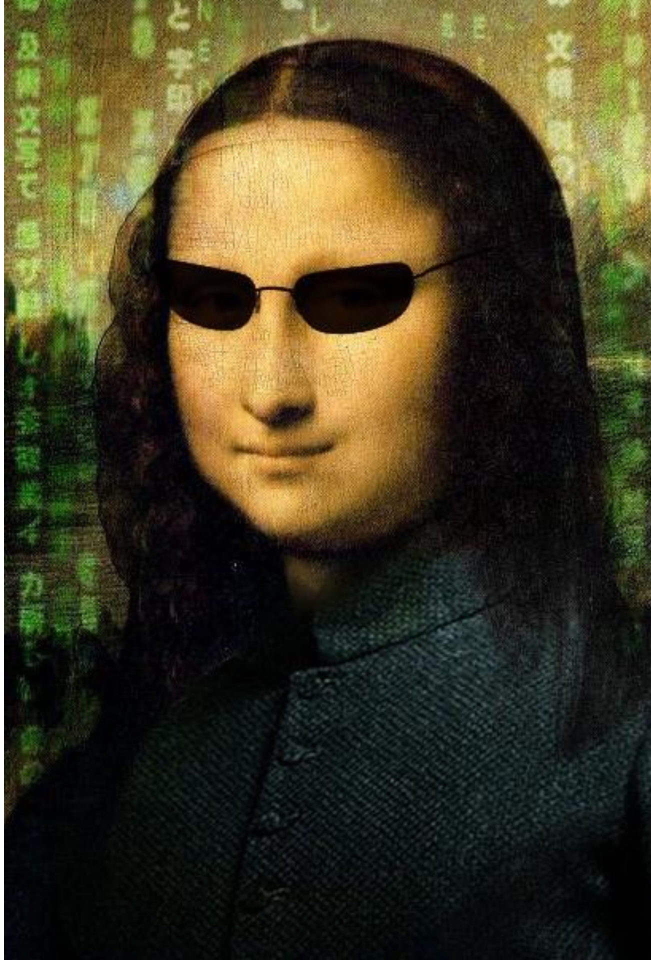
Robust Face Recognition with Occlusions in both Reference and Query Images

Xingjie Wei, Chang-Tsun Li, Yongjian Hu
Department of Computer Science,
University of Warwick
x.wei@warwick.ac.uk
http://warwick.ac.uk/xwei



THE UNIVERSITY OF WARWICK

Outline

- Face recognition with occlusions
- Current methods
- Three occlusion cases
- Our methods
- Experimental results

Face recognition with occlusions

- Intra-class variations > inter-class variations
- Causes imprecise registration of faces

→→ poor recognition performance !

Face recognition with occlusions

- Why is it so difficult?
- No prior knowledge of occlusion
- location,size,shape,texture -- unpredictable!

Current methods

Reconstruction based methods

[Wright et al. PAMI09, Yang et al. ECCV'10, Zhang et al. ICCV'11]

- An occluded probe image is represented as a linearly combination of unoccluded gallery images
- The probe image is assigned to the class with the minimal reconstruction error

Assume the gallery/training images are ***clean***

Three occlusion cases

Table 1. Three typical occlusion cases in the real world.

	Gallery	Probe	Scenarios
Uvs.O:	Unoccluded	Occluded	Access control, boarder check
Ovs.U:	Occluded	Unoccluded	Suspect detection, shoplifter recognition
Ovs.O:	Occluded	Occluded	

Current methods

Very limited work considers the existence of occlusions in both gallery and probe sets

- [Jia et al. FG'08,CVPR'09]
  - Proposed a reconstruction based method in, as well as an improved SVM

  Depend on an occlusion mask trained through the use of skin colour

- [Chen et al. CVPR'12]
  - Uses the low-rank matrix recovery to remove the occlusions

  Requires faces to be well registered in advance

WARWICK

Our method

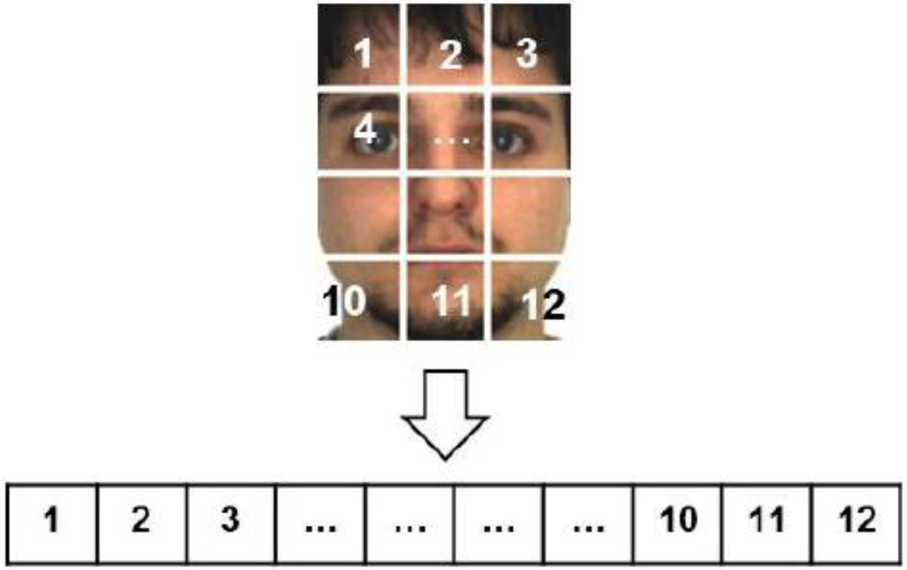
- Dynamic *Image-to-Class* Warping
  - An image → a patch sequence
  - Matching → the *Image-to-Class* distance

✓No occlusion detection

✓No training phase

Our method

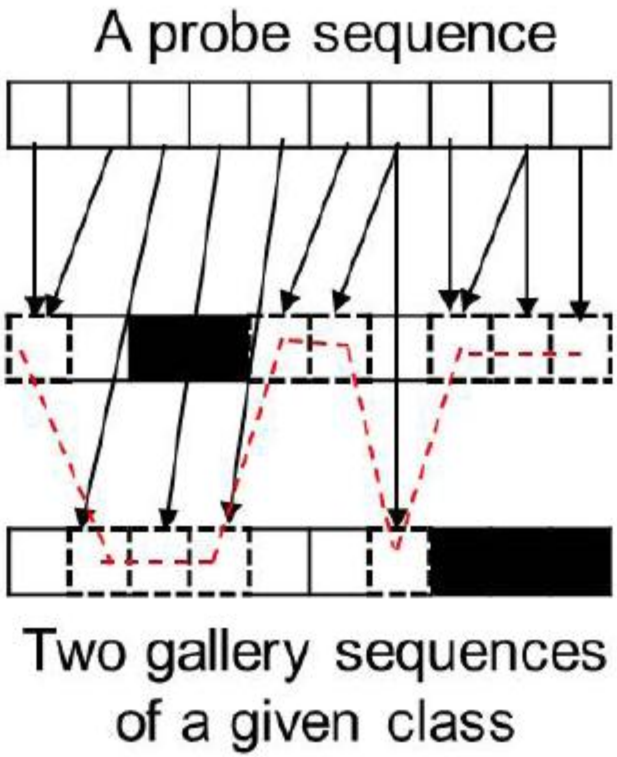
- Face representation
  - Natural order: *forehead, eyes, nose and mouth to chin* does not change despite occlusion or imprecise registration



WARWICK

Our method

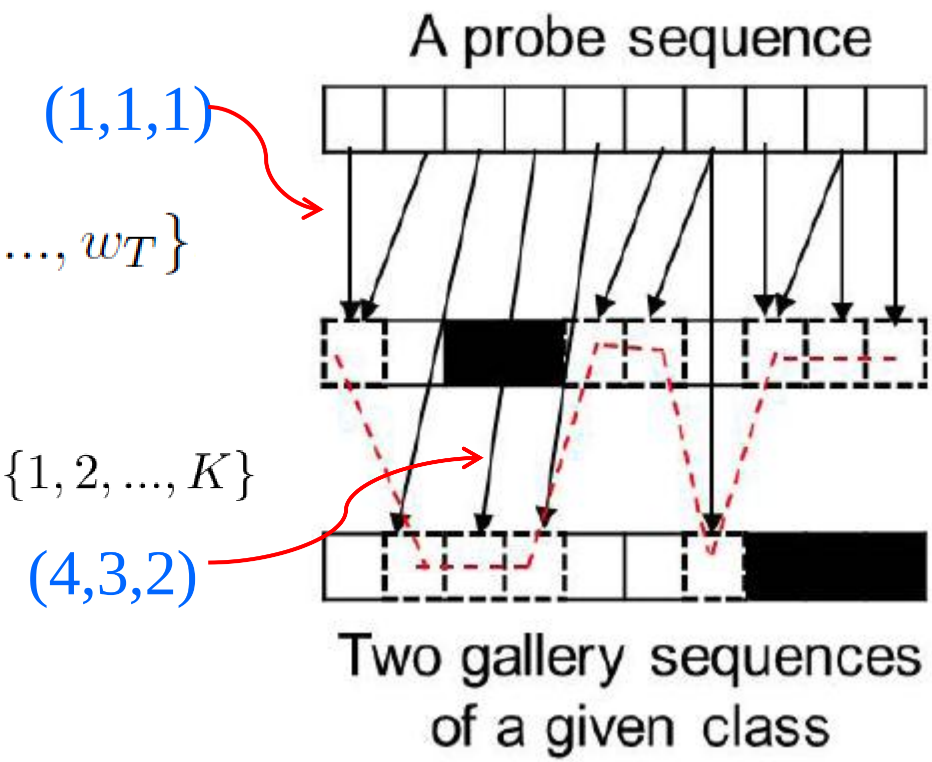
- *Image-to-Class* distance
  - from a probe sequence to all the gallery sequences of an enrolled class
  - each patch in the probe sequence can be matched to a patch from different gallery sequences

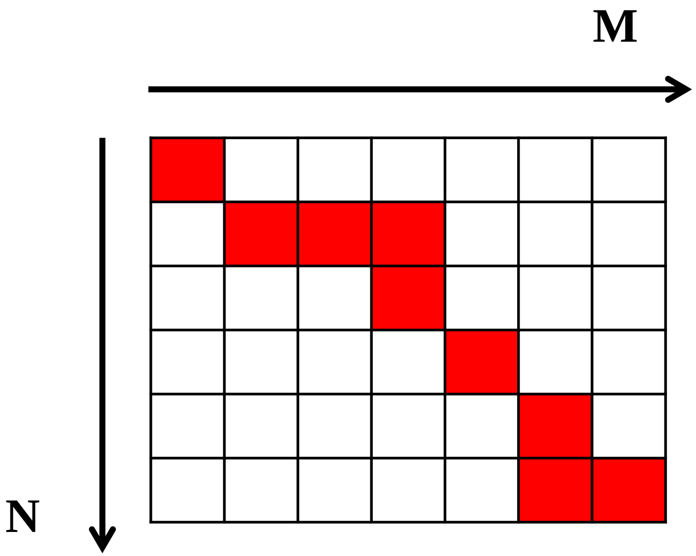


Our method

Warping path: $\boldsymbol{W} = \{w_1, w_2, \dots, w_T\}$

$$w_t : (m_t, n_t, k_t) \in \{1, 2, \dots, M\} \times \{1, 2, \dots, N\} \times \{1, 2, \dots, K\}$$



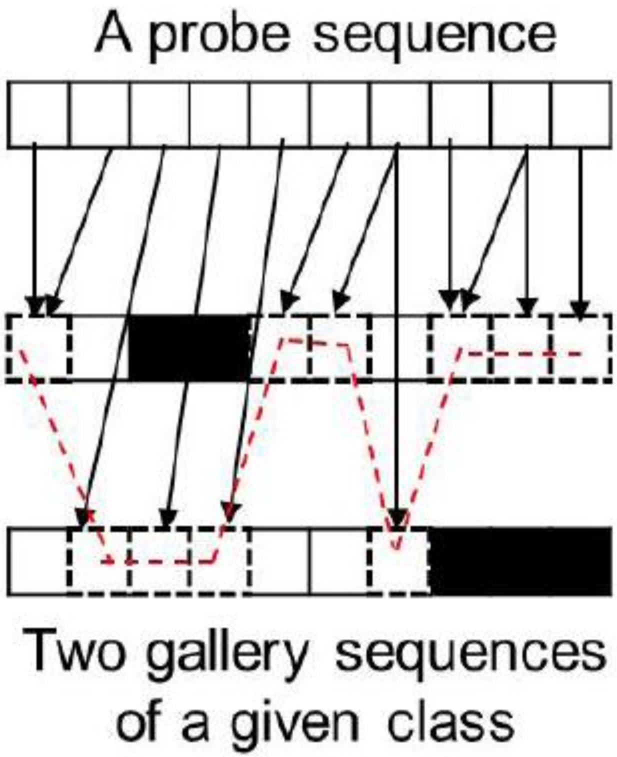


$$\max(M, N) \leqslant T \leqslant M + N - 1$$

Our method

Warping path: $\boldsymbol{W} = \{w_1, w_2, ..., w_T\}$

$w_t : (m_t, n_t, k_t)$



- Constraints:
  - Boundary

    $m_1 = n_1 = 1, m_T = M, n_T = N$
  - Continuity

    $m_t - m_{t-1} \leqslant 1, n_t - n_{t-1} \leqslant 1$
  - Monotonicity

    $m_{t-1} \leqslant m_t, n_{t-1} \leqslant n_t$
  - Window constraint: $|m_t - n_t| \leqslant l$

Maintains the order of facial features

Our method

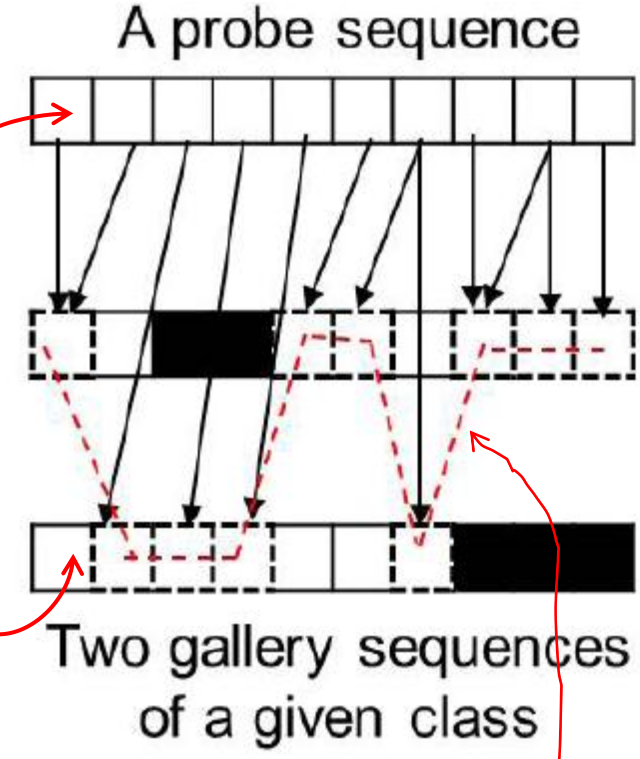
- Local cost

$$C_{w_t} = C_{m_t,n_t,k_t} = \|\boldsymbol{p}_m - \boldsymbol{g}_{kn}\|_2$$

- Optimal overall cost

$\boldsymbol{p}_m$

P



G

$\boldsymbol{g}_{kn}$

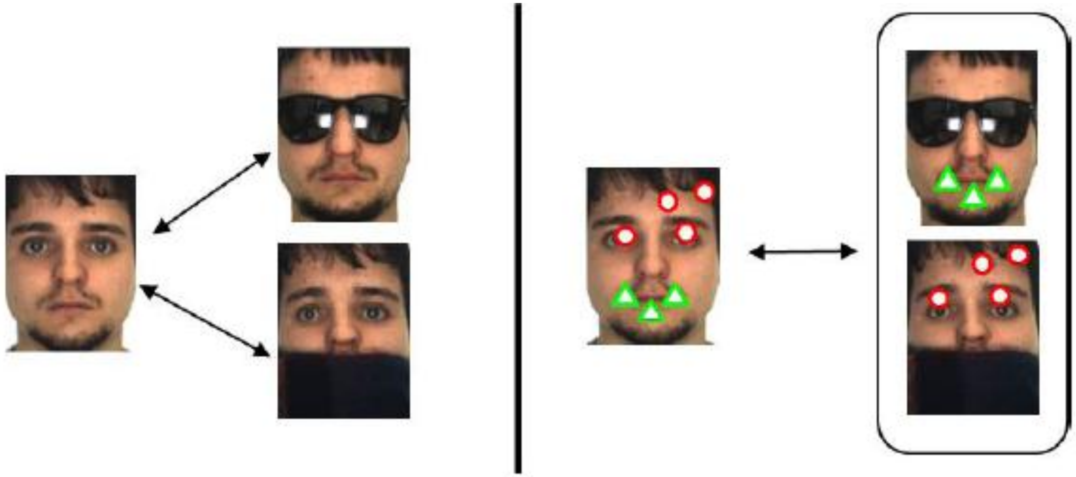
$$\boxed{DICW(\boldsymbol{P},\boldsymbol{G}) = \min_{w_t \in \boldsymbol{W}} \sum_{t=1}^{T} C_{w_t}}$$

the optimal $\boldsymbol{W}$

WARWICK

Our method

- Why does it work?
  - Tries every possible warping path and select the one with minimal overall cost
  - Exploits the information from different gallery images and reduce the effect of occlusions



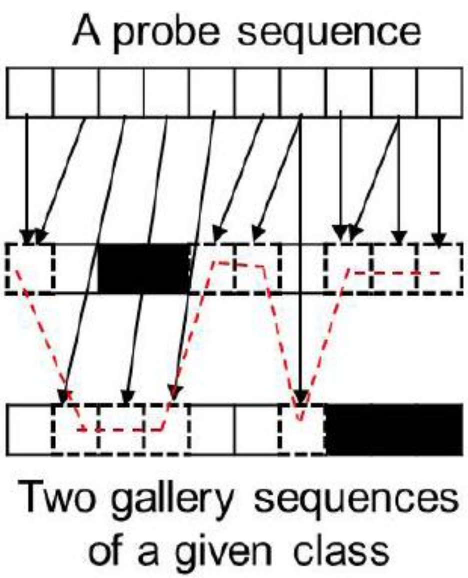
WARWICK

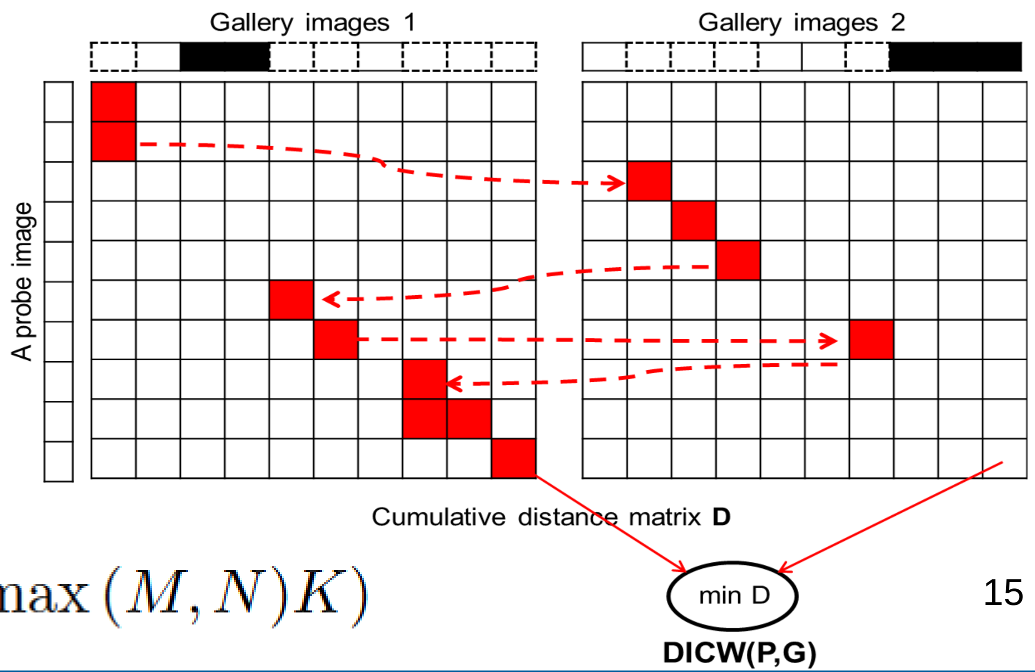
Our method

Dynamic Programming (DP)

$$D_{m,n,k} = \min\begin{pmatrix} D_{\{(m-1,n-1)\}\times\{1,2,\dots,K\}} \\ D_{\{(m-1,n)\}\times\{1,2,\dots,K\}} \\ D_{\{(m,n-1)\}\times\{1,2,\dots,K\}} \end{pmatrix} + C_{m,n,k}$$

$$DICW(\boldsymbol{P},\boldsymbol{G}) = \min_{w_t \in \boldsymbol{W}} \sum_{t=1}^{T} C_{w_t} \Rightarrow \boxed{DICW(\boldsymbol{P},\boldsymbol{G}) = \min_{k\in\{1,2,\dots,K\}} \boldsymbol{D}_{M,N,k}}$$





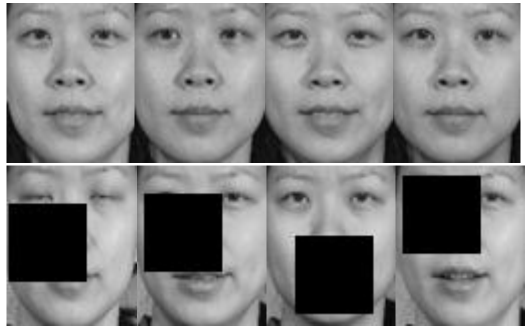
$$O(l\max(M,N)K)$$

Experiments

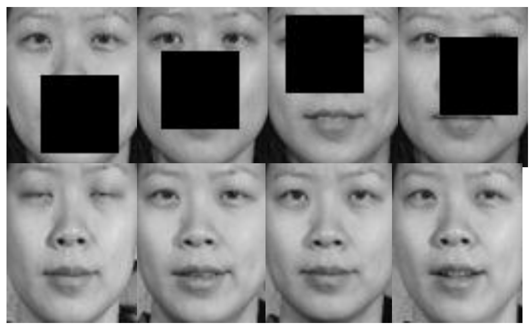
- The FRGC database
  - 44,832 images with different illuminations and expressions, 100 subjects selected
- The AR database
  - >4000 images with real disguise, 100 subjects selected
- Realistic images
  - >2000 frontal view faces of strangers on the streets, 80 subjects selected

Synthetic occlusions

- **Uvs.O**
  - Gallery
  - Probe



- **Ovs.U**
  - Gallery
  - Probe

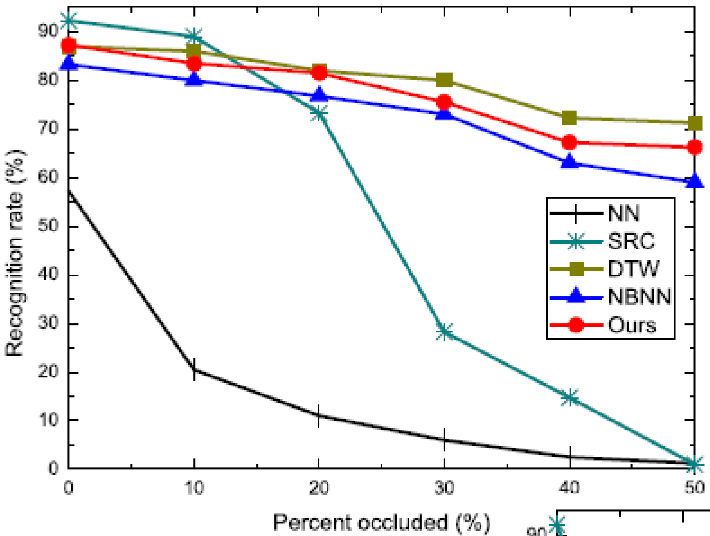


- **Ovs.O**
  - Gallery
  - Probe

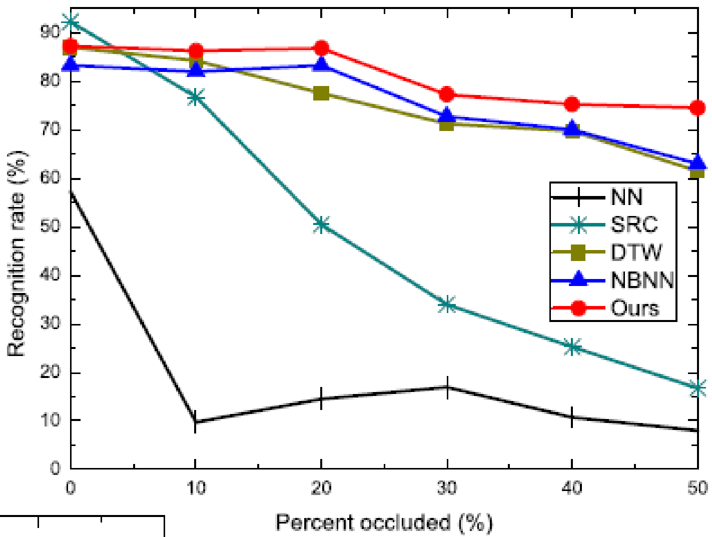
Block size: 0%~50% of the image

Block location: random and unknown to the algorithm

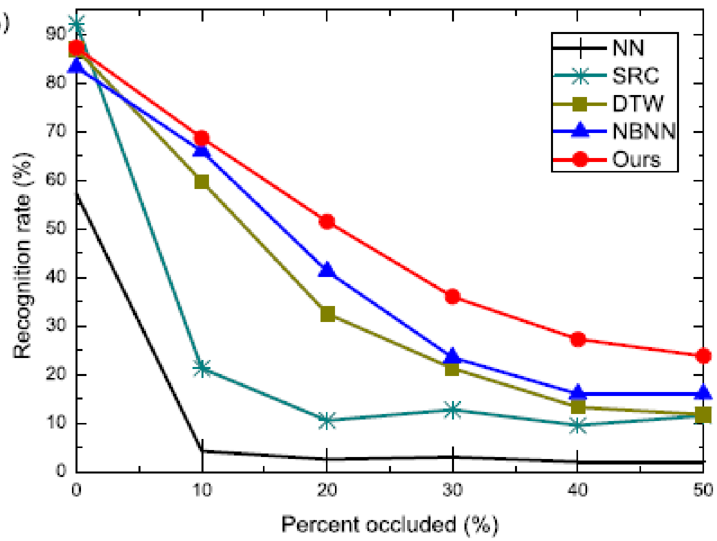
Synthetic occlusions



Uvs.O

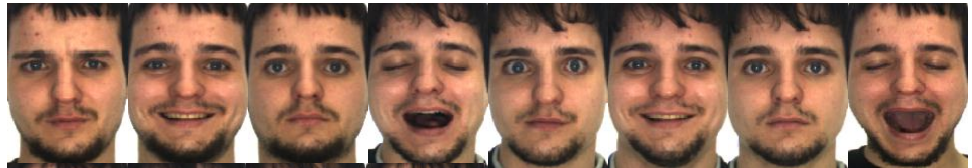
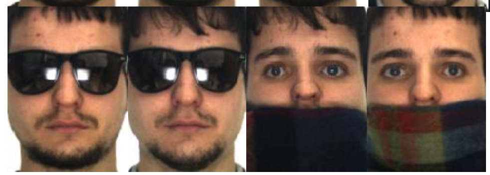
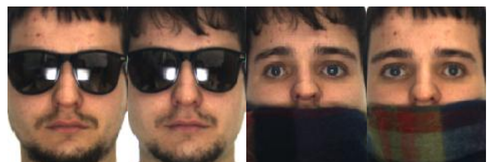
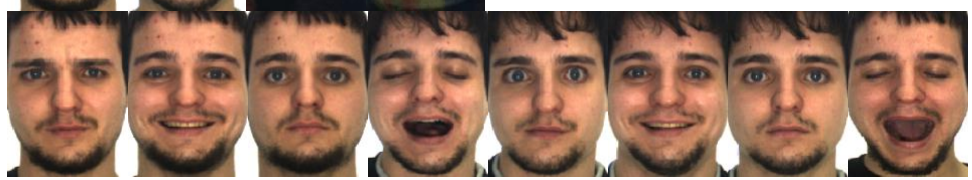
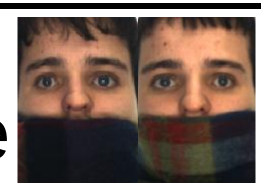
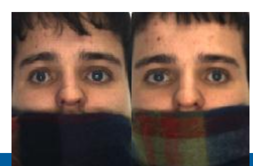


Ovs.U



Ovs.O

Real disguises

- **Uvs.O**
  - Gallery 
  - Probe 
- **Ovs.U**
  - Gallery 
  - Probe 
- **Ovs.O**
  1. Gallery  Probe 
  2. Gallery  Probe 

Real disguises

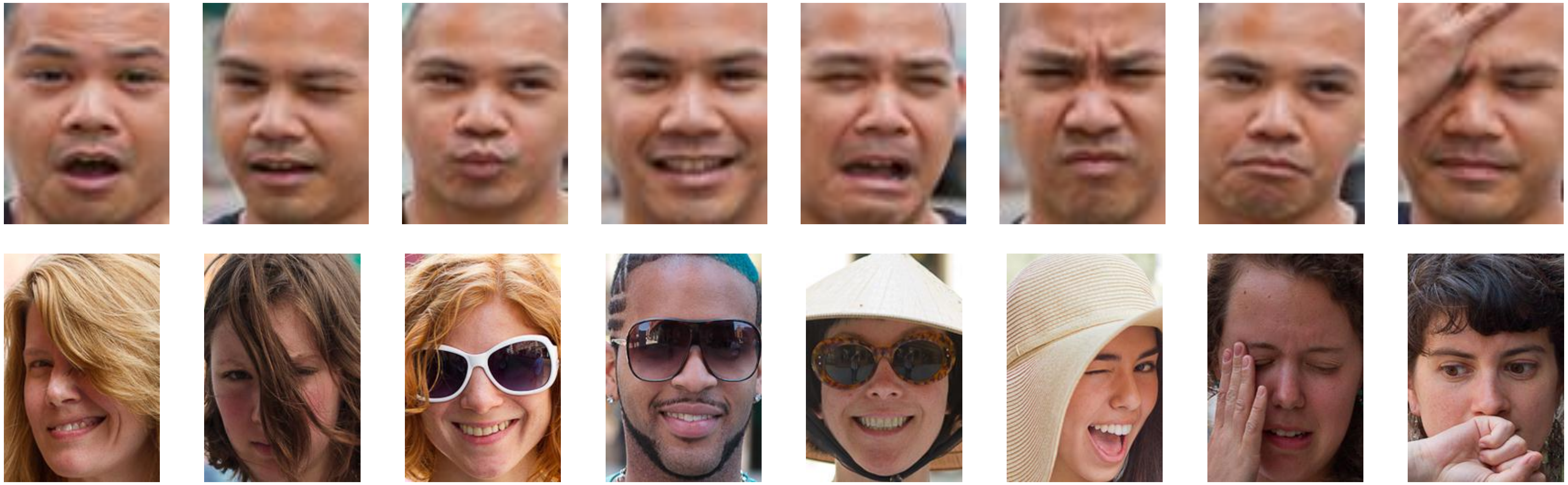
Table 2. Recognition rates (%) on the AR database

		NN	SRC	DTW	NBNN	Ours
Uvs.O: Gallery-unoccluded						
Probe	sunglasses	69.5	87.0	99.0	96.5	**99.5**
	scarf	11.5	59.5	96.5	95.5	**98.0**
Ovs.U: Gallery-occluded						
Probe	unoccluded	42.6	85.7	87.7	94.4	**94.6**
Ovs.O: Gallery-occluded						
Probe	sunglasses	5.5	18.0	55.0	49.0	**56.0**
	scarf	5.5	10.0	**61.5**	52.5	55.5

Realistic images

Table 3. Recognition rates (%) using the realistic images

NN	SRC	DTW	NBNN	Ours
62.5	69.4	64.4	68.1	**75.6**



WARWICK

Thank you

- Questions ?

- Xingjie Wei
- x.wei@warwick.ac.uk
- http://warwick.ac.uk/xwei
- Department of Computer Science, University of Warwick

WARWICK