中国科学技术大学
University of Science and Technology of China

Single-Image Face Recognition via Sparse Illumination Transfer

Dr. Liansheng Zhuang

http://staff.ustc.edu.cn/~lszhuang

Outline

中国科学技术大学
University of Science and Technology of China

- Motivation
- Sparse Representation Classification
- Sparse Illumination Transfer
- Single-Sample Face recognition
- Experiments
- Conclusion

[1] Liansheng ZHUANG, Allen Y. Yang, Zihan Zhou, S. Shankar Sastry, Yi Ma. "Single-Sample Face Recognition with Image Corruption and Misalignment via Sparse Illumination Transfer", in CVPR 2013.

[2] Liansheng ZHUANG, Tsung-Han Chan, Allen Y. Yang, S. Shankar Sastry, Yi Ma. "Sparse Illumination Learning and Transfer for Single-Sample Face Recognition with Image Corruption and Misalignment", 114(2), IJCV, 2015.

Motivation

- very lively in 2015

-Image Face Recognition via Sparse Illumination Transfer

Motivation

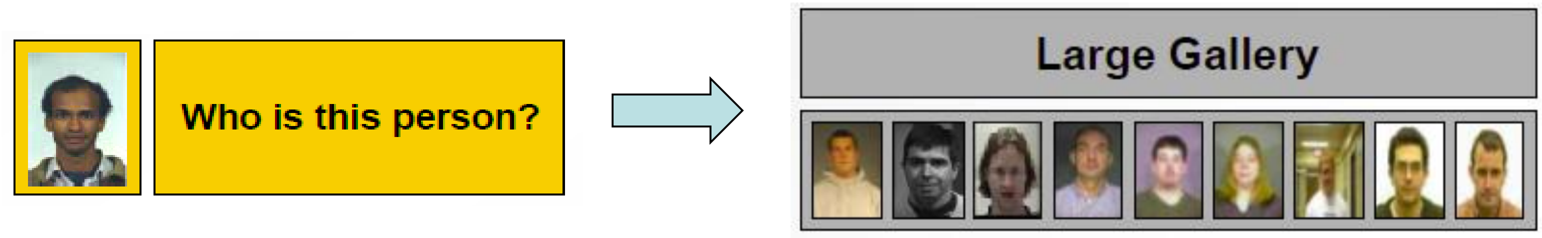
- very lively in 2015

1:1 Face Verification ---> Deep Learning

1:N Face Recognition ---> ?

Motivation

- Face Recognition Problem:



Setting: Only one frontal image under arbitrary lighting is available for each subject in the gallery!

- ✓ **Surveillance**
- ✓ **Internet face recognition**
- ✓ **Access control**
- ✓ **……**

Very Challenging but more practical !

Motivation

- Why is face recognition so difficult?
  - Pose
  - Illumination
  - Expression
  - Aging(time lapse)
  - Occlusion (glasses, accessories, markup, …)
  - …….

Motivation

中国科学技术大学
University of Science and Technology of China

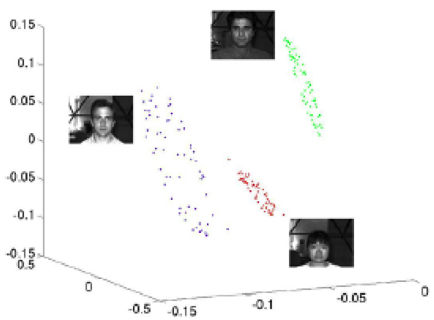
- Why is face recognition so difficult?
  - ✓ Pose
  - ✓ Illumination
  - ■ Expression
  - ■ Aging(time lapse)
  - ✓ Occlusion (glasses, accessories, markup, …)

Sparse representation based Classification (SRC) provides a good framework for these factors.

Overview of SRC

- Face subspace model [Belhumeur et al. ’97, Basri & Jacobs ‘03]

Assume $\mathbf{b}$ belongs to Class i from K classes.



$$\begin{aligned} \mathbf{b} &= \alpha_{i,1}\mathbf{a}_{i,1} + \alpha_{i,2}\mathbf{a}_{i,2} + \cdots + \alpha_{i,n_1}\mathbf{a}_{i,n_i}, \\ &= A_i\alpha_i. \end{aligned}$$

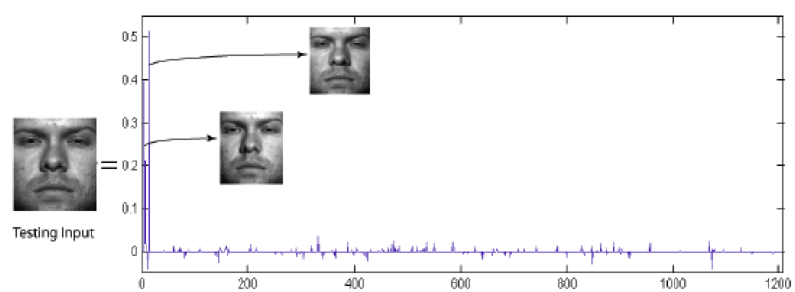
where $A_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \cdots, \mathbf{a}_{i,n_i}]$.

- Nevertheless, class i is unknown. We solve a bigger problem: **b** = **A**x

- Sparse representation problem:

$$(P_0): \quad \mathbf{x}^* = \arg\min \|\mathbf{x}\|_0 \quad \text{subj. to} \quad A\mathbf{x} = \mathbf{b}$$

$$(P_1): \quad \mathbf{x}^* = \arg\min \|\mathbf{x}\|_1 \quad \text{subj. to} \quad A\mathbf{x} = \mathbf{b}$$



x* encodes membership through its nonzero coefficients!

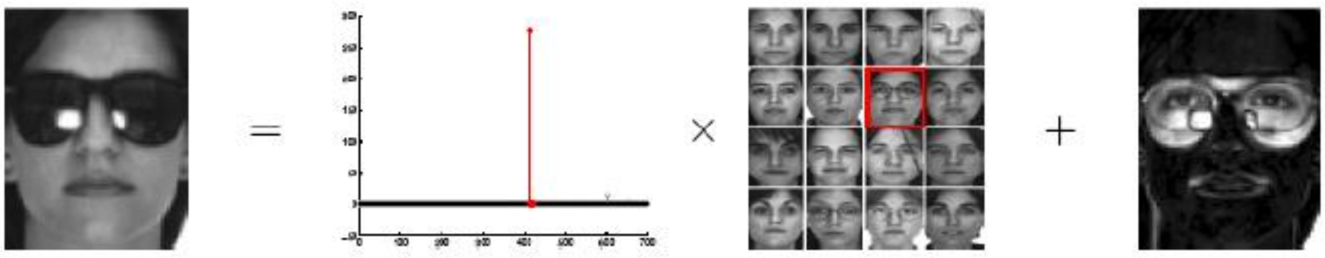
Ref: Wright, AY, Sastry, Ma, Robust face recognition via sparse representation. IEEE PAMI, 2009

Overview of SRC

- To deal with Occlusion, Corruption and Disguise

Sparse representation + sparse error

$$\mathbf{b} = A\mathbf{x} + \mathbf{e}$$



- To deal with misalignments (caused by pose variation)

Seek an image transformation τ (affine or homography)

$$\hat{\tau} = \underset{\boldsymbol{x}_i, \boldsymbol{e}, \tau}{\operatorname{argmin}} \|\boldsymbol{e}\|_1 \quad \text{subj. to} \quad \boldsymbol{b} \circ \tau = A_i \boldsymbol{x}_i + \boldsymbol{e},$$

$$\Downarrow$$

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}, \boldsymbol{e}} \|\boldsymbol{x}\|_1 + \|\boldsymbol{e}\|_1$$

$$\text{subj. to} \quad \boldsymbol{b} = \left[A_1 \circ \tau_1^{-1}, \cdots, A_L \circ \tau_L^{-1}\right] \boldsymbol{x} + \boldsymbol{e}$$

Deformable SRC

Ref: Wagner, et al., Towards a practical face recognition system. IEEE PAMI, 2012

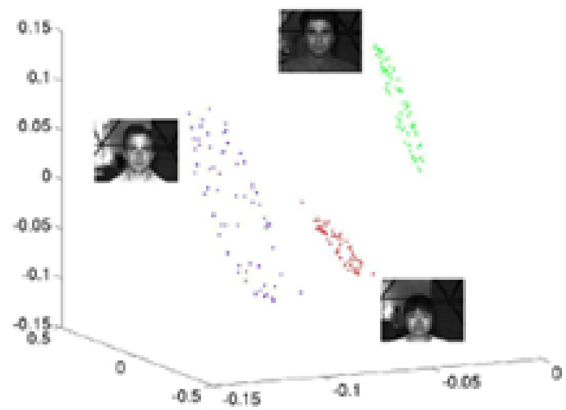
Overview of SRC

- Deformable SRC:

$$\mathbf{b} \circ \tau = Ax + e.$$

- Pose
- Illumination
- Occlusion
- Disguise
- Corruption

$$\begin{aligned} \mathbf{b} &= \alpha_{i,1}\mathbf{a}_{i,1} + \alpha_{i,2}\mathbf{a}_{i,2} + \cdots + \alpha_{i,n_i}\mathbf{a}_{i,n_i} \\ &= A_i\alpha_i. \end{aligned}$$



Shortcoming: Require many images for each subject!

Sparse Illumination Transfer

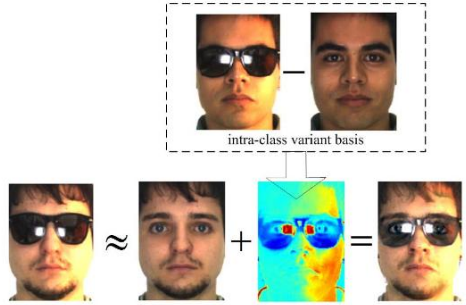
中国科学技术大学
University of Science and Technology of China

Core Idea: Compensate the missing illumination information.

- Test Image = Gallery Image + (Inter-subj. variation + Intra-subj. variation) (Ref. [1],[2])

$$b = a + [\, V_{\text{inter}} + \underline{V_{\text{intra}}} \,]$$

- ✓ Expression
- ✓ Illumination
- ✓ Occlusion
- ✓ ……



$$\Longrightarrow \quad b = a + V_{\text{intra}}$$

If a & b are from the same subj.

- Face Variations for different subjects share much similarity.

$\Longrightarrow$ Learn generic intra-class variations from an additional training set. (Ref. [3],[4])

- Model the illumination variations with a subspace model.

$$\Longrightarrow \quad b = a + V_{\text{intra}} = a + \boxed{C}y$$

[Belhumeur et al. '97, Basri & Jacobs '03]

Reference:

[1] B. Moghaddam, *et al.*, Bayesian face recognition, Patter Recognition, 2000. [2] Dong Chen, et al., Bayesian Face Revisited: A Joint Formulation, in ECCV 2012.

[3] Weihong Deng, *et al.*, Extended SRC, IEEE T-PAMI, 34(9), 2012. [4] X. Cao, *et al.*, A Practical Transfer Learning Algorithm for Face Verification, in CVPR 2013.

Sparse Illumination Transfer

Core Idea: Compensate the missing illumination information.

- Model the illumination variations with a subspace model.

$$b = a + V_{\text{intra}} = a + \boldsymbol{C}y$$

Questions:

$$I = \mathbf{B}s \Longrightarrow b - a = \boldsymbol{C}y$$

✓ Why linear model? — Illumination Cone (ref[1]) + Lambert Illumination Model

✓ What does $\boldsymbol{C}$ model? — Illumination Directions

✓ What should y be like? — Sparse!

Reference:

A. S. Georghiades, Peter N. Belhumeur, Dvid J. Kriegman, “From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose”, IEEE T-PAMI, 23(6), 2001.

Sparse Illumination Transfer

中国科学技术大学 University of Science and Technology of China

Definition: SIT is an image-based Re-rendering method that can transfer the illumination information from a source image ***b*** to a target image **a**, so that **a** and **b** are under the same illumination condition.

Let ***a*** and ***b*** be two well-aligned images, SIT contains two steps:

Step 1: $\boldsymbol{y}^* = \arg\min_{\boldsymbol{y},\boldsymbol{e}} \|\boldsymbol{y}\|_1 + \lambda \|\boldsymbol{e}\|_1, \quad \text{subj. to} \quad \boldsymbol{b} - \boldsymbol{a} = C\boldsymbol{y} + \boldsymbol{e}$

where ***C*** is the additional illumination dictionary

Step 2: $\tilde{\boldsymbol{a}} \doteq (\boldsymbol{a} + C\boldsymbol{y}^*)$

Example of SIT Results

Sparse Illumination Transfer

Definition: SIT is an image-based Re-rendering method that can transfer the illumination information from a source image ***b*** to a target image **a**, so that **a** and **b** are under the same illumination condition.

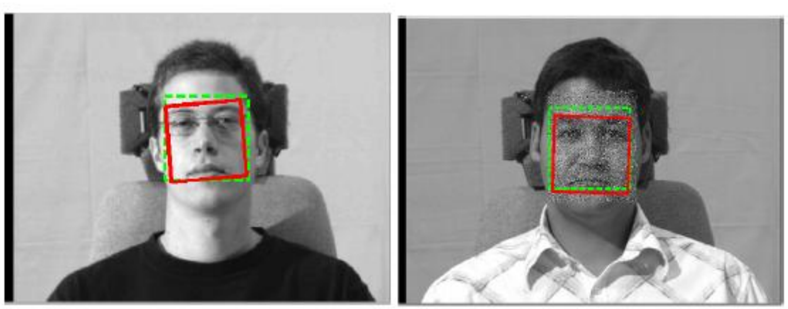
- Let **a** and **b** be two misaligned images, we solve

Step 1: $$\arg\min_{\tau,\boldsymbol{y},\boldsymbol{e}} \|\boldsymbol{y}\|_1 + \lambda\|\boldsymbol{e}\|_1,$$

$$\text{subj. to} \quad \underline{\boldsymbol{b}\circ\tau} - \boldsymbol{a} = C\boldsymbol{y} + \boldsymbol{e}$$

Taylor expand it, and get a linear convex problem! (Ref.[1])

Step 2: $$\tilde{a} \doteq (\boldsymbol{a} + C\boldsymbol{y}) \circ \tau^{-1}$$



Single-sample alignment results

Example of SIT results

Reference:
[1] A. Wagner, et al., Towards a practical face recognition: Robust pose and illumination via Sparse Representation, PAMI 34(2), 2012.

Single Sample Face Recog.

University of Science and Technology of China

- Let $A = [a_1, a_2, \ldots, a_k]$ be gallery images of k subjects, and b be a well-aligned query image.

$$b = a_i + V_{\text{intra}} + e$$

$$\Longrightarrow \quad b = \boldsymbol{A} x + C y + e$$

Ideally, only one element is non-zero. So, both x and y are sparse.

Therefore, we can do face recognition by solving the following problem.

$$\begin{aligned} (\boldsymbol{x}^\star, \boldsymbol{y}^\star) &= \arg\min_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{e}} \|\boldsymbol{x}\|_1 + \|\boldsymbol{y}\|_1 + \lambda \|\boldsymbol{e}\|_1 \\ \text{subj. to} \quad & \boldsymbol{b} = A\boldsymbol{x} + C\boldsymbol{y} + \boldsymbol{e}. \end{aligned}$$

This is the problem which ESRC solves. Note here that, b must be well-aligned!

Reference:

[1] Weihong Deng, *et al.*, Extended SRC: Undersampled Face Recognition via Intraclass Variant Dictionary, PAMI, 34(9), 2012.

Single Sample Face Recog.

- Let A = [$a_1, a_2, \ldots, \boldsymbol{a}_k$] be gallery images of k subjects, and b be a misaligned query image.

$$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\tau}) = \arg\min_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{e},\tau} \|\boldsymbol{x}\|_1 + \|\boldsymbol{y}\|_1 + \lambda\|\boldsymbol{e}\|_1$$
$$\text{subj. to} \quad \boldsymbol{b} \circ \tau = A\boldsymbol{x} + C\boldsymbol{y} + \boldsymbol{e}.$$

$$\Downarrow$$

Step 1:
$$(\hat{\tau}_i, \hat{x}_i, \hat{\boldsymbol{y}}_i) = \arg\min_{\tau_i, x_i, \boldsymbol{y}_i, \boldsymbol{e}} \|\boldsymbol{y}_i\|_1 + \lambda\|\boldsymbol{e}\|_1$$
$$\text{subj. to} \quad \boldsymbol{b} \circ \tau_i = \boldsymbol{a}_i x_i + C\boldsymbol{y}_i + \boldsymbol{e}.$$

Step 2:
$$\tilde{\boldsymbol{a}}_i \doteq (\boldsymbol{a}_i + \underline{C\boldsymbol{y}_i}) \circ \tau_i^{-1}$$

Step 3:
$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x},\boldsymbol{e}} \|\boldsymbol{x}\|_1 + \lambda_1\|\boldsymbol{e}\|_1,$$
$$\text{subj. to} \quad \boldsymbol{b} = \tilde{A}\boldsymbol{x} + \boldsymbol{e}$$

where $\tilde{A} = [\tilde{\boldsymbol{a}}_1, \cdots, \tilde{\boldsymbol{a}}_L]$

Experiments

Data Settings:

- Training images are from CMU Multi-PIE session 1, including 249 subjects. Each subject randomly selects one image under arbitrary illumination.
- Testing images are from CMU Multi-PIE session 1 & session 2. For session 1, each subject randomly selects one image, and all subjects are used. For session 2, each subject randomly selects two images, and only 166 shared subjects from session 1 are used.
- Illumination Variation Dictionary comes from YaleB. We only use the first subject with 65 aligned frontal images (64 illuminations + 1 ambient illumination).

Experiments --- Face Alignment

中国科学技术大学
University of Science and Technology of China

Align a test image *b* to a training image *a*:

$$\arg\min_{\tau, \boldsymbol{y}, \boldsymbol{e}} \|\boldsymbol{y}\|_1 + \lambda \|\boldsymbol{e}\|_1,$$

$$\text{subj. to} \quad \boldsymbol{b} \circ \tau - \boldsymbol{a} = C\boldsymbol{y} + \boldsymbol{e}$$

Criterion:

$$\left| \|\boldsymbol{e}\|_1 - \|\boldsymbol{e}_0\|_1 \right| \leq 0.01 \|\boldsymbol{e}_0\|_1$$

Comparison methods:

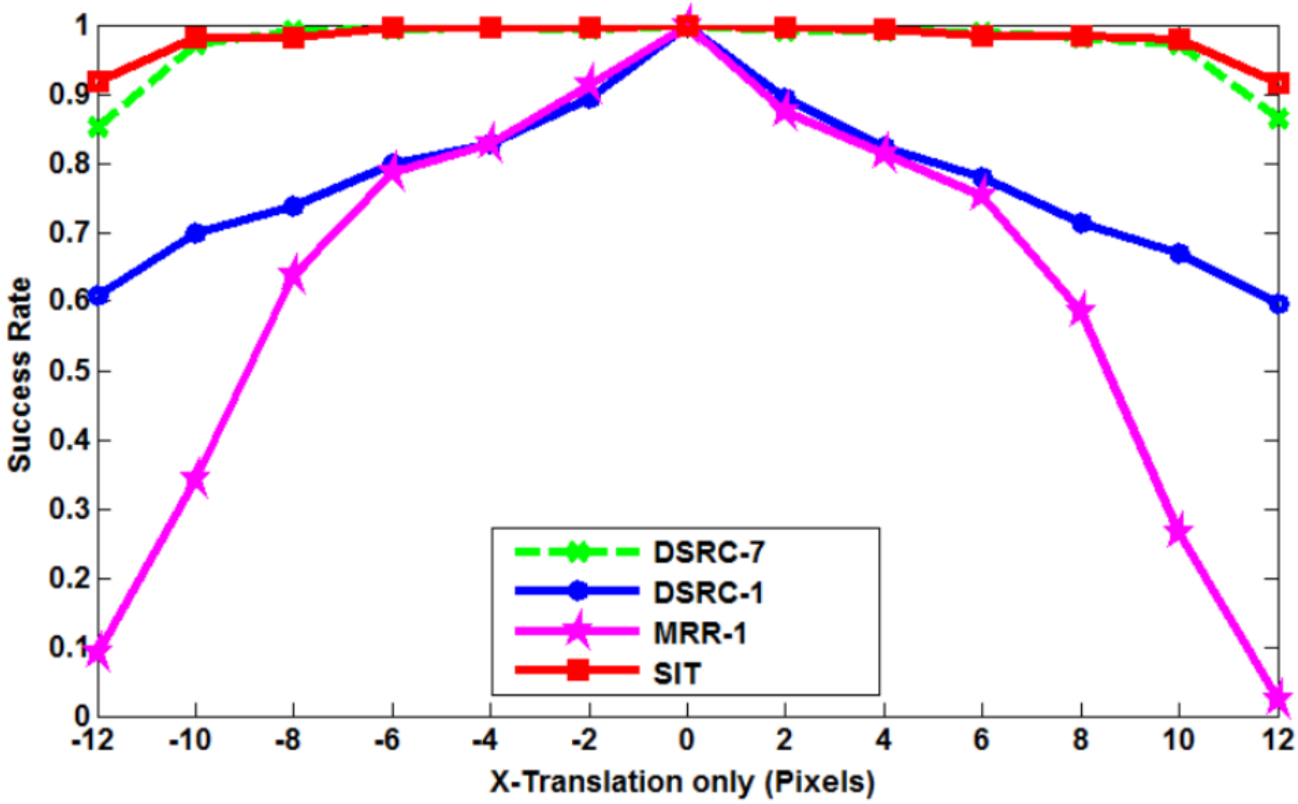
[1] A. Wagner, et al., Toward a practical face recognition: Robust pose and illumination via sparse representation. PAMI, 34(2), 2012. ➔ **DSRC**

[2] M. Yang, L. Zhang and D. Zhang. Efficient misalignment-robust representation for real-time face recognition. ➔ **MRR**

Experiments --- Face Alignment

中国科学技术大学
University of Science and Technology of China

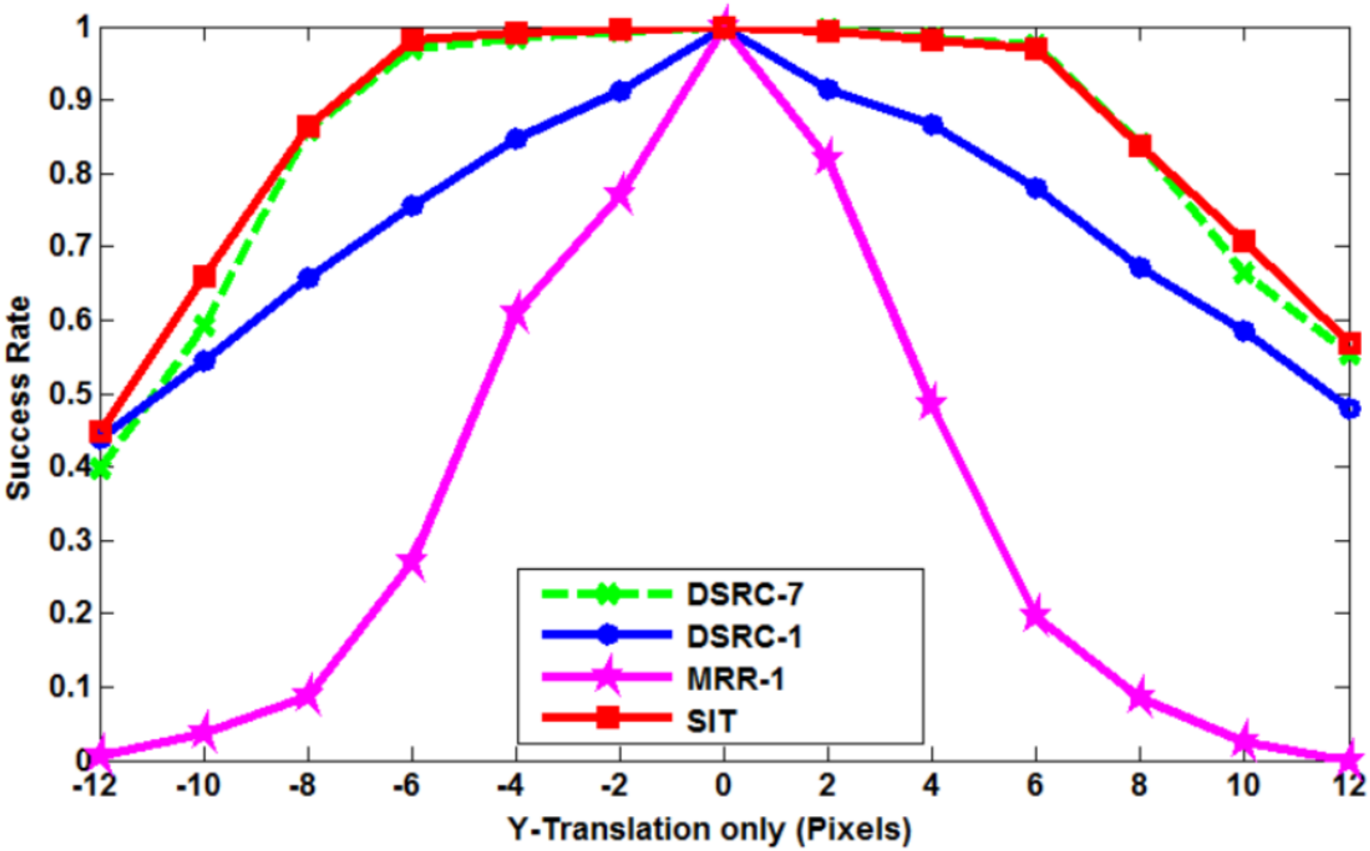
- The translation ranges from [-12, 12] pixels with a step of 2 pixels.



Simulation on 2D Alignment with *X*-translation only

Experiments --- Face Alignment

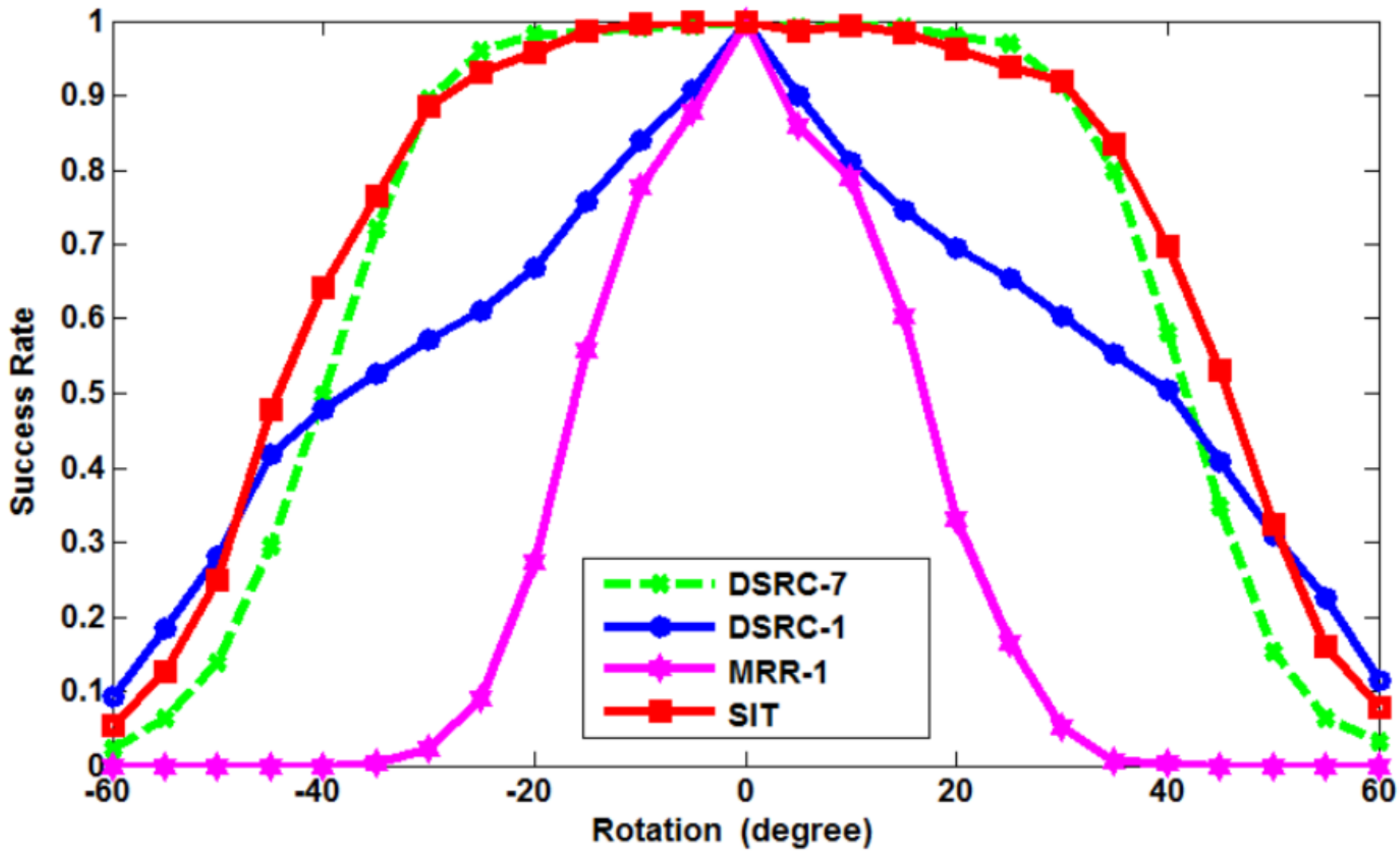
- The translation ranges from [-12, 12] pixels with a step of 2 pixels.



Simulation on 2D Alignment with *Y*-translation only

Experiments --- Face Alignment

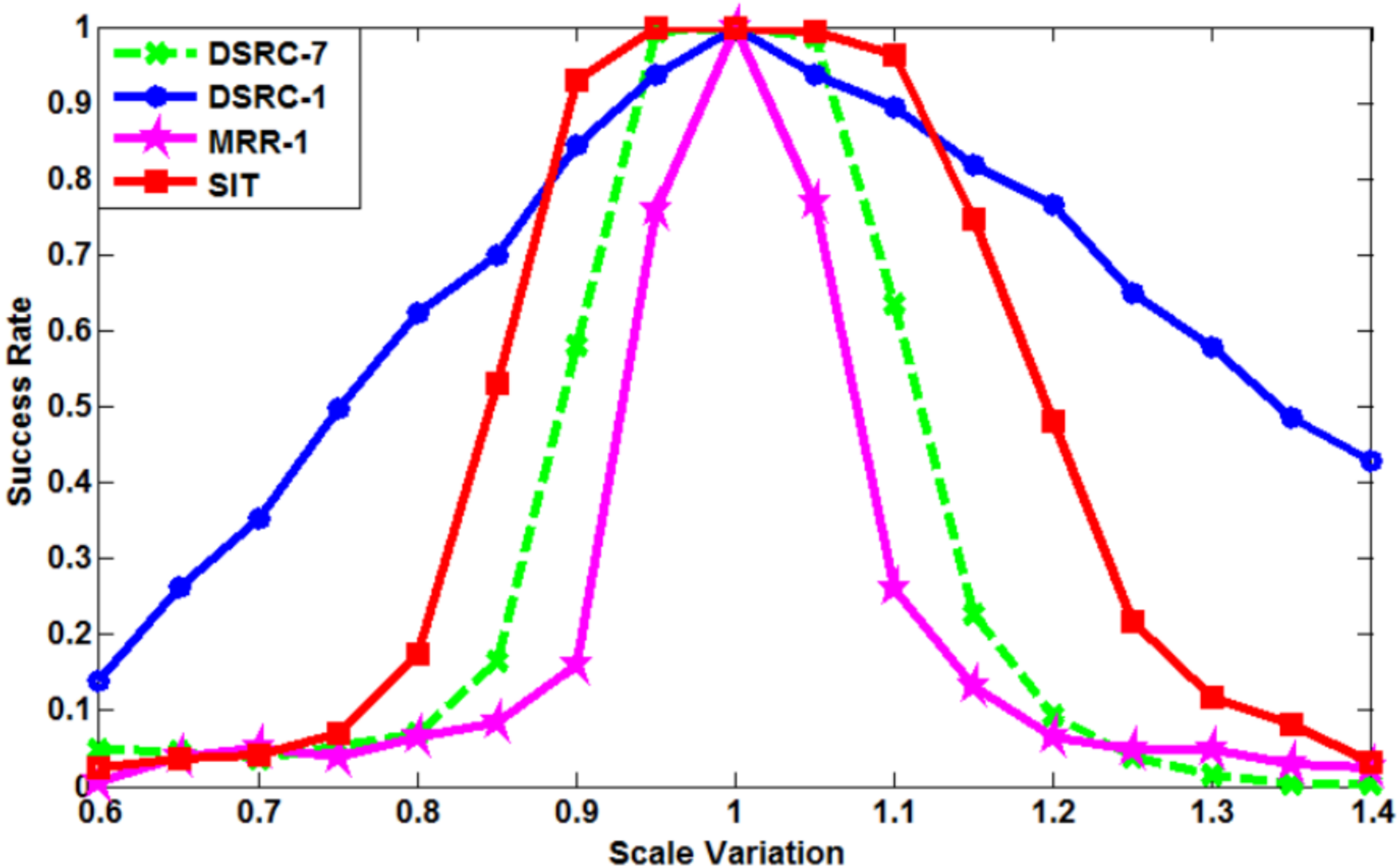
- The rotation ranges from [-60, 60] degree with a step of 20 degree.



Simulation on 2D Alignment with *rotation*-translation only

Experiments --- Face Alignment

- The scale ranges from [0.6, 1.4] degree with a step of 0.1.



Simulation on 2D Alignment with *scale*-translation only

Experiment – Face Alignment

- SIT works well under a broad range of 2D deformation, particularly when the translation in x or y direction is less than 20% of the eye distance (10 pixels) and when the in-plane rotation is less than 30◦

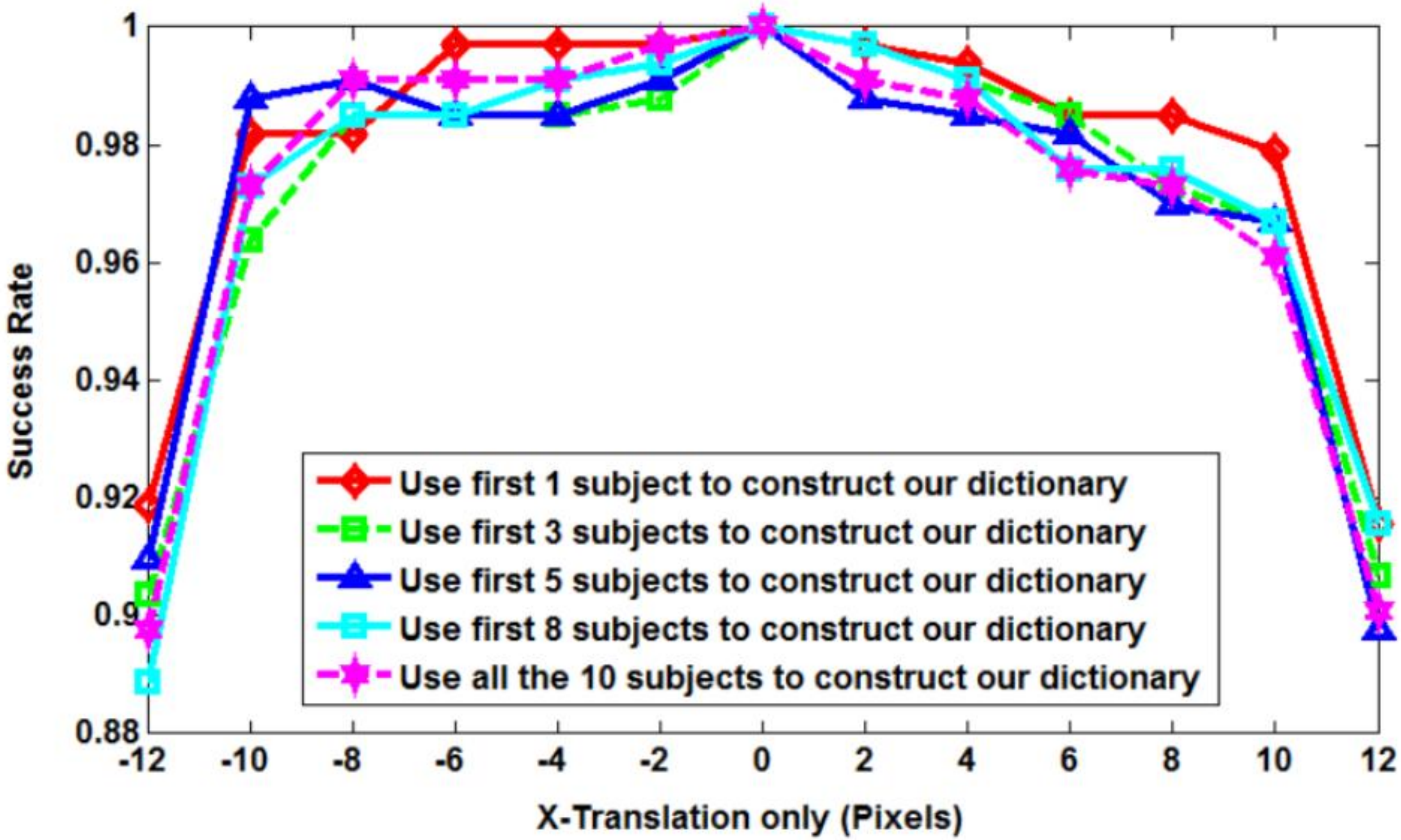
- SIT outperforms both DSRC-1 and MRR-1 when the same setting is used. The obvious reason is that DSRC and MRR were not designed to handle the single-sample scenario.

- SIT silightly outperforms DSRC-7, where DSRC-7 has access to seven training images of different illumination conditions. Furthermore, the SIT dictionary is derived from a single subject class from the unrelated YaleB database. It validates that illumination examples of a well-chosen subject are sufficient for SIT alignment.

Simulation on 2D Alignment under four type of deformation

Experiments --- Face Alignment

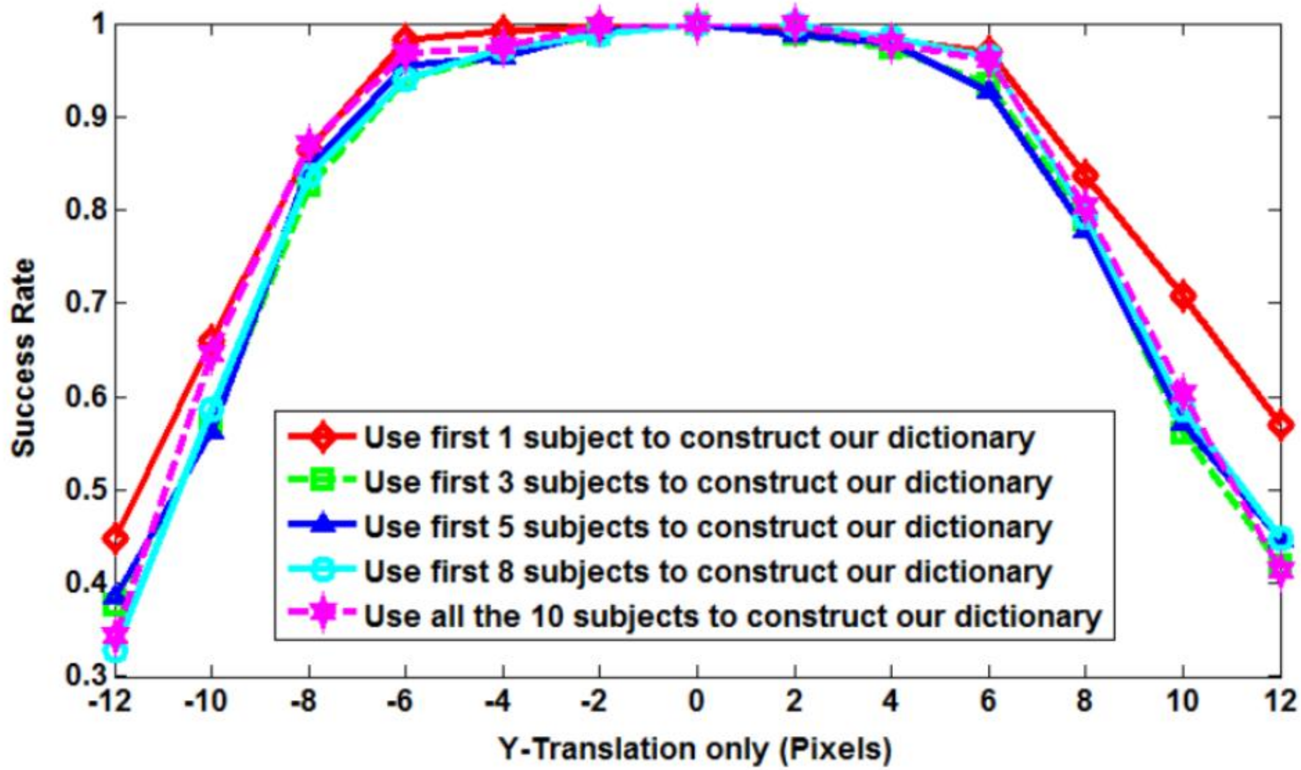
- Face alignment with a multiple-subject dictionary.



Experiments --- Face Alignment

中国科学技术大学
University of Science and Technology of China

- Face alignment with a multiple-subject dictionary.



Experiments

Data Settings:

- Training images are from CMU Multi-PIE session 1, including 249 subjects. Each subject randomly selects one image under arbitrary illumination.
- Testing images are from CMU Multi-PIE session 1 & session 2. For session 1, each subject randomly selects one image, and all subjects are used. For session 2, each subject randomly selects two images, and only 166 shared subjects from session 1 are used.
- Additional illumination Dictionary comes from YaleB. YaleB contains 10 subjects with 9 poses. Given a pose, each subject has 64 illuminations plus one ambient illumination. We only use the frontal images under various illumination.

Compared methods:

- Standard SRC (TPAMI 2009)
- Extended SRC (TPAMI 2012)
- Deformable SRC (DSRC) (TPAMI 2012)
- MRR: Efficient Misalignment-Robust Representation for Real-Time Face Recognition (ECCV 2012)

Experiments-- Single-Sample Recog

1. Demonstrate the improvement of SRC and ESRC with the illumination transfer.

Table 1. Single-sample recognition accuracy via manual alignment.

Method	Session 1 (%)	Session 2 (%)
SRC_M	88.0	53.6
$ESRC_M$	89.6	56.6
SRC_M + SIT	91.6	59.0
$ESRC_M$ + SIT	**93.6**	**59.3**

2. Comparison in the full pipeline of alignment plus recognition.

Table 2. Single-sample alignment + recognition accuracy.

Method	Session 1 (%)	Session 2 (%)
DSRC	36.1	35.7
MRR	46.2	34.6
SIT	**81.9**	**68.7**

3. Robustness under Random Corruption

To limit this variability, in this experiment, we use Multi-PIE Session 1 for both training and testing, although the images should never overlap. We use all the subjects in Session 1 as the training and testing sets. For each subject, we randomly select one frontal image with arbitrary illumination for testing. Various levels of image corruption from 10% to 40% are randomly generated in the face region.

Table 3. Recognition rates (%) under various random corruption.

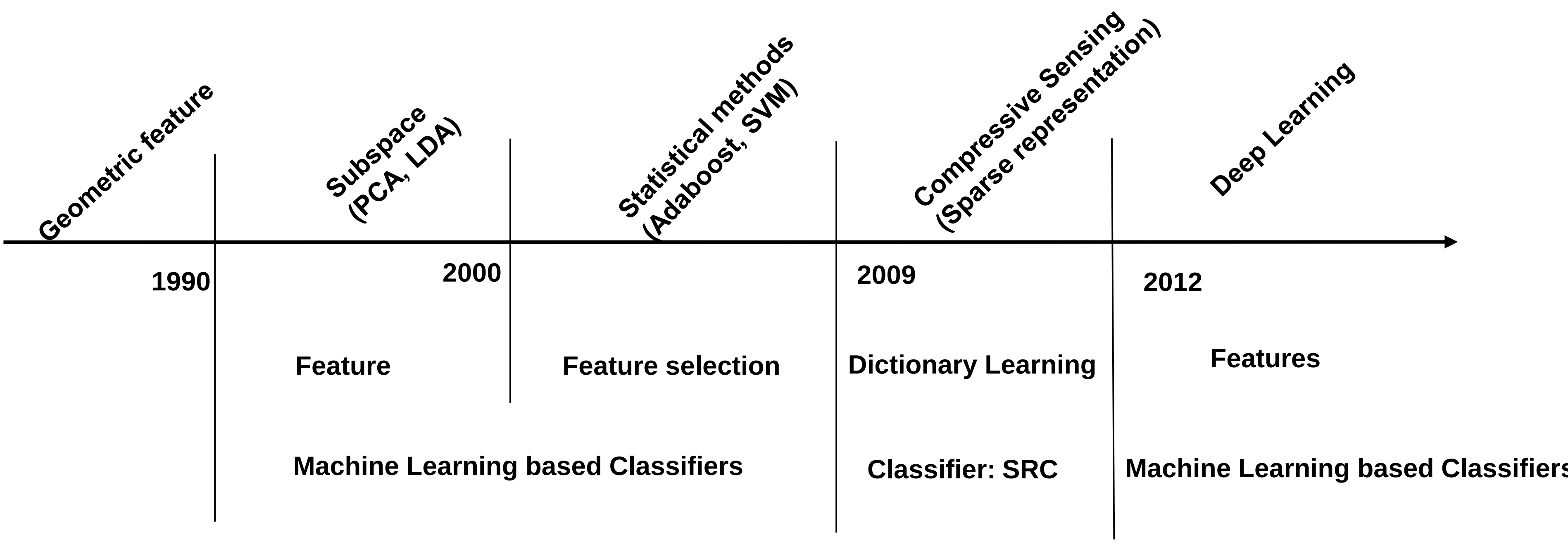
Corruption	10%	20%	30%	40%
DSRC	32.9%	31.7%	28.9%	24.1%
MRR	24.9%	14.5%	11.7%	9.2%
SIT	**75.5%**	**72.7%**	**67.1%**	**55.8%**

Conclusion

- We have introduced a new image-based re-rendering method, Sparse Illumination Transfer.
- Based on SIT, we have also presented a novel face recognition algorithm specifically designed for single-sample alignment and recognition.

Discussion: SRC vs DL

Why Sparse representation based Classification (SRC) ?



Advantages of SRC:

* **Accuracy → Theory Guarantee**
* **Scalable → Solve linear convex problem**
* **Robust** → Sparse Error

Model based vs. data-driven based

Discussion: dictionary learning

How to learning an efficient dictionary?

$$\min_{V,C,S,E} \|S\|_0 + \|E\|_F \text{ subj. to } D = V \otimes \mathbf{1}^T + CS + E,$$

Top row is the Ad-hoc illumination variation dictionary. Second row is the learned illumination variation dictionary.

Discussion: dictionary Learning

Three dictionary learning methods:

1. Ad-hoc method was used in our CVPR 2013 paper.
2. SVDL was proposed by Meng Yang in ICCV 2013.
3. SILT was proposed in our IJCV 2015 paper.

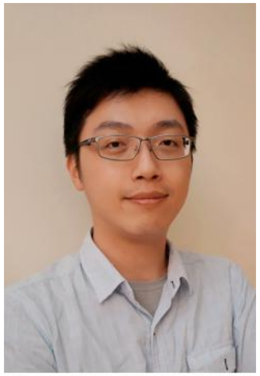
Table 1 Single-sample recognition accuracy via manual alignment. The atom size is fixed to 80

Method	Session 1 (%)	Session 2 (%)
SRC_M	88.0	53.6
$ESRC_M$	89.6	56.6
SILT + SRC_M	92.8	59.0
SILT + $ESRC_M$	**93.2**	**59.3**
$SVDL_M$	70.3	41.6

Yang, M., Gool, L.V., & Zhang, L. (2013). Sparse variation dictionary learning for face recognition with a single training sample per person. In *Proceedings of the IEEE international conference on computer vision*.

Thank you!

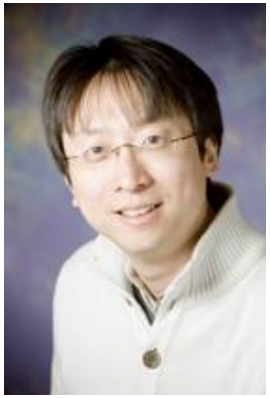
Question?



Tsung-Han Chan

Allen Y. Yang

Zihan Zhou



Yi Ma

Shankar S. Sastry