### Siamese-structure DNN Recognizing Changes in Facial Expression According to the Degree of Smiling

K. KONDO, T. NAKAMURA, Y. NAKAMURA (KYOTO UNIVERSITY) S. SATOH (NII)

# Background

#### Automatic QOL evaluation

- Evidence-based lifestyle support in aging or stress-ridden society
- Recognizing 'Smile' for estimating 'happiness'

#### Problem

- Ambiguous facial expressions frequently occur
- Only one image does not achieve confident estimation



Smile ?

Not Smile ?

### Our approach



t-1

t+1

Smiling intensity is increasing

t



. . .

How much accurately a computer recognizes such facial expression changes ?

# Network for recognizing 'Smile'



Input images

Feature extractor

Comparator

Recognition result

### Dataset

- Face capture :
  - when listening to a ratio program or enjoying a conversation
  - captured from 7 directions

















# Labeling

- Annotators sequentially compared two images
  - Starting from two images t and t+1 frames
  - If smiling intensity change, register the image pair with a label
  - If not, comparison proceeds to t and t+2 frames



# Dataset for subtle change in facial expression

Intermediate images between the annotated pairs



A rectangle label is given to an image with greater smile

	situation	pairs	acc.
Person 1	Radio	290	0.993
Person 1	Talk	482	0.998
Person 2	Talk	150	0.993

Contribution map (for the predicted label)

Contribution map (for the other label)

#### Ground truth Predicted



- Training with frontal face images in dataset A
- Test with frontal face images in dataset B



### Results : reasons of incorrect estimation

# 1. Little visual pattern deformation



2. Smiling intensity comparison is difficult



#### 3. Not well trained ?



#### All combinations within each image pairs

















Not a monotonic change (=Ground truth label is rather incorrect)

Which two images should be compared is an important issue

- Leave-one-out cross validation 
  Direction-fold cross validation

Person 1 talk	0.918		0.886		0.857	0.837		0.929		0.90
	0.922	0.95 I	0.961	0.956	0.900	0.878	0.946	0.95 I	0.989	0.87
	0.933		0.972		0.950	0.889		0.956		0.93
Person 2 talk	0.902		0.963		0.927	0.902		0.927		-
	0.963	0.951	1.000	0.951	0.963	-	-	0.943	0.890	0.92
	0.963		1.000		1.000	-		0.932		0.932
Person 3 talk	0.806		0.964		0.824	0.973		0.982		0.964
	0.986	0.986	0.986	0.995	0.982	0.973	0.991	0.991	0.995	0.982
	0.964		0.982		0.982	0.982		0.991		0.99

Enough robust to capturing directions

#### Person 1



#### Person 2



#### Person 3



### Future works

- Image section issue ( how to pick two images from a sequence )
- Smiling intensity difference/similarity issue
- During talk or eating issue
- Absolute rating of smiling intensity
- Facial expression with mixture of multiple emotions