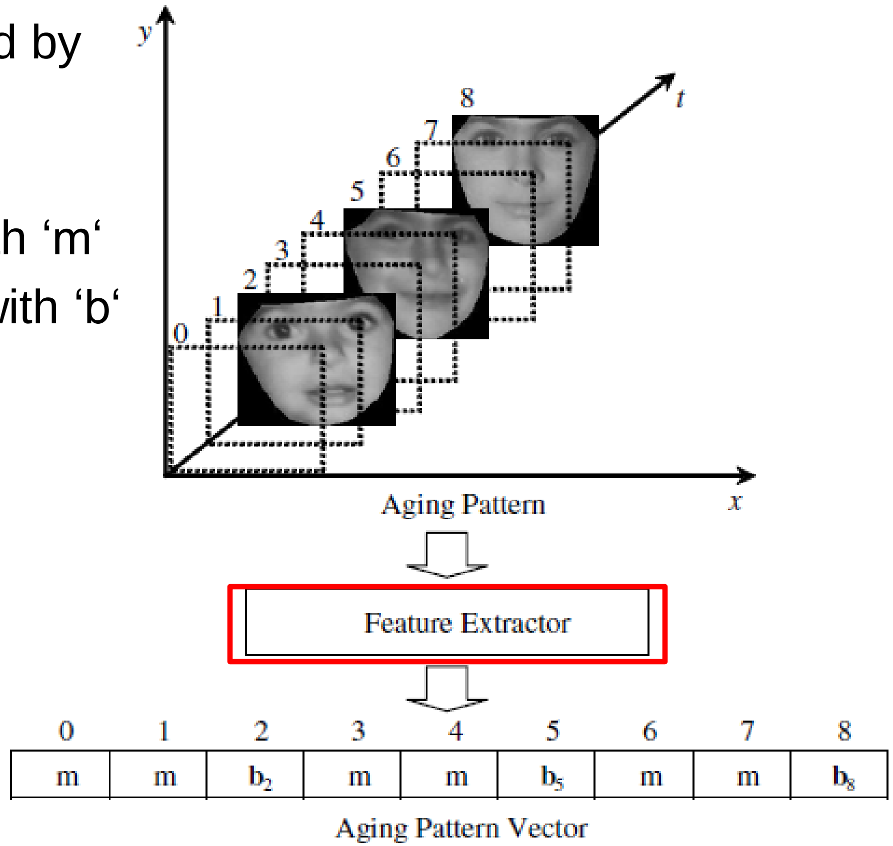
- *Definition 1.* An aging pattern is a sequence of personal face images sorted in time order

- All face images come from the same person
- And are arranged by time



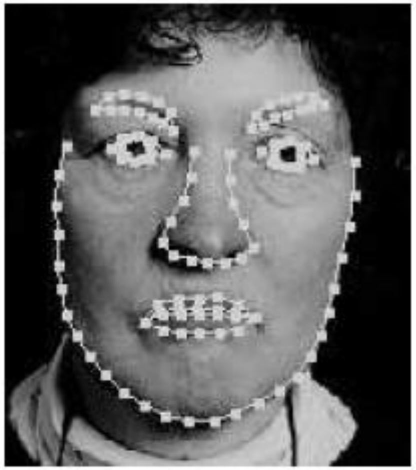
Aging Pattern(2)

- Feature Extraction
  - Images are transformed into feature vector
  - Feature vectors are extracted by Active Appearance Model
  - Missing parts are marked with ‘m‘
  - Available parts are marked with ‘b‘



Feature Extraction

- Active Appearance Model
  - Requires training set of labeled images
  - Appearance model relies on landmark points
  - Combines shape and texture variations



$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c}$$

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c}$$

$\mathbf{c}$ - Vector of appearance parameter

$\bar{\mathbf{x}}$ - The mean shape

$\bar{\mathbf{g}}$ - The mean texture

$\mathbf{Q}_s, \mathbf{Q}_g$ - The Modes of variations

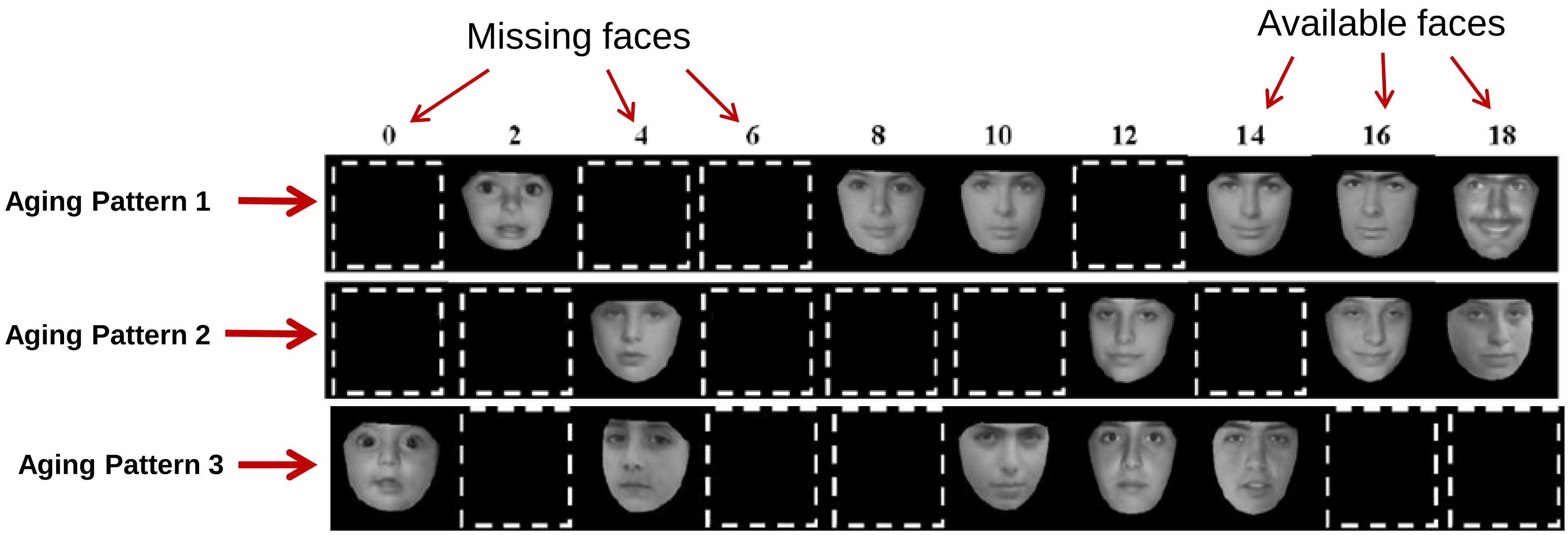
0	1	2	3	4	5	6	7	8
m	m	b_2	m	m	b_5	m	m	b_8

Aging Pattern Vector

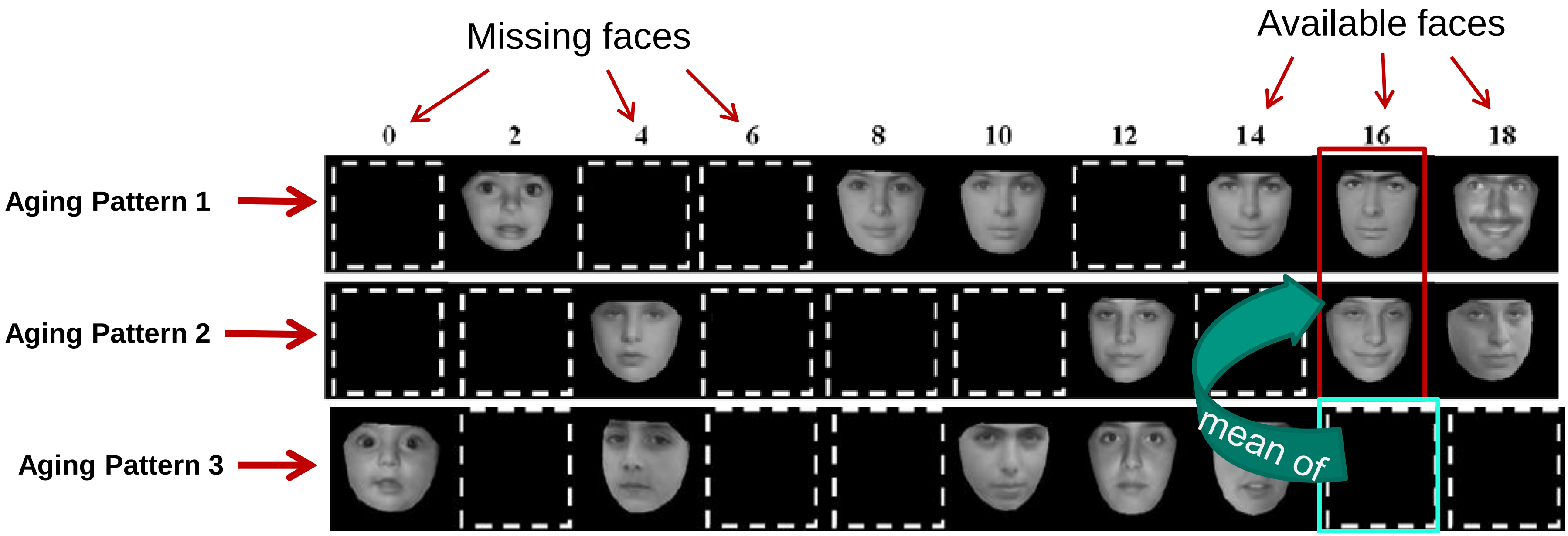
The AGES Algorithm

- AGES - Learning
- Age Estimation

AGES – Learning



AGES – Missing Faces - Initialization



Missing Faces → Initialization

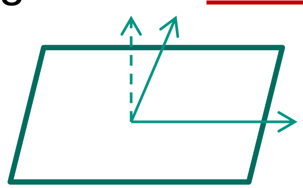
AGES – Learning Algorithm

Missing Faces→ Initialization

Aging Pattern

PCA

Aging Pattern Subspace



W, μ

AGES – Learning Algorithm

Missing Faces→ Initialization

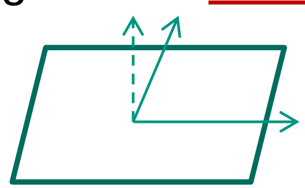
Aging Pattern

PCA

Projection

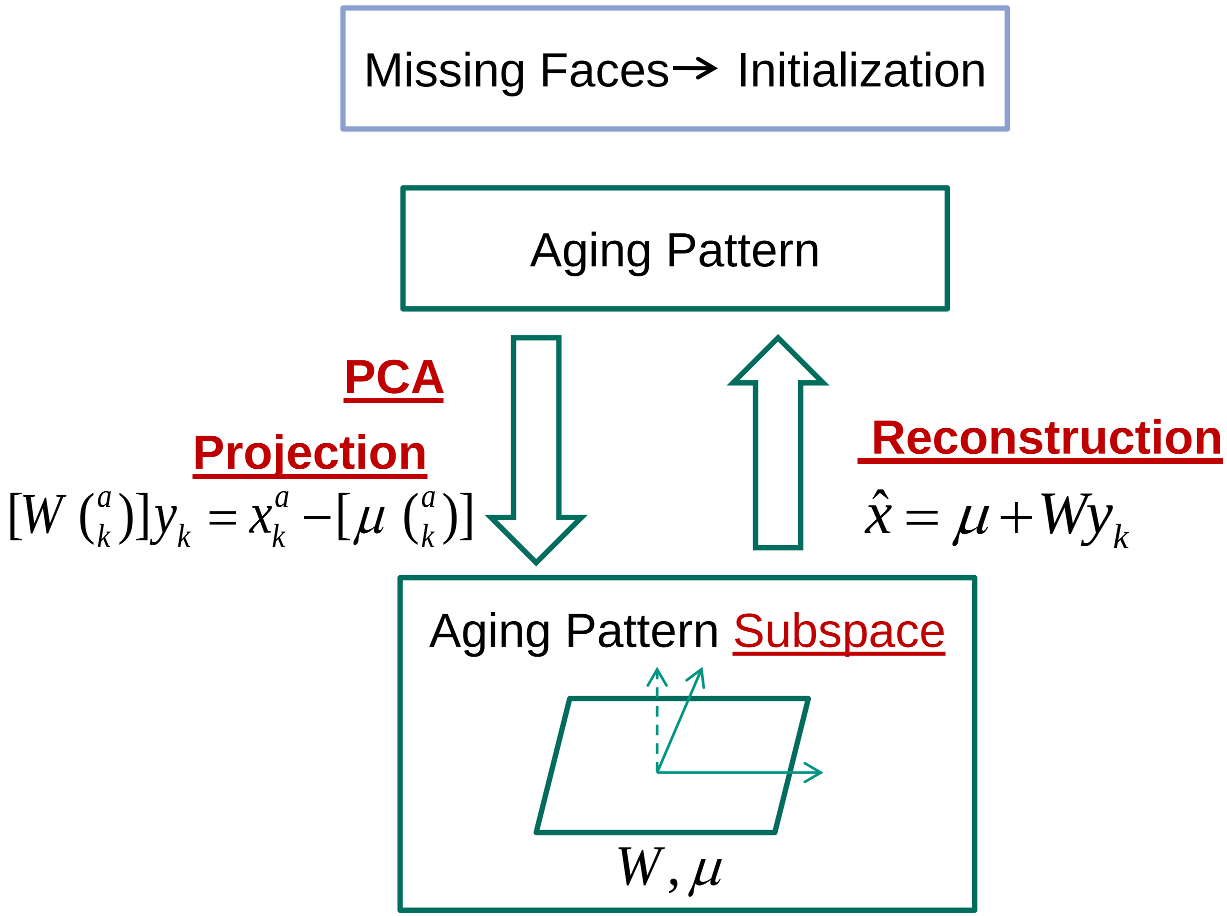
$$[W\binom{a}{k}]y_k = x_k^a - [\mu\binom{a}{k}]$$

Aging Pattern Subspace



W, μ

AGES – Learning Algorithm



AGES – Learning Algorithm

Missing Faces→ Initialization

Aging Pattern

PCA

Projection

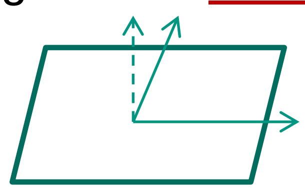
$$[W(^{a}_{k})]y_k = x^{a}_{k} - [\mu(^{a}_{k})]$$

$$x_k^m \leftarrow \hat{x}_k^m$$

Reconstruction

$$\hat{x} = \mu + W y_k$$

Aging Pattern Subspace



W, μ

AGES – Learning Algorithm

Missing Faces→ Initialization

Aging Pattern

PCA

Projection

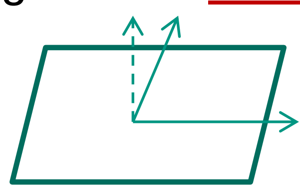
$$[W\tbinom{a}{k}]y_k = x_k^a - [\mu\tbinom{a}{k}]$$

$$x_k^m \leftarrow \hat{x}_k^m$$

Reconstruction

$$\hat{x} = \mu + W y_k$$

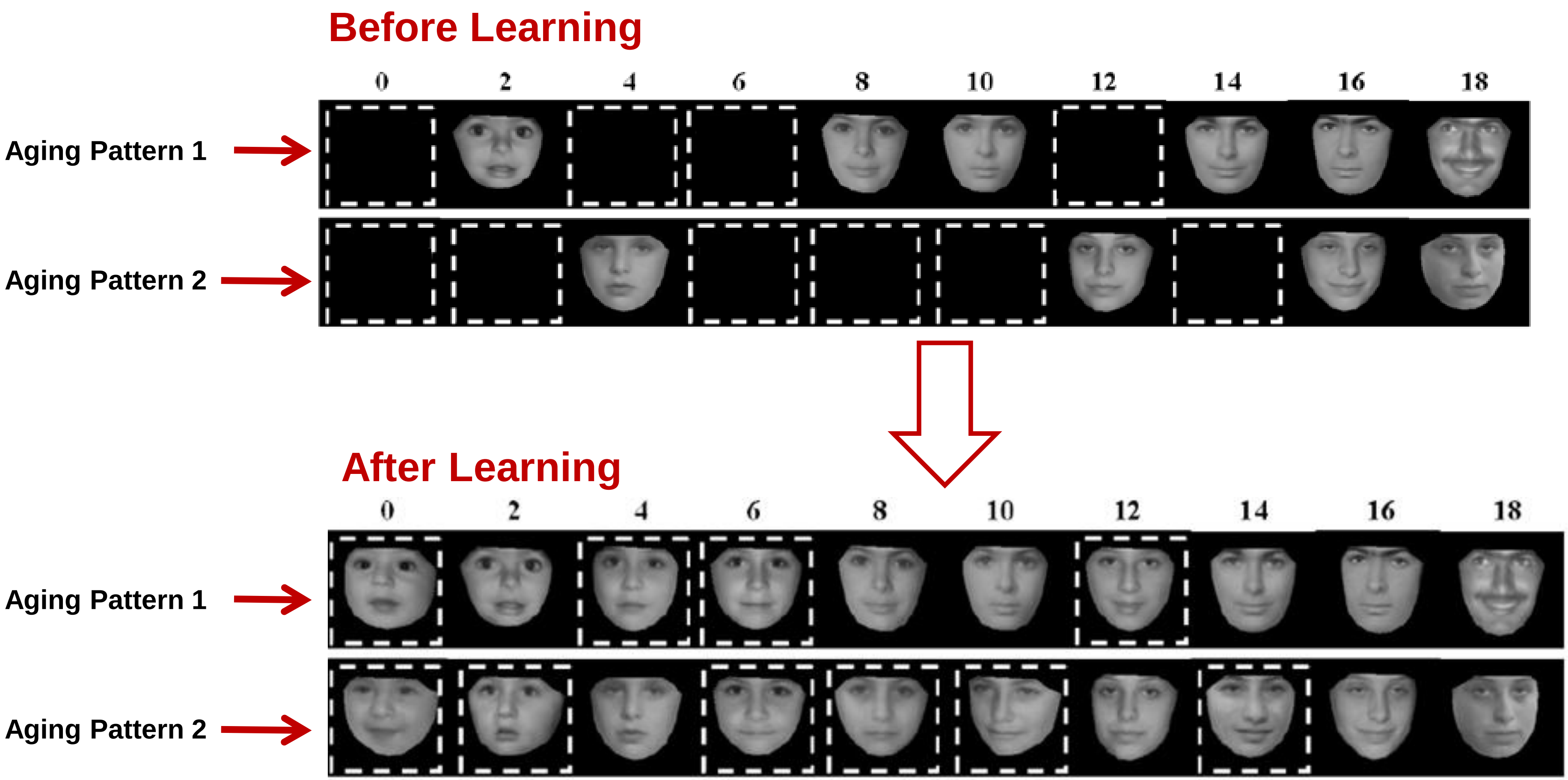
Aging Pattern Subspace



$$W, \mu$$

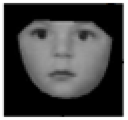
Minimizing Reconstruction Error

AGES - Learning



Age Estimation

New Test Image



Introduction

AGES

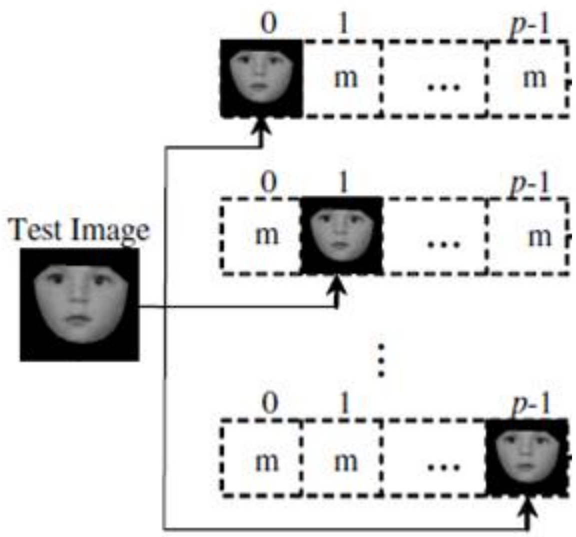
LLD

Experiments

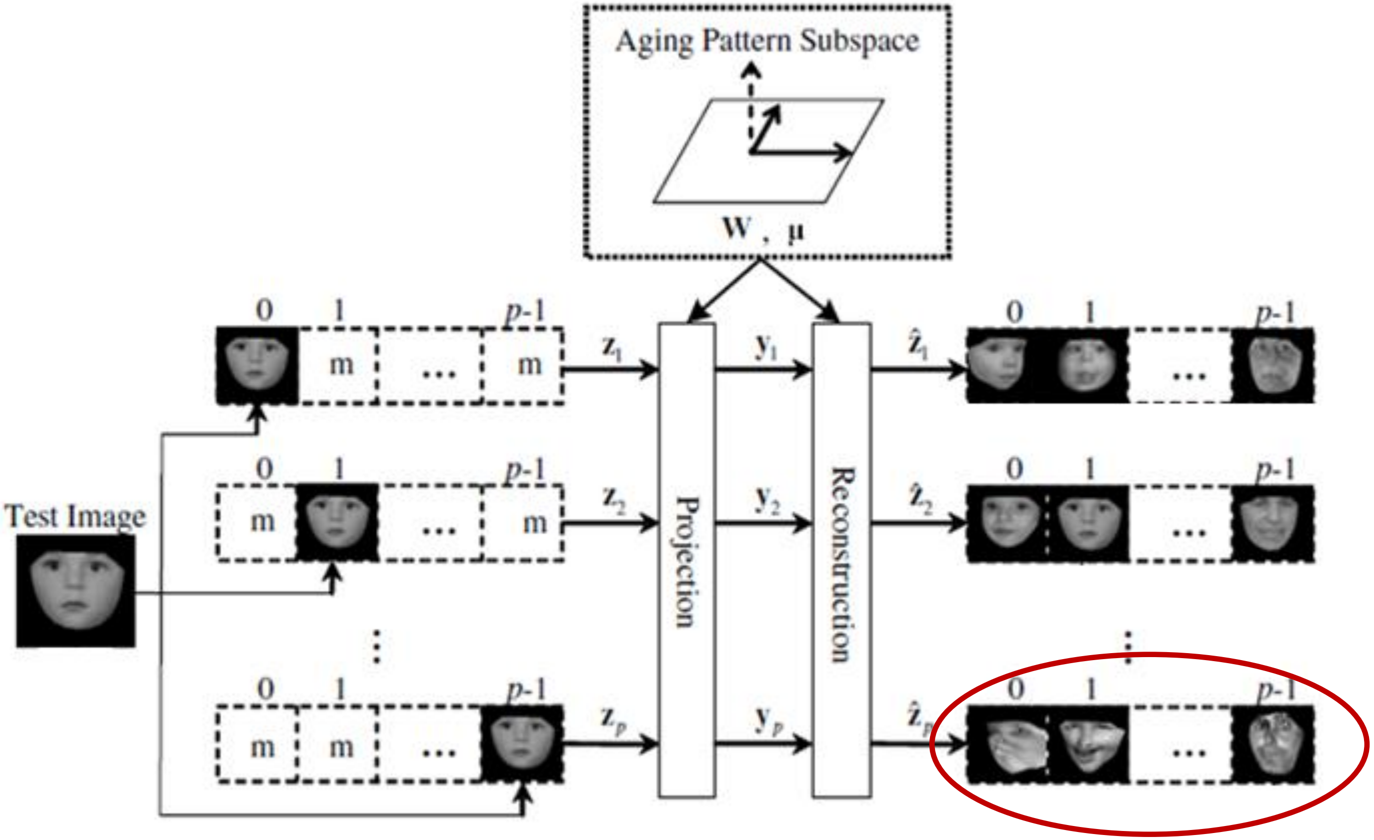
Conclusion

Age Estimation(2)

Generating aging patterns



Age Estimation(3)



Introduction

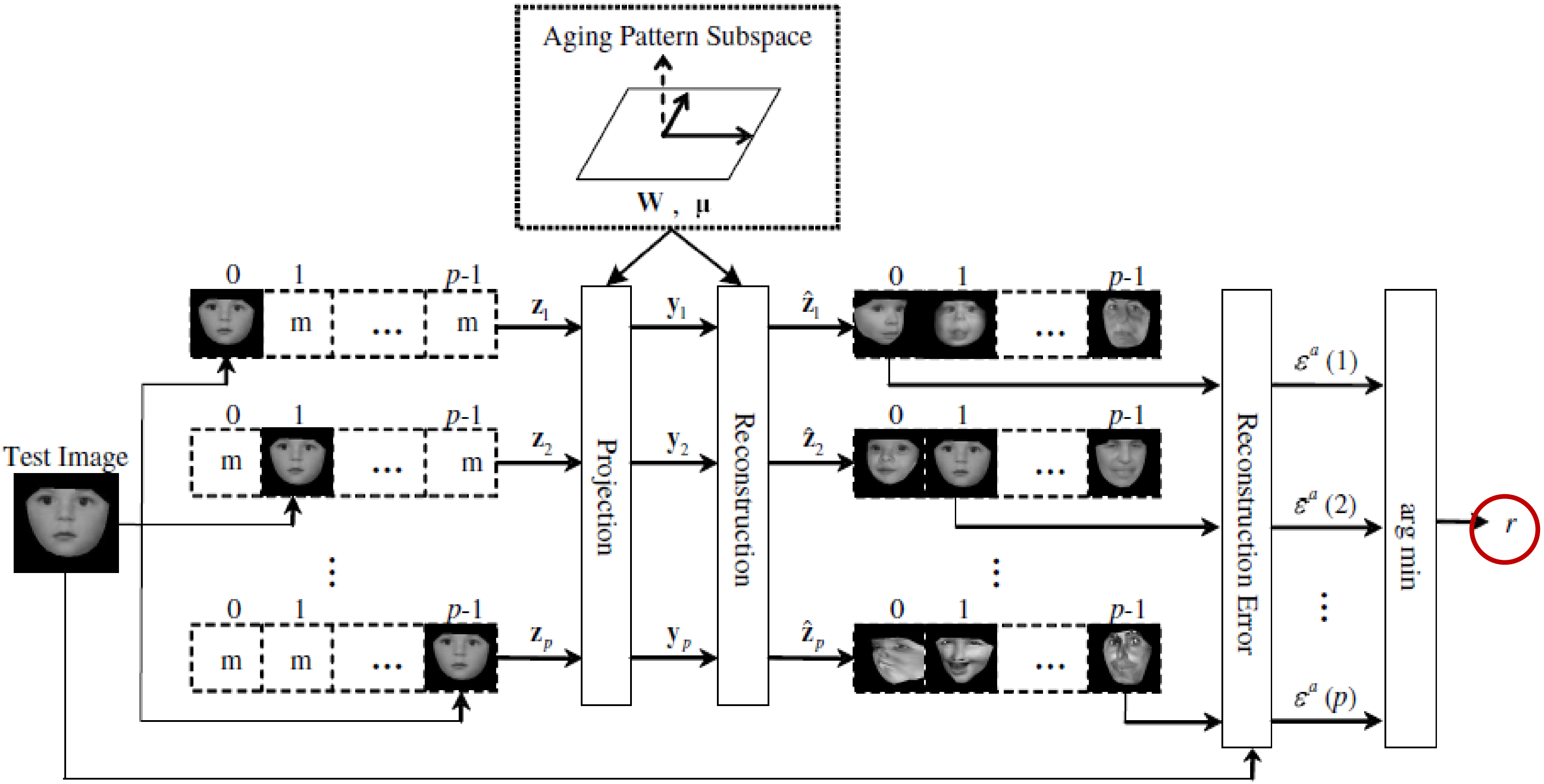
AGES

LLD

Experiments

Conclusion

Age Estimation(4)



AGES - Challenges

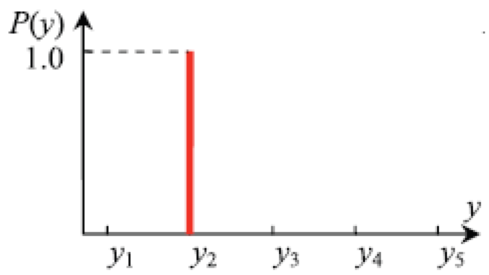
- Lack of sufficient training data
  - The images at the higher ages are especially rare
- Many missing values in aging pattern vector

LLD – Learning from Label Distribution

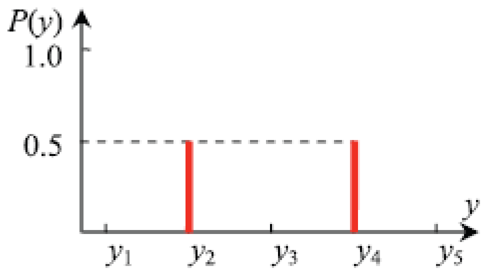
- Relieves the problem of insufficient training samples
- Additional knowledge → **Close ages look quite similar**
- Label distribution rather than a single label for each image

Different cases of label distribution

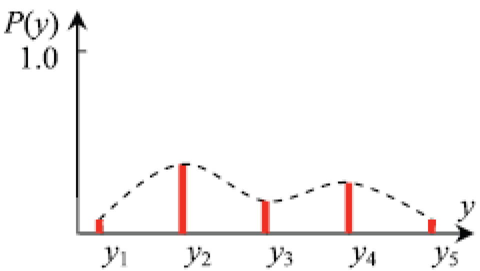
Single Label



Two Labels



General Case

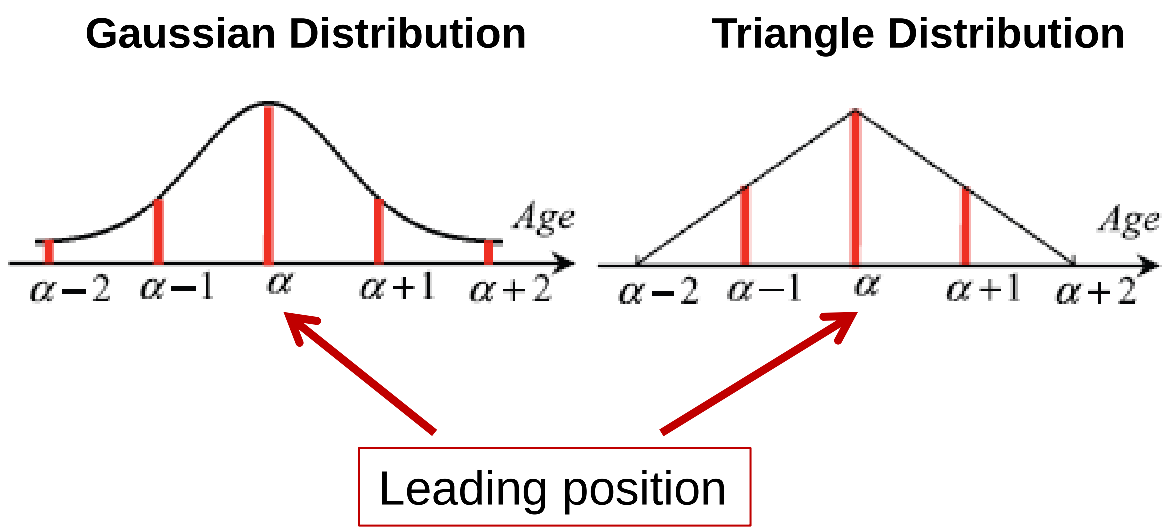


$P(y) \rightarrow$ The *proportion of* y in a full class description of the instance

$$y \rightarrow age; \ \sum_y P(y) = 1$$

Label Distribution

- Typical cases of label distribution



LLD – Learning from Label Distribution

- Input: A Trainig set

$$S = \{(\underbrace{x_1}_{\text{Instance}}, \underbrace{P_1(y)}_{\text{Label distribution}}), \ldots, (x_N, P_N(y),)\}$$

- Goal: Learn $p(y \mid x; \theta)$ as similar as possible to $p(y)$
- The features of image x are extracted by AAM

LLD – Learning from Label Distribution

- Similarity measure ⟶ Kullback-Leibler divergence

$$\theta^* = \arg\min_{\theta} D_{KL}(P_i(y) \| p(y | x_i; \theta))$$

Improved Iterative Scaling

$$\theta^i \leftarrow \theta^{i-1} + \Delta$$

Iteratively Improves θ

LLD – Age Estimation

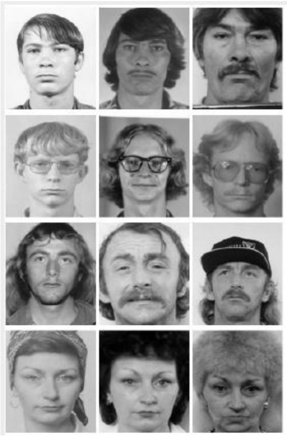
- Given a new face image x'
- Calculate $p(y \mid x')$
- If expected class label for x' is single

  $y* = \arg\max_y p(y \mid x')$
- Otherwise if multiple labels are allowed
  - A threshold is used to select multiple labels

Aging Face Databases

- FG-NET
  - Face images: 1,002
  - Subjects: 82
  - Age range: 0 to 69
  - Variations: pose, illumination, expression, etc.

- MORPH
  - Face images: 1,724
  - Subjects: 515
  - Age range: 15 to 68
  - Variations: pose, illumination, occlusion, etc.



Age Range Distribution

Age Range	FG-NET (%)	MORPH (%)
0-9	37.03	0
10-19	33.83	24.71
20-29	14.37	47.34
30-39	7.88	18.94
40-49	4.59	6.47
50-59	1.50	1.85
60-69	0.80	0.69

Compared Methods

- Human A

- Human B

51 Images from FG-NET

Evaluation Measurements

- Mean Absolute Error(MAE)

$$MAE = \sum_{k=1}^{M} | \overline{age}_k - age_k | / M$$

  - Indicates the avarage performance of the age estimator
- Cumulative Score

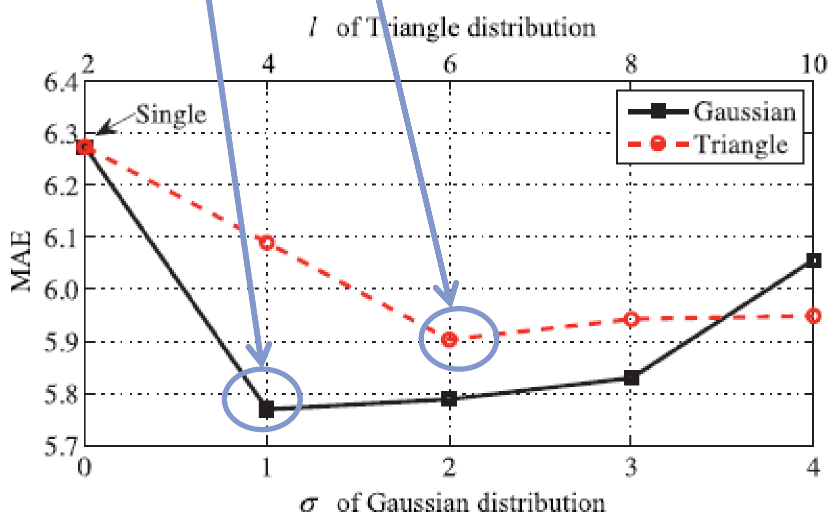
$$CumScore(l) = M_{e \le l} / M \times 100\%$$

  - Indicates the accurancy of the age estimator

Results - MAE

Method	IIS-LLD			AGES	AGES$_{lda}$	WAS	AAS	kNN	BP	C4.5	SVM	Human Observers	
	Gaussian	Triangle	Single									HumanA	HumanB
MAE	**5.77**	**5.90**	**6.27**	**6.77**	**6.22**	**8.06**	14.83	8.24	11.85	9.34	**7.25**	8.13	6.23

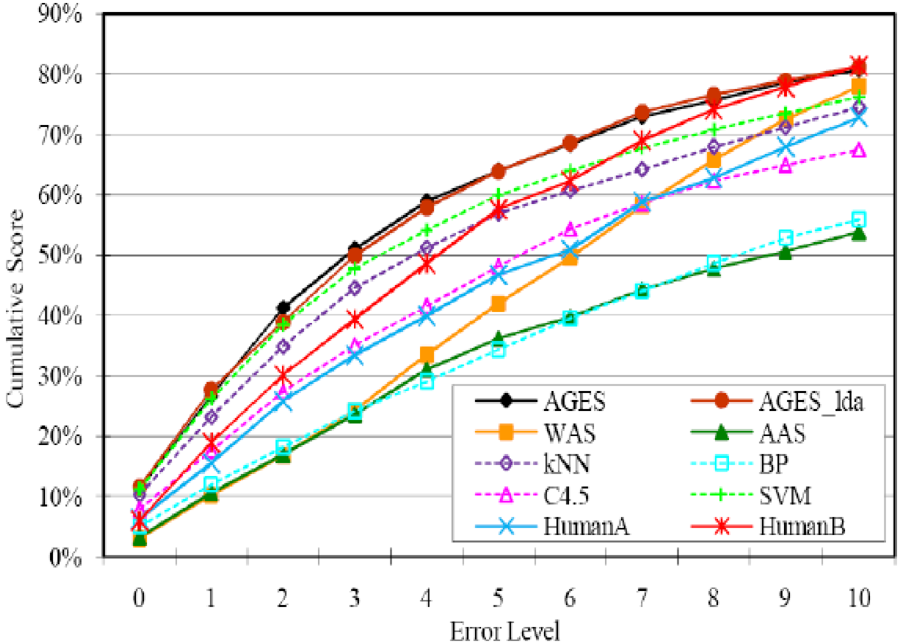
FG-NET - Leave-One-Person-Out



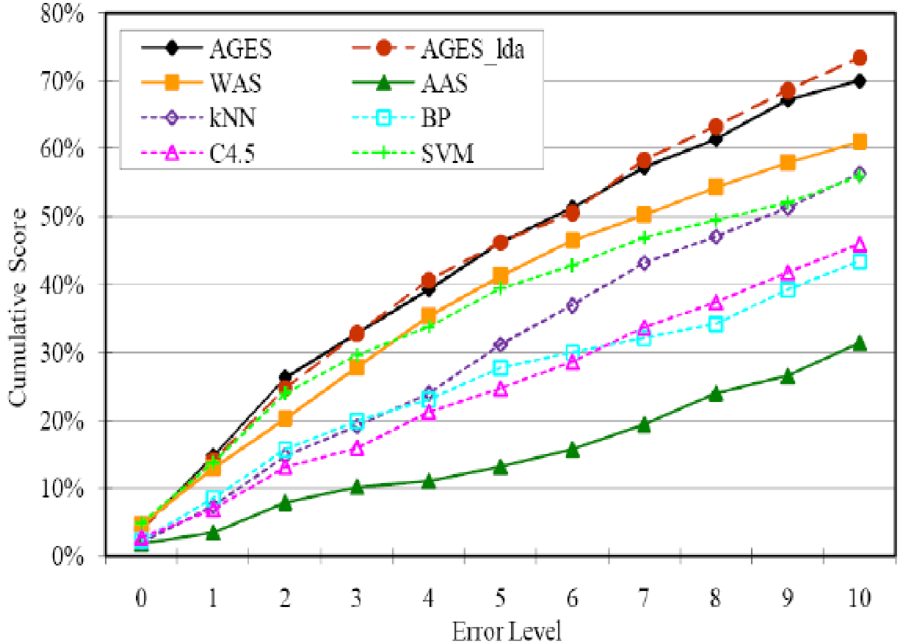
Results – MAE – Different Age Ranges

Range	#Samples	IIS-LLD			AGES
		Gaussian	Triangle	Single	
0-9	371	2.83	2.83	3.06	2.30
10-19	339	5.21	5.17	4.99	3.83
20-29	144	6.60	6.39	6.72	8.01
30-39	79	11.62	11.66	12.10	17.91
40-49	46	12.57	15.78	18.89	25.26
50-59	15	21.73	22.27	27.40	36.40
60-69	8	24.00	26.25	32.13	45.63

Results – Cumulative Score



FG-NET - LOPO



MORPH – Test Set

Conclusion and Discussion

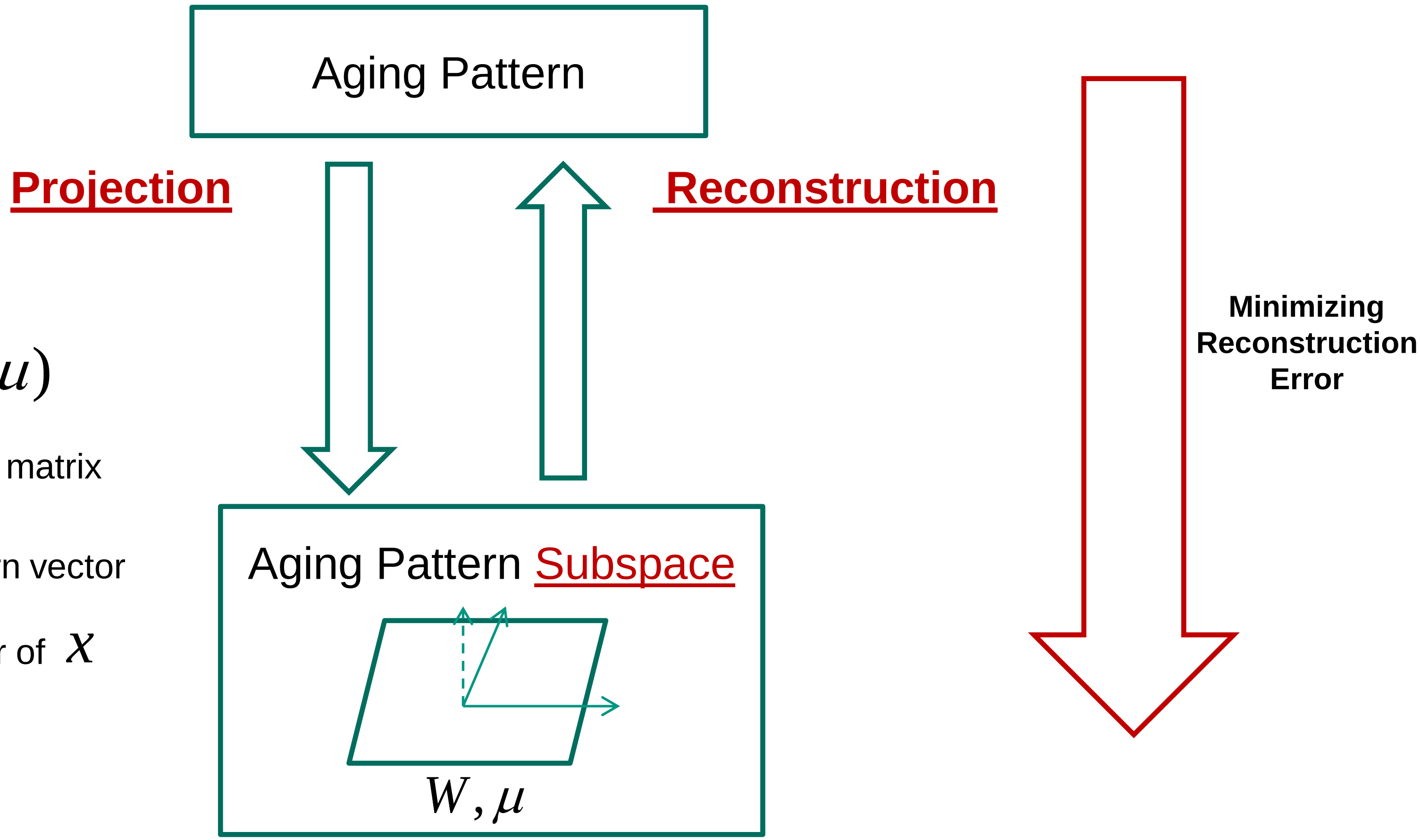
- AGES → an effective algorithm for learning and age estimation
- AGES Problem → insufficient training data
- LLD relieves the problem of insufficient training data
- AGES and IIS-LLD show better results than the compared approaches
- Sufficient data → AGES performs better than LLD otherwise LLD is better

- Future Work
  - Special feature extractor for age estimation
  - Voice, hair, gait can be considered

Referneces

- Geng, X.; Zhou, Z.-H.; Zhang, Y.; Li, G.; and Dai, H. Learning from facial aging patterns for automatic age es- timation. In Proc. the 14th **ACM Int'l Conf. Multimedia, 2006**.
- Xin Geng, Zhi-Hua Zhou, and Kate Smith-Miles. Automatic age estimation based on facial aging patterns**. *IEEE TPAMI, 2007.***
- T. Cootes, G. Edwards, and C. Taylor, Active Appearance Models, **IEEE Trans. Pattern Analysis and Machine Intelligence, June 2001.**
- Xin Geng, Kate Smith-Miles, Zhi-Hua Zhou. Facial Age Estimation by Learning from Label Distributions. ***AAAI'10, Atlanta, GA, 2010.***
- Yun Fu, Guodong Guo and Thomas S. Huang, Fellow, Age Synthesis and Estimation via Faces: A Survey. **IEEE November 2010**

AGES – Learning Algorithm



PCA

$$y = W^T (x - \mu)$$

W^T - Transponse matrix

x - Aging pattern vector

μ - Mean vector of x

AGES – Learning Algorithm

- **Initialisation**

  $i \leftarrow 0; \quad x_k^m \leftarrow [\mu_k^m]; \ \mu_k^m$ - mean vector

  Apply PCA to get W_0 and μ_0

AGES – Learning Algorithm

- **Initialization**

  $i \leftarrow 0; \quad x_k^m \leftarrow [\mu_k^m]; \ \mu_k^m$ - mean vector

  Apply PCA to get W_0 and μ_0

- **Projection**
  - Estimate $\longrightarrow y_k$

  $$[W_i(^a_k)]y_k = x_k^a - [\mu_i(^a_k)]$$

- **Reconstruction**
  - Reconstruct $\longrightarrow x_k$

  $$\hat{x} = \mu + Wy_k$$

  $$x_k^m \leftarrow \hat{x}_k^m$$

AGES – Learning Algorithm

- **Initialization**

  $i \leftarrow 0; \quad x_k^m \leftarrow [\mu_k^m]; \; \mu_k^m$ - mean vector

  Apply PCA to get W_0 and μ_0

- **Projection**
  - Estimate $\longrightarrow y_k$

    $[W_i(_k^a)]y_k = x_k^a - [\mu_i(_k^a)]$

- **Reconstruction**
  - Reconstruct $\longrightarrow x_k$

    $\hat{x} = \mu + W y_k$

    $x_k^m \leftarrow \hat{x}_k^m$

- **PCA**

  Apply PCA to get W_{i+1} and μ_{i+1}

  $i \leftarrow i+1$

AGES – Learning Algorithm

- **Initialization**

  $i \leftarrow 0; \quad x_k^m \leftarrow [\mu_k^m]; \ \mu_k^m$ - mean vector

  Apply PCA to get W_0 and μ_0

- **Projection**
  - Estimate $\longrightarrow y_k$

  $$[W_i(_k^a)]y_k = x_k^a - [\mu_i(_k^a)]$$

- **Reconstruction**
  - Reconstruct $\longrightarrow x_k$

  $$\hat{x} = \mu + W y_k$$

  $$x_k^m \leftarrow \hat{x}_k^m$$

- **PCA**

  Apply PCA to get W_{i+1} and μ_{i+1}

  $i \leftarrow i+1$

Repeat until reconstruction error < θ

AGES – Learning Algorithm

- Reconstruction Error

$$\bar{\varepsilon}^{a} = \frac{1}{N}\sum_{k=1}^{N}(x_k^a - \hat{x}_k^a)^T(x_k^a - \hat{x}_k^a)$$

- Our Goal is to find

  W and μ that minimize reconstruction error

LLD – Learning From Label Distribution

- Similarity measure $\longrightarrow$ Kullback-Leibler divergence

$$\theta^* = \arg\min_{\theta} \sum_i \sum_y \left(P_i(y) \log \frac{P_i(y)}{p(y \mid x_i;\theta)}\right) = \arg\max_{\theta} \sum_i \sum_y P_i(y) \log p(y \mid x_i;\theta) \quad (1)$$

LLD – Learning From Label Distribution

- Similarity measure $\longrightarrow$ Kullback-Leibler divergence

$$\theta^* = \arg\min_{\theta} \sum_i \sum_y \left(P_i(y) \log \frac{P_i(y)}{p(y \mid x_i; \theta)}\right) = \arg\max_{\theta} \sum_i \sum_y P_i(y) \log p(y \mid x_i; \theta) \quad (1)$$

- **Case 1** $\longrightarrow$ Single label

$P_i(y) = \delta(y, y_i), \delta(.,.)$ - Kronecker function

  - Consequently (1) can be simplified to the maximum lakelihood criterion
- **Case 2** $\longrightarrow$ Multi-label (equal probabilities)
  - Consequently (1) can be simplified to

$$\theta^* = \arg\max_{\theta} \sum_i \frac{1}{p_i} \sum_y \log p(y \mid x_i; \theta)$$

LLD – Learning From Label Distribution

- **Case 3** $\rightarrow$ Multiple labels (different probabilities)

  $f_k(x,y)$ - feature function, $\tilde{p}(x,y)-$ empirical joint distribution

  Expected value of $f_k(x,y)$ w.r.t. $\tilde{p}(x,y)$

  $$\tilde{f}_k = \sum_{x,y} \tilde{p}(x,y) f_k(x,y)$$

  Expected value of $f_k(x,y)$ w.r.t. $\tilde{p}(x)$ and $p(x \mid y;\theta)$

  $$\hat{f}_k = \sum_{x,y} \tilde{p}(x) p(y \mid x;\theta) f_k(x,y)$$

  Maximum Entropy model

  $$H = -\sum_{x,y} \tilde{p}(x) p(x \mid y;\theta) \log p(y \mid x;\theta)$$

  Subject to constrain $\hat{f}_k = \tilde{f}_k$

LLD – Learning From Label Distribution

- Maximum entropy model has the exponential form

$$p(y \mid x;\theta) = \frac{1}{Z}\exp\left(\sum_k \theta_k f_k(x,y)\right)$$

Different Age Ranges - MAE

Method	0-5	6-30	31-69
AGES	1.87	4.88	24.97
$AGES_r$	1.17	4.48	7.93

FG-NET - LOPO