On the use of Radon Transform for Facial Expression Recognition

Nicolas Tsapatsoulis, Yannis Avrithis and Stefanos Kollias

National Technical University of Athens

Department of Electrical and Computer Engineering

Problem Statement

- *Facial expression recognition* based on two instances of a face in the same emotional state
- Face *detection* and *registration* to account for translation, scaling and rotation variations
- Spatiotemporal expression description based on *Radon transform* of motion vectors between neutral and 'apex' condition of the expression
- Radon curve *normalization* for translation, scaling & resolution invariant representation
- Expression *classification* using correlation criterion and neural network classifier

Facial Expression Recognition

- Human interaction consists of transmitted messages either *explicit* or *implicit*
- Expression recognition permits categorization of active & spontaneous expressions, interpretation of *implicit messages*, and understanding of *emotional states* using visual cues
- Recognition approaches divided into
  - *Static* : using a single face instance
  - *Semi-static* : using two instances, neutral and 'apex'
  - *Dynamic* : using several frames (usually 3-15)

Assumptions / Constraints

- *Semi-static* approach adopted, using two images representing a face in its neutral condition and the 'apex' of the expression
- Only one face with significant scaling in expression images - *no multiface detection* required
- The face is the same along the expression - *no aging or personal variations* like make-up, eyeglasses or beard

Face Detection and Registration (1)

- Let $M(u,\theta)$ be a *face template* at orientation θ and scale $u(h,v)$ described by horizontal scaling h and vertical scaling v
- If F is frame containing a face at arbitrary location, scale & rotation and $A \subset F$ a frame area, the *minimum correlation* between A and M is obtained at scale u and orientation θ:

$$r(u,\theta) = \min_{A \subset F} \left\{ \frac{\left| A - M(u,\theta) \right|}{\rho \cdot a \cdot b} \right\}$$

where $\rho = 1 - c \cdot |h/v - 2/3|$, a = mean(A) and b = mean(M)

Face Detection and Registration (2)

- Best scale & orientation obtained by

  $[U,\Theta] = \text{argmin}\{r(u,\theta)\}$
- Final area A^* detected using template $M(U,\Theta)$ and transformed to *standard coordinates* using scaling & rotation according to U, Θ.
- Detection applied to frame of neutral condition - parameters U, Θ *re-used* for detection in ‘apex’ of the expression
- *Registration* based on facial features (eyes, nose and mouth) not implemented - only required for face recognition

Optical Flow Estimation (1)

- *Optical flow* derived directly from facial pixel values $p_0(x,y)$, $p_1(x,y)$ of image frames F_0, F_1 corresponding to neutral and 'apex' conditions
- Motion vectors $\hat{\mathbf{v}}(x,y) = (\hat{v}_x, \hat{v}_y)$ obtained using *block matching* :

$$\hat{\mathbf{v}}(x,y) = \underset{(v_x,v_y)\in Q}{\arg\min} \sum_{l=-n}^{n} \sum_{m=-n}^{n} \left| d(x+l, y+m; v_x, v_x) \right|$$

where $d(x,y;\ v_x,v_y) = p_0(x,y) - p_1(x-v_x, y-v_y)$
and $Q = \{-q,..,q\} \times \{-q,..,q\}$ is a search area

Optical Flow Estimation (2)

- Motion vectors only calculated for image blocks with significant error between neutral and ‘apex’ images, based on image dependent *thresholding*
- *Logarithmic search* employed for block matching to reduce computational load
- *Median filtering* performed on motion vector phase and magnitude (directional filtering) to achieve motion vector smoothness and discard estimation noise

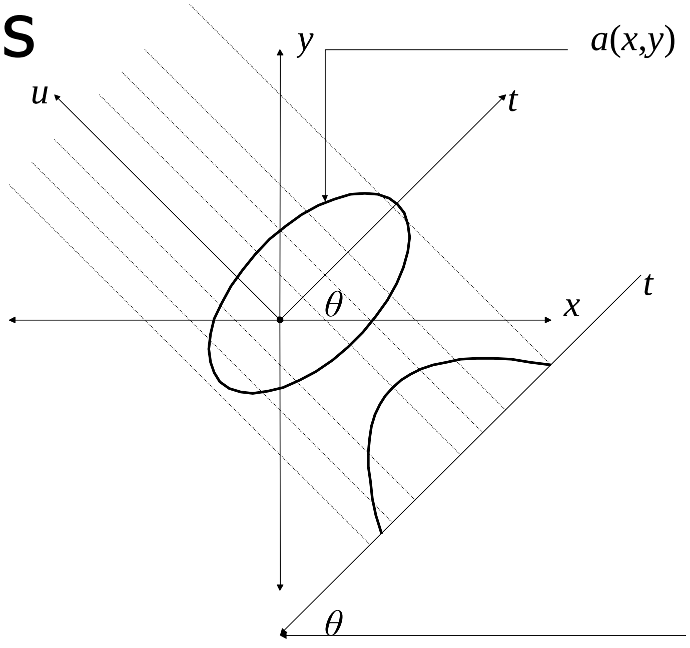
Radon Transform

- Spatio-temporal expression representation obtained through *Radon transform* of motion vectors at different angles:

$$R(\theta) = \sum_{u=-\infty}^{\infty} a_k(x,y)\Big|_{x=t\cos\theta-u\sin\theta,\, y=t\sin\theta+u\cos\theta}$$

where m.v. are expressed as

$$\hat{\mathbf{v}}_k(x,y) = a_k(x,y)e^{j\phi_k(x,y)}$$



- Expressions characterized by Radon transform projections on angles 0° and 90°, called '*signatures*'

Signature Normalization (1)

- Let $R = [r_i]$, $i \in F = \{0,..,L\text{-}1\}$ be the Radon transform at angle 0° or 90°. *Vertical scale normalization* is performed first, as

$$s_i = r_i \left(\frac{1}{L} \sum_{k=0}^{L-1} r_k^2 \right)^{-1/2}, \quad i = 0, \ldots, L-1$$

- *Horizontal normalization* is performed next, by removing zero values from the left and right edges of $S = [s_i]$, $i \in F$:

$$z_i = s_{i+i_L}, \quad i = 0, \ldots, i_R - i_L$$

where $i_L = \min\{i \mid i \in F'\}$, $i_R = \max\{i \mid i \in F'\}$ and $F' = \{i \in F : s_i > T\}$

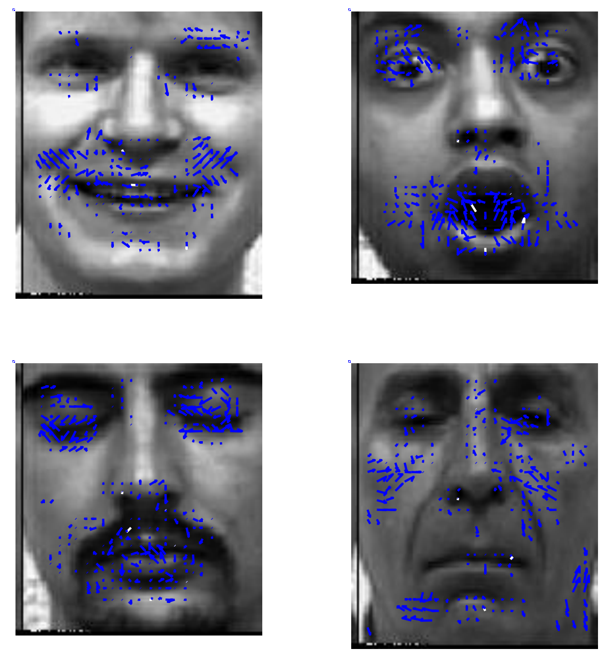
Signature Normalization (2)

- *Re-sampling* of vector $Z = [z_i]$, $i \in F$ at K points using linear interpolation gives the normalized vector $N = [n_i]$, $i = 0,..,K-1$
- The normalization process:
  - is *invariant* to translation, scaling and resolution of signature vectors
  - consists of *linear* operations - no information is lost
  - is applied to all signatures in the same way - *no comparison* or *matching* is required, and normalized signatures can be directly used with any classification mechanism

Experimental Results

- 75 images from the Yale database used, corresponding to expressions ‘normal’, ‘happy’, ‘surprised’, ‘sad’ and ‘sleepy’

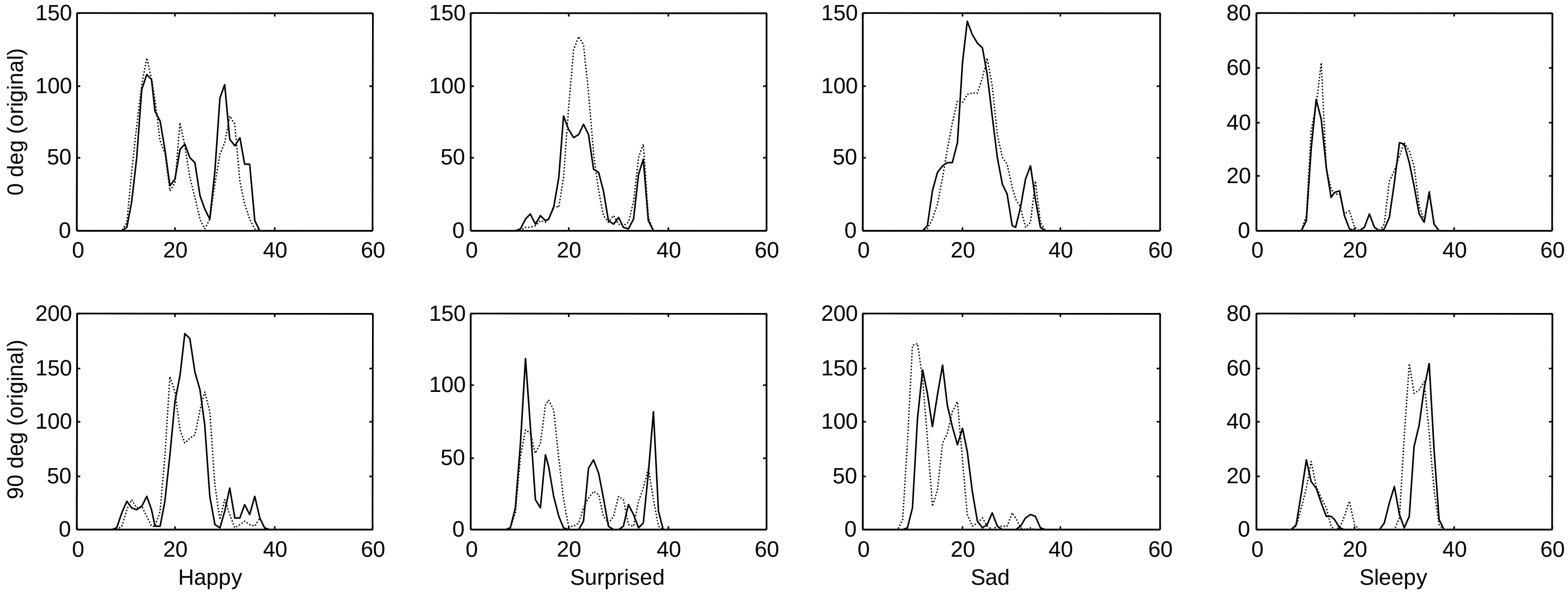
Neutral expressions after face detection & registration



‘Apex’ expressions with estimated motion vectors

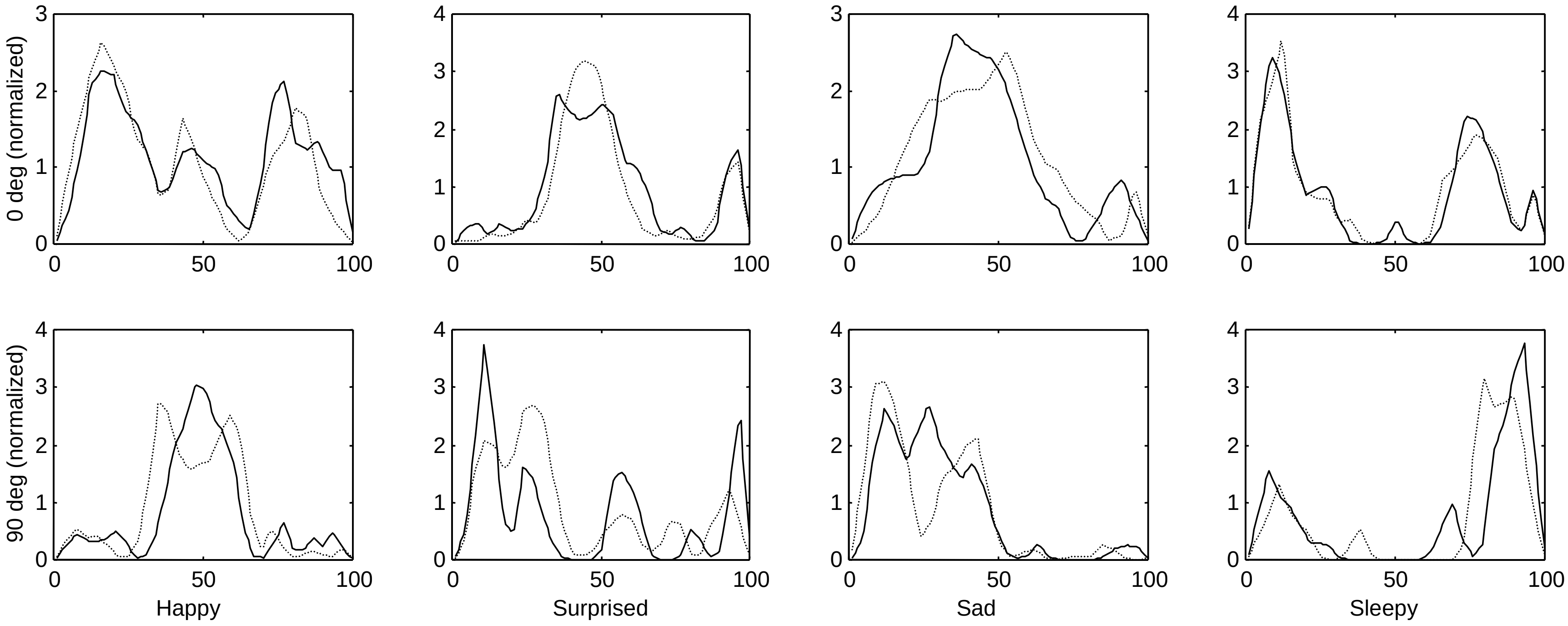
Radon Transform Results

- *Radon transform curves* for two different subjects at 0° and 90°, corresponding to four expressions:

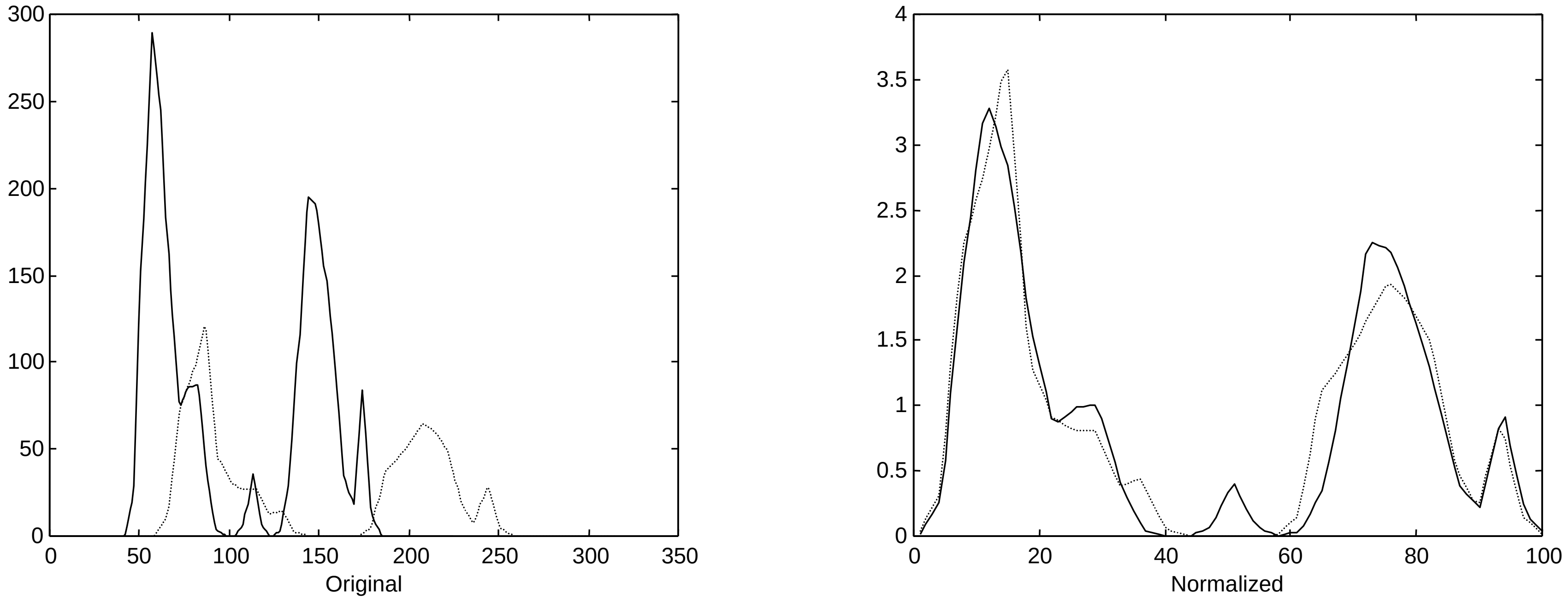


Curve Normalization Results

- *Normalized signatures* for two different subjects at 0° and 90°, corresponding to four expressions:



Curve Normalization Results



Original and normalized signatures of two subjects corresponding to expression 'sleepy' at 0 degrees

Classification using Correlation Coefficients

	Happy A	Happy B	Surprised A	Surprised B	Sad A	Sad B	Sleepy A	Sleepy B
Happy A	1.0000	0.7859	-0.1193	-0.1906	-0.0225	0.1323	-0.0194	0.0262
Happy B	0.7859	1.0000	-0.1619	-0.1461	0.0791	0.1750	-0.0550	-0.0107
Surprised A	-0.1193	-0.1619	1.0000	0.7564	0.6560	0.5561	-0.1663	-0.2771
Surprised B	-0.1906	-0.1461	0.7564	1.0000	0.8014	0.5475	-0.2752	-0.3388
Sad A	-0.0225	0.0791	0.6560	0.8014	1.0000	0.8411	-0.3254	-0.3196
Sad B	0.1323	0.1750	0.5561	0.5475	0.8411	1.0000	-0.2686	-0.2383
Sleepy A	-0.0194	-0.0550	-0.1663	-0.2752	-0.3254	-0.2686	1.0000	0.7842
Sleepy B	0.0262	-0.0107	-0.2771	-0.3388	-0.3196	-0.2383	0.7842	1.0000

Neural Network Classifier

- MLP network with 200 *input units*, four *output units* and one *hidden layer* with 20 units
- 20 signatures used for training set, 8 used for validation set and 32 for testing
- After 400 learning cycles of *backpropagation*, the NN could recognize all expressions of the training set and 7 out of 8 from the validation set
- The NN *generalized* well to the test set, with 87.5% overall correct classification

NN Classification Results

Expression	Happy	Sadness	Surprised	Sleepy
Happy	8	0	1	0
Sadness	0	5	0	0
Surprise	0	2	7	0
Sleepy	0	1	0	8
Success	100%	62.5%	87.5%	100%

Conclusions - Further Work

- Facial expression recognition achieved through *Radon transform* of optical flow between neutral and ‘apex’ conditions of the expression
- *Normalized signatures* used directly for classification either with *correlation* or *neural network*, with promising results
- The method is currently applied on a *larger database* to evaluate its efficiency
- An extension is under investigation for expression recognition using *video sequences*