Human identification at a distance via gait recognition

Liang Wang

Center for Research on Intelligent Perception and Computing (CRIPAC)

National Lab of Pattern Recognition (NLPR)

Institute of Automation, Chinese Academy of Sciences (CASIA)

Outline

1. Introduction and overview

2. Traditional approaches for gait-based human identification
- History and databases
- Gait representation and learning algorithms

3. Deep networks for gait-based human identification
- Cross-view gait based human identification with deep CNNs

4. How to build a practical gait-based human identification system?
- End-to-end deep network for gait segmentation & recognition
- System demo

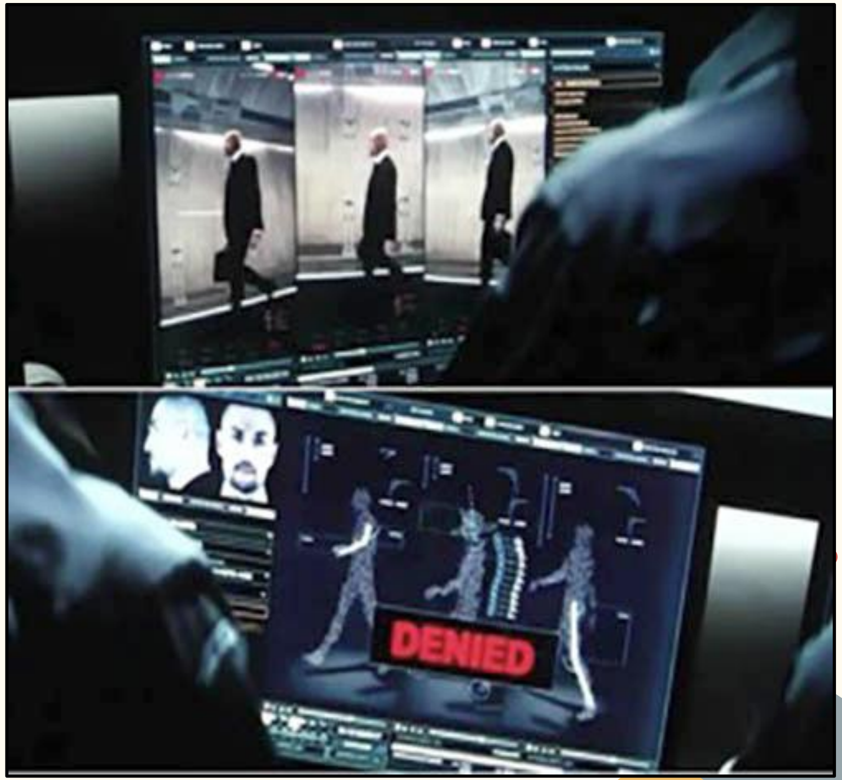
5. Open questions and discussion

Outline

1. Introduction and overview
2. Traditional approaches for gait-based human identification • History and databases • Gait representation and learning algorithms
3. Deep networks for gait-based human identification • Cross-view gait based human identification with deep CNNs
4. How to build a practical gait-based human identification system? • End-to-end deep network for gait segmentation & recognition • System demo
5. Open questions and discussion

What is Gait Recognition?

GAIT is a kind of behavioral biometric feature, whose raw data are video sequences presenting walking people. The goal of gait recognition is to identify people based on their gait features.



Movie “Mission Impossible 5”

Is gait recognition necessary?

Short distance

Cooperative

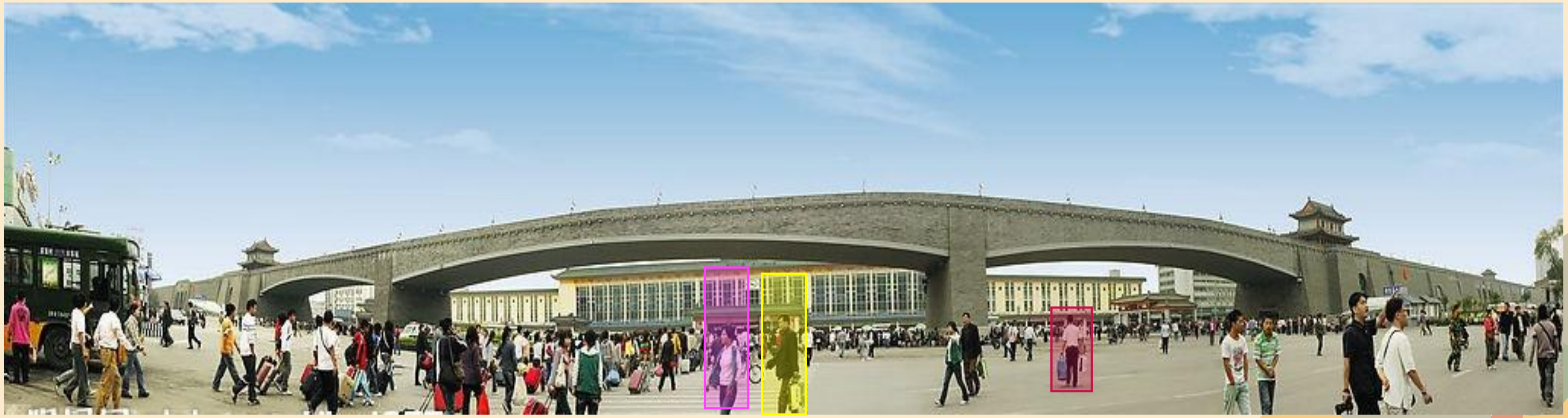
Fingerprint

Iris

Face

Long distance

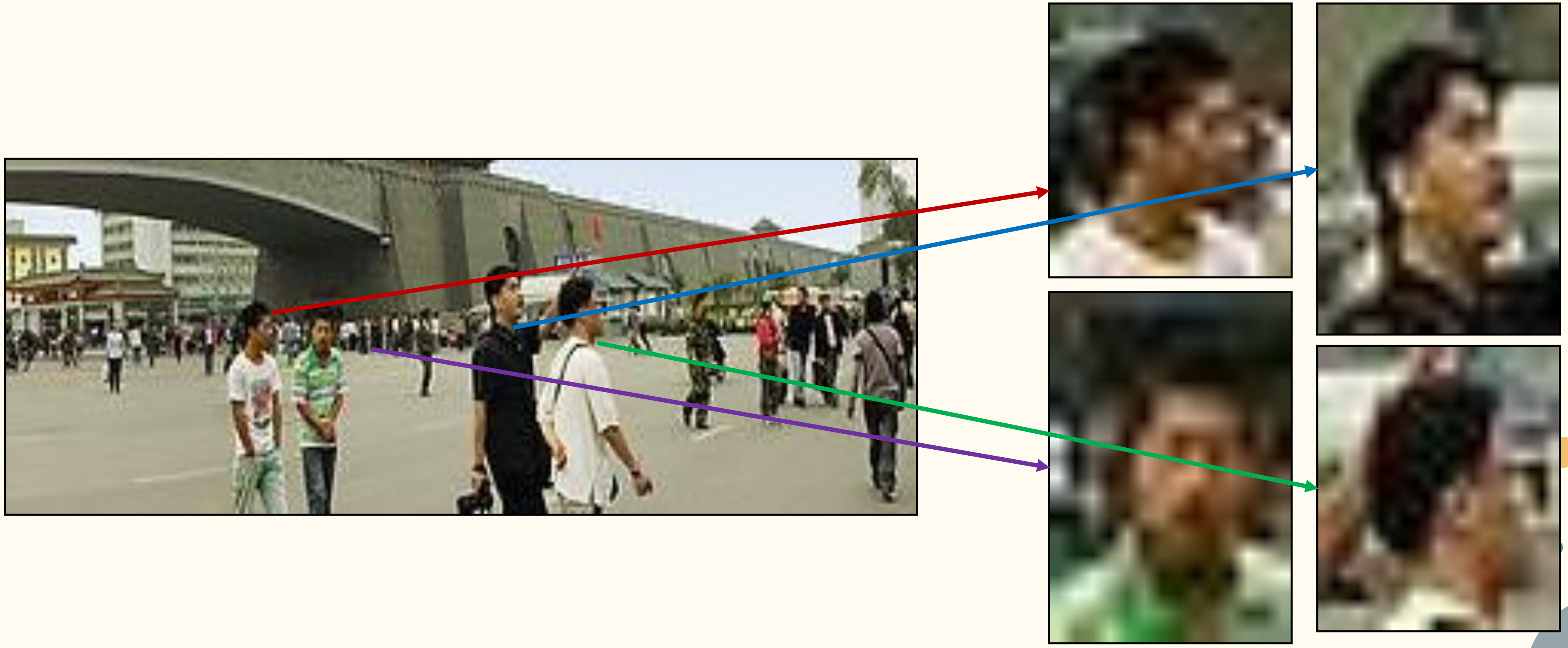
Uncooperative



Gait

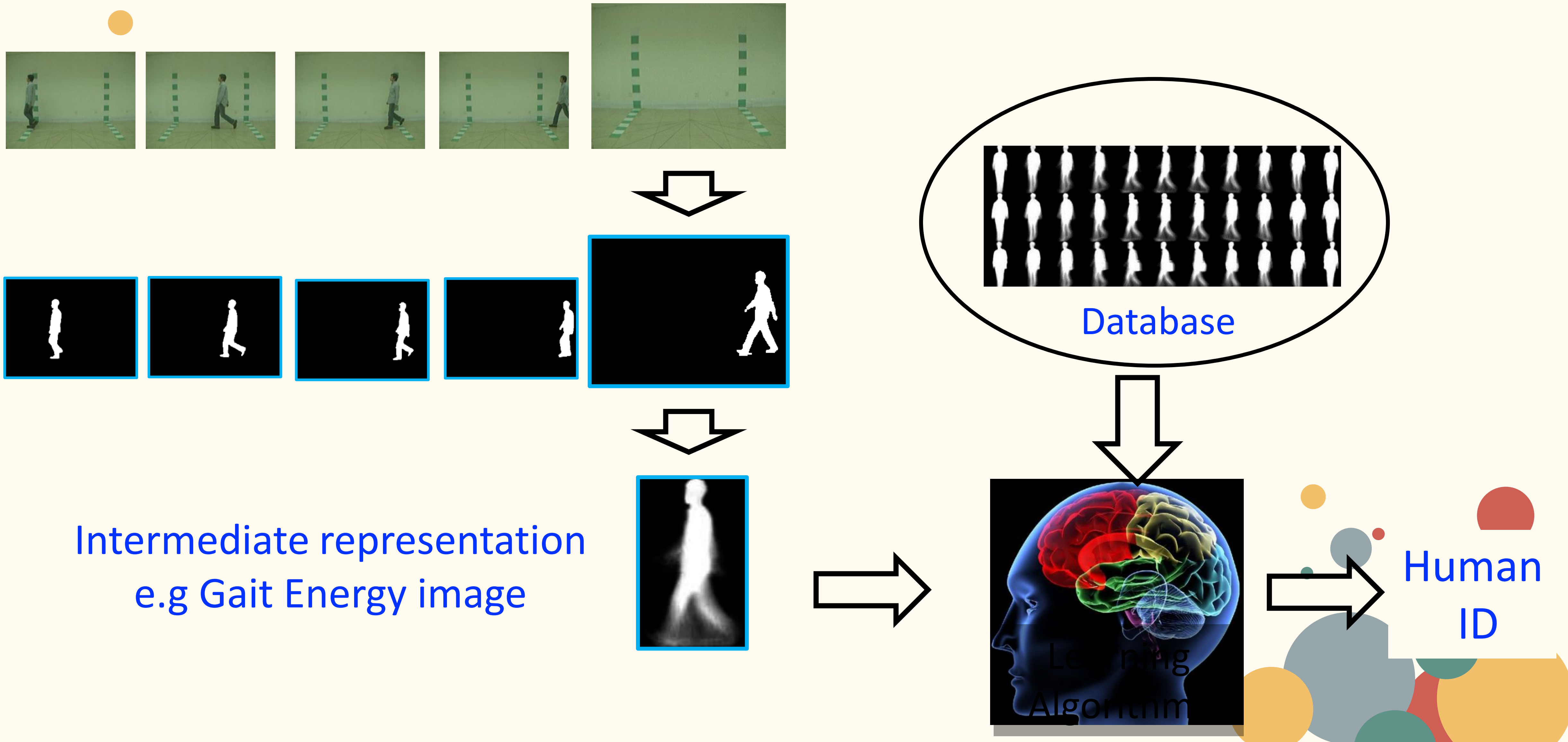
Is gait recognition necessary?

As a biometric, gait is still available at a distance when other biometrics are obscured or at too low resolution. Therefore, we need gait recognition.



Advantages: insensitive to distance, resolution, view, illumination

How does a gait recognition system work?



Applications of gait recognition

Suspect searching


Access control

Airport security

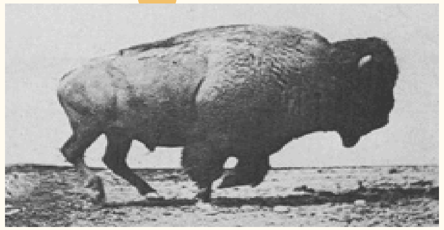
Robotics & Smart home

Outline

1. Introduction and overview
2. Traditional approaches for gait-based human identification • History and databases • Gait representation and learning algorithms
3. Deep networks for gait-based human identification • Cross-view gait based human identification with deep CNNs
4. How to build a practical gait-based human identification system? • End-to-end deep network for gait segmentation & recognition • System demo
5. Open questions and discussion

History of gait recognition:

[Slide Credit: Mark Nixon]

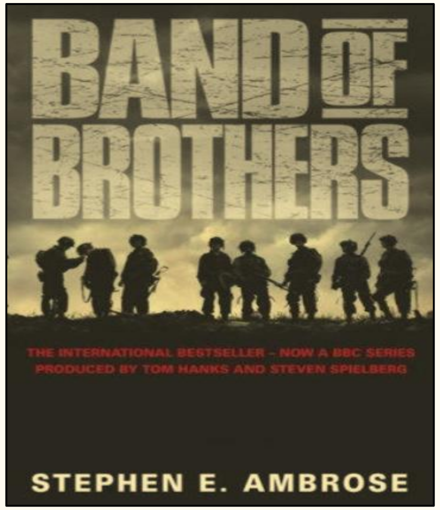


~350 BC 1500s 1600s

- Aristotle (~350 BC): The first to analyze gait. “On the gait of animals”
- Leonardo da Vinci (~1500): movement sketches
- Borelli (1600s): Father of biomechanics, study the mechanical principles of locomotion. *‘De Motu Animalium’*

History of gait recognition:

[Slide Credit: Mark Nixon]



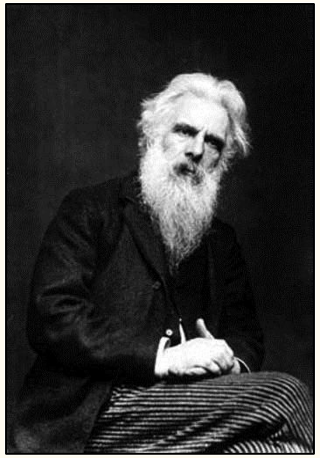
1600s

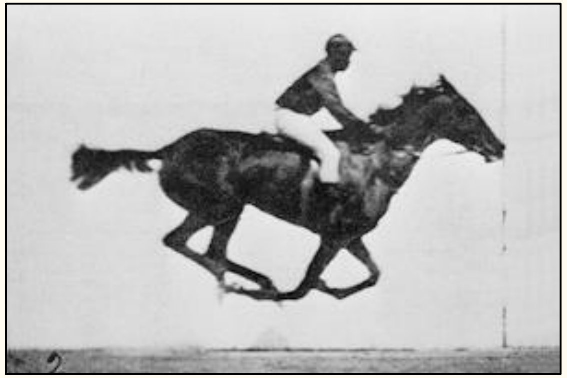
Shakespeare observed recognition:

- “High’st Queen of state; Great Juno comes; I know her by her gait” [The Tempest]
- “For that John Mortimer....in face, in gait in speech he doth resemble” [Henry IV/2]

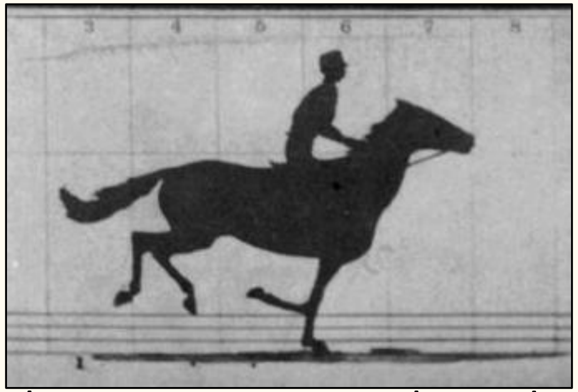
Other literature: e.g. Band of Brothers: “I noticed this figure coming, and I realized it was John Eubanks from the way he walked”

History of gait recognition:





Galloping horse, animated in 2006, using photos by Eadweard Muybridge



The Horse in Motion by Eadweard Muybridge. running at a 1:40 pace. Frames 1-11 used for animation

1800s

Eadweard Muybridge (1830-1904):

- Pioneering work in photographic studies of motion and motion-picture projection.
- Studied horses (1872):whether all four feet of a horse were off the ground at the same time while trotting
- Studied movement (1884)

History of gait recognition:

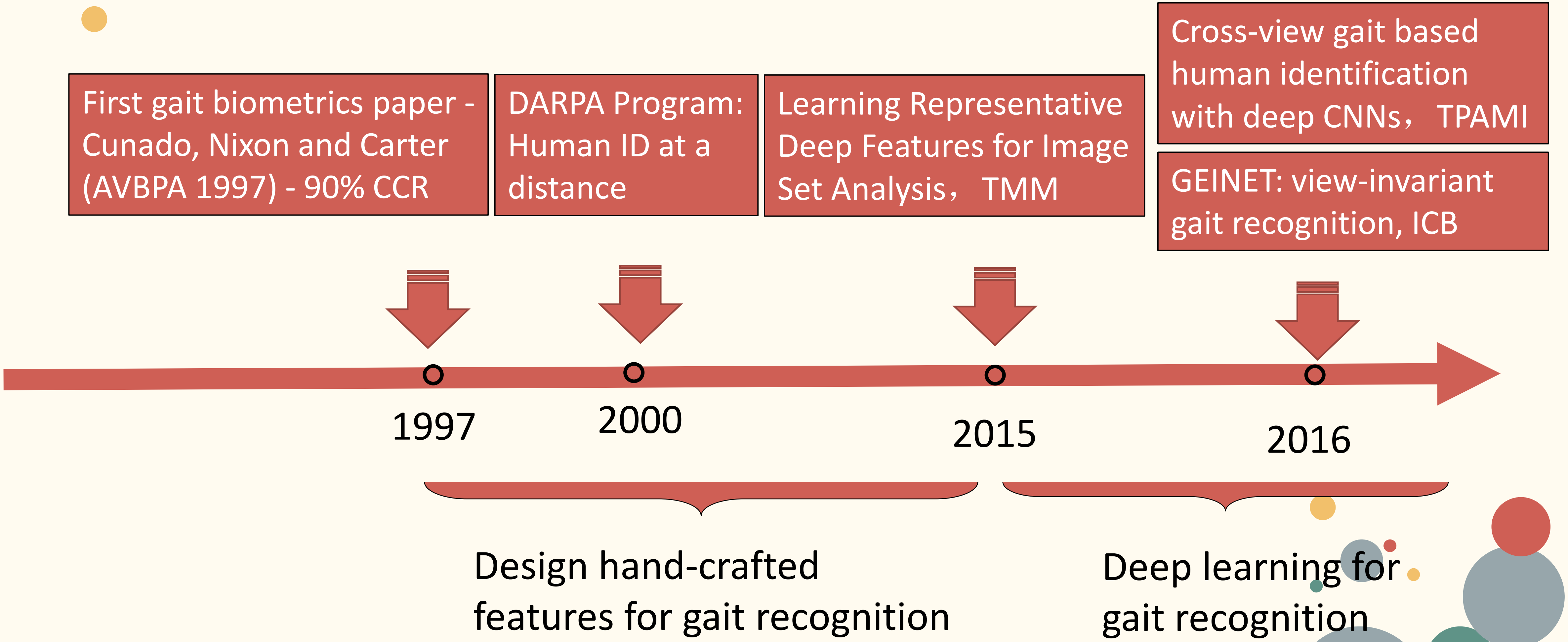
- Murray (1964): Produced standard movement patterns for pathologically normal people, suggesting the uniqueness of gait for individuals. *'Walking Patterns of Normal Man'* *'Gait As a Total Pattern of Movement'.*

1964, 1973, 1977

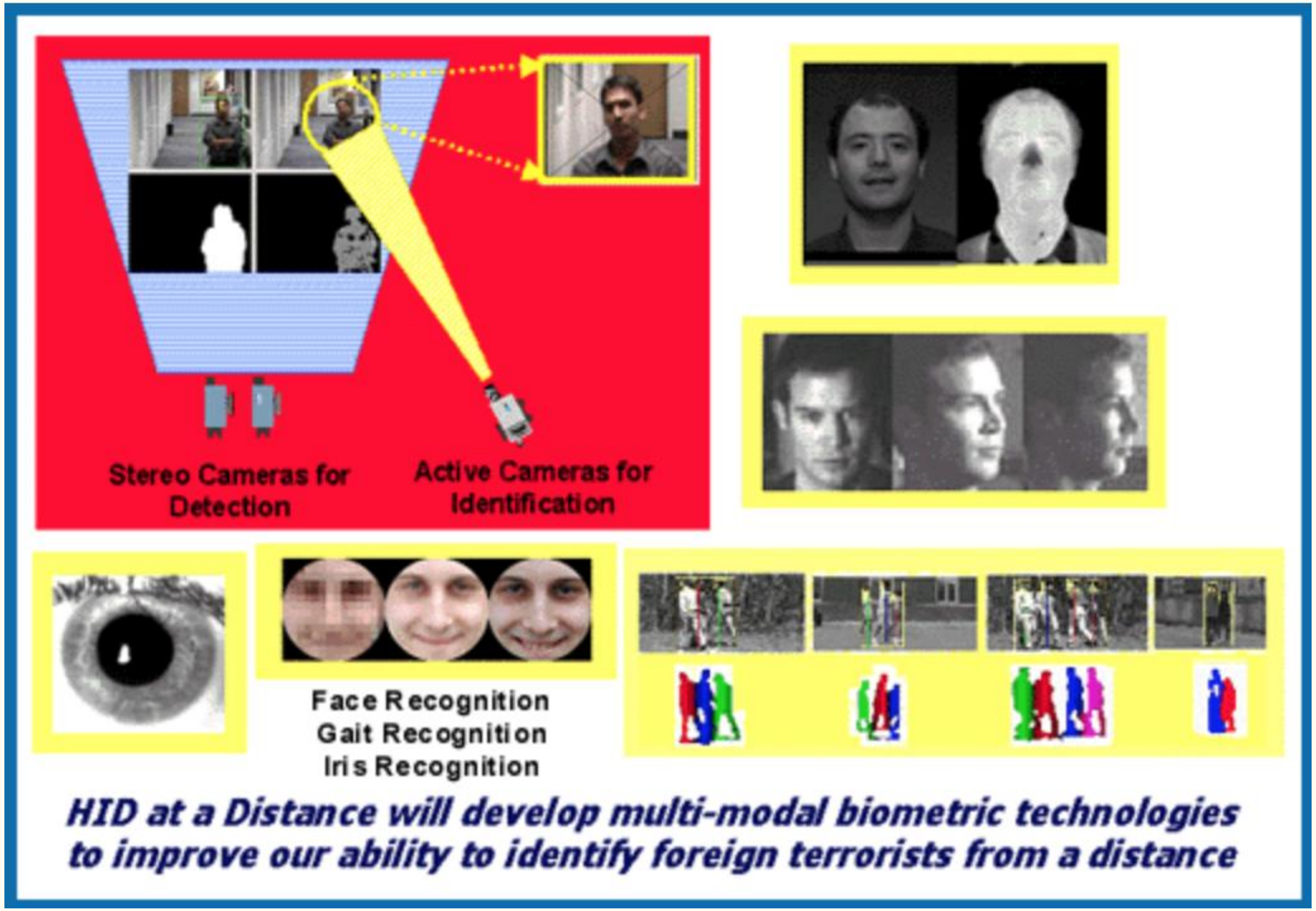
- Johansson(1973): Studied visual perception of motion patterns and suggested that 'biological motion' has far higher complexity than mechanical motions, and presented point-light displays to simulate human gait. *'Visual Perception of Biological Motion and a Model for its Analysis'*

- Cutting & Kozlowski (1977): Announced that humans can recognize friends of a person solely by their gait with 70-80% accuracy.*'Recognizing friends by their walk: Gait perception without familiarity cues'*

History of gait recognition:



DARPA program: Human ID at a distance



The DARPA program motivated the research on gait recognition

Released gait databases

Name	Subjects	Sequences	Covariates	Viewpoints	Indoor(I)/ Outdoor(O)
CMU MoBo (30)	25	600	Y	6	I (Treadmill)
Georgia Tech (31)	15	268	Y	-	O
	18	20	Y	-	-
HID-UMD (32)	25	100	N	1	O
	55	222	Y	2	O
SOTON Small Database (33)	12	-	Y	3	I
SOTON Large Database (34)	115	2,128	Y	2	I/O
SOTON Multimodal (35)	>300	>5,000	Y	12	I
SOTON Temporal (36)	25	2,280	Y	12	I
USF HumanID (23)	122	1,870	Y	2	O
CASIA A (37)	20	240	Y	3	I
CASIA B (38)	124	1,240	Y	11	I
CASIA C (39)	153	1,530	Y	1	O
OU-ISIR, Treadmill A (40)	34	612	Y	1	I (Treadmill)
OU-ISIR, Treadmill B (41)	68	2,764	Y	1	I (Treadmill)
OU-ISIR, Treadmill C (42)	200	200	Y	25	I (Treadmill)
OU-ISIR, Treadmill D (43)	185	370	N	1	I (Treadmill)
OU-ISIR, LP (44)	4,007	7,842	N	2	I
TUM-IITKGP (45)	35	850	Y	1	O
TUM-GAID (46)	305	3,370	Y	1	O
WOSG (47)	155	684	Y	8	O

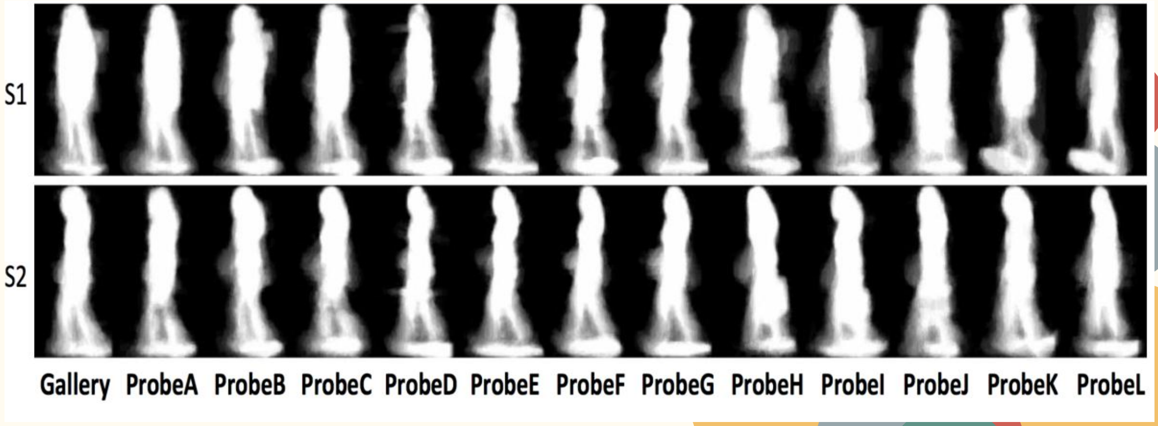
Widely used benchmarks in the community

a) CASIA-B
b) USF HumanID
c) OU-ISIR, Large Population

[Makihara *et al*. 2015]

USF Human ID database

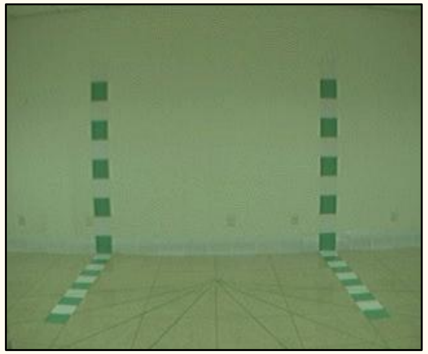
Details	
Indoor/Outdoor	outdoor
# of subjects	122
# of carrying conditions	2 (w/wo briefcase)
# of walking conditions	2 (shoe types)
# of viewpoints	2 (left/right)
# of backgrounds	2 (grass/concrete)
# of time instants	2



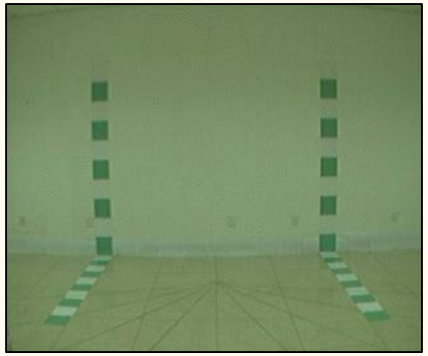
GEIs of two subjects under different conditions. The obtained GEIs are more noisy and of lower quality due to the complex backgrounds

CASIA-B database

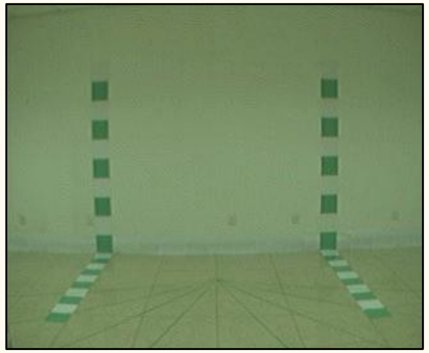
Details	
Indoor/Outdoor	indoor
# of subjects	124
# of carrying/walking conditions	3
# of viewpoints	11



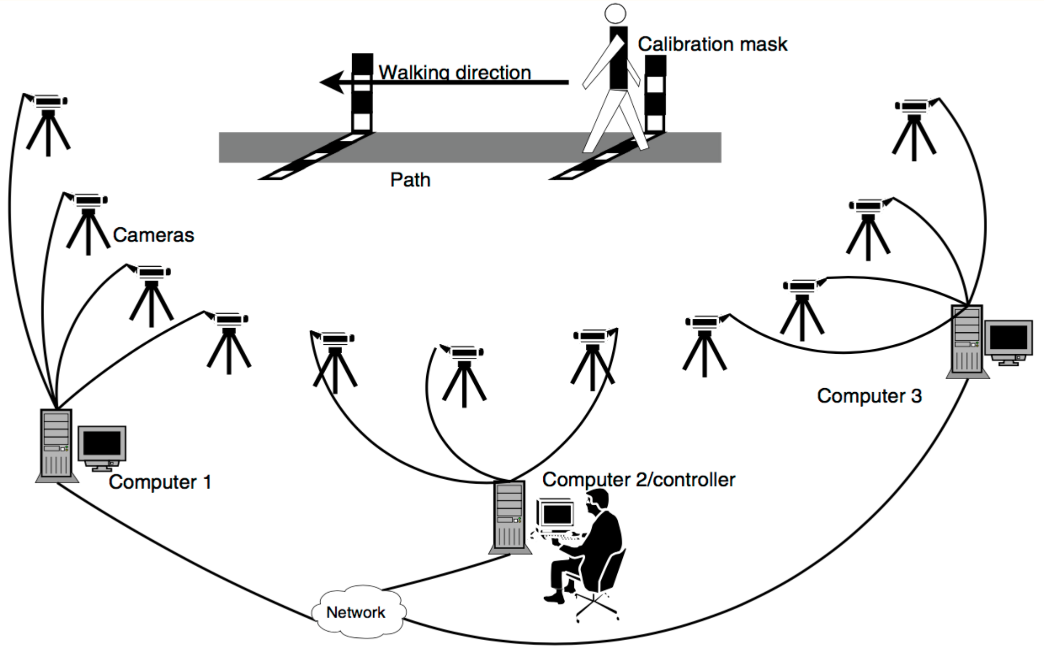
Normal Walk

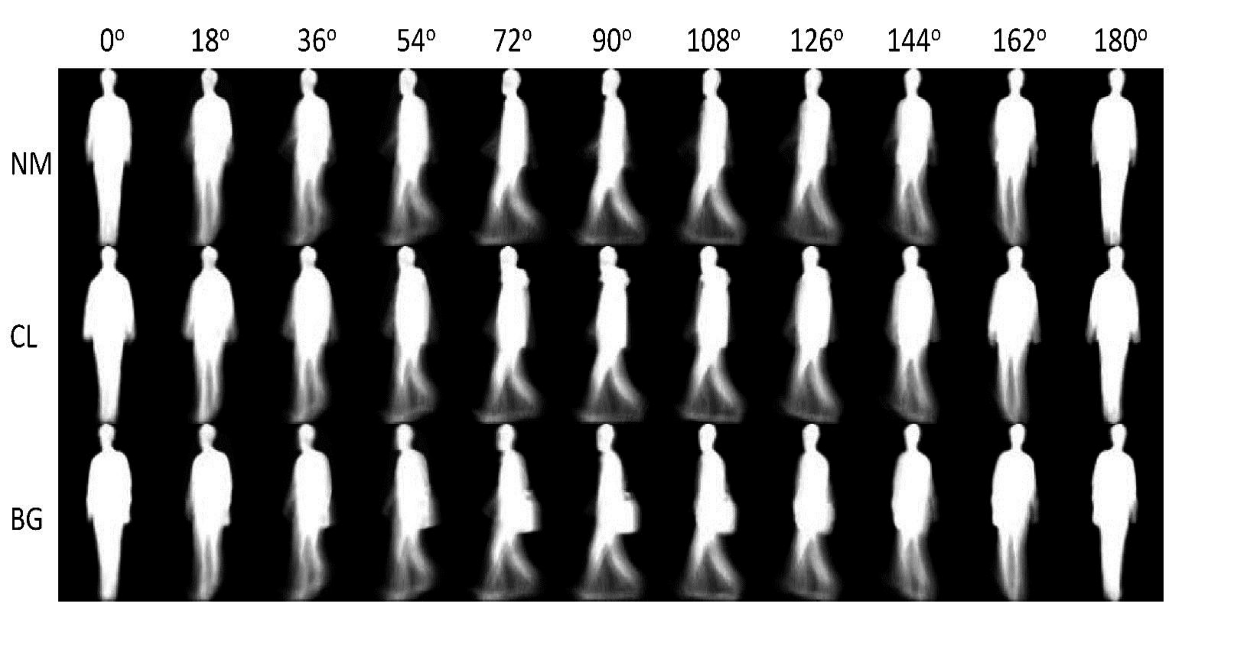


Wearing Coats



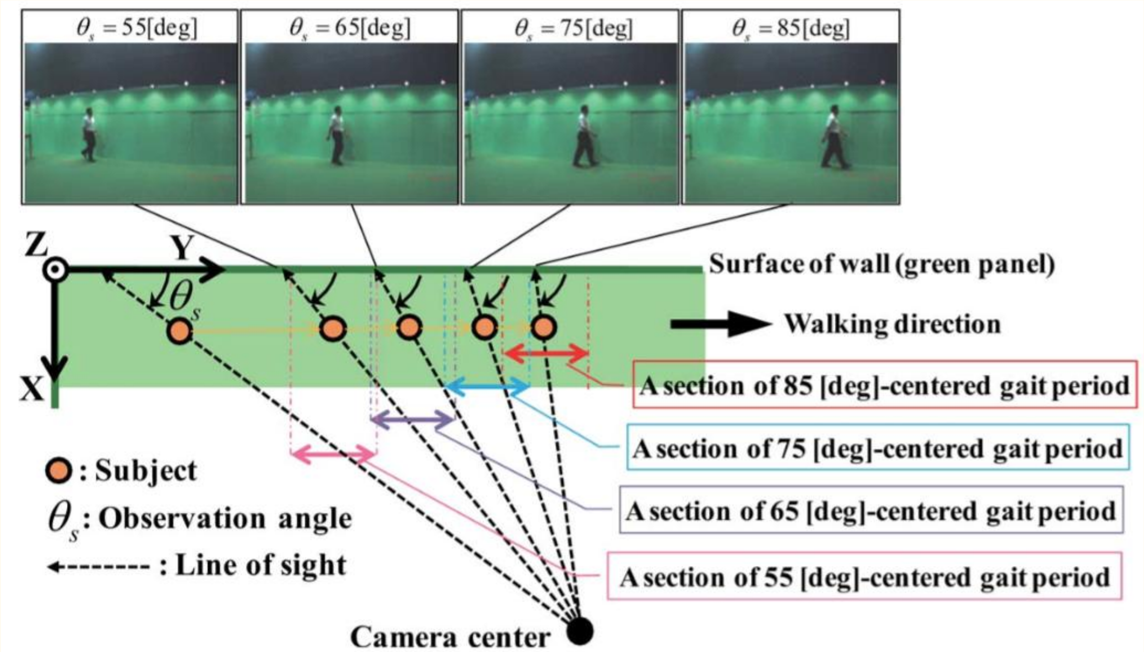
Carrying bags

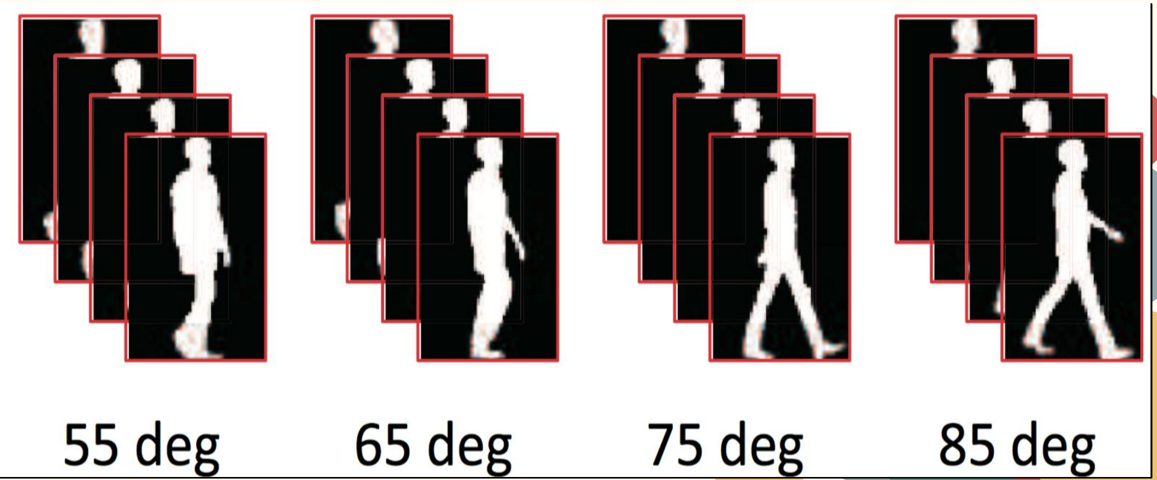




OU-ISIR database, Large population dataset

Details	
Indoor/Outdoor	indoor
# of subjects	4,007(v1), 4,016(v2)
Age range	1-94 years old
# of walking conditions	1
# of viewpoints	4 (55,65,75,85)
# of backgrounds	1



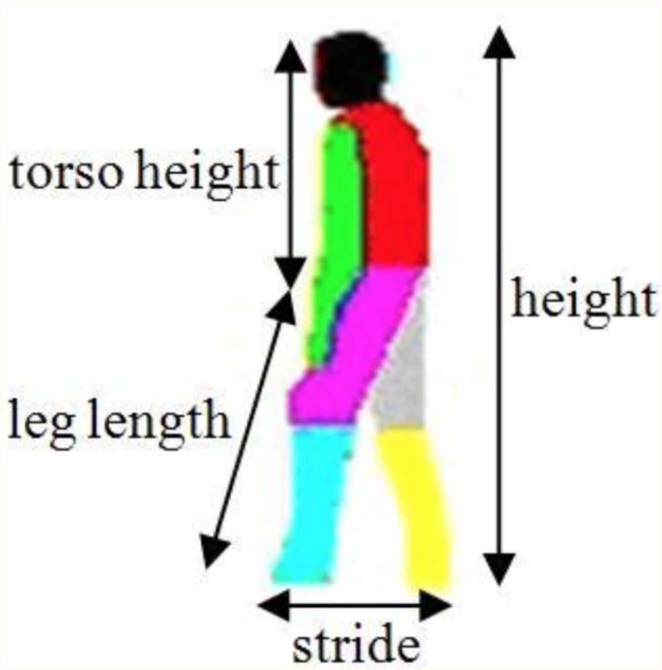
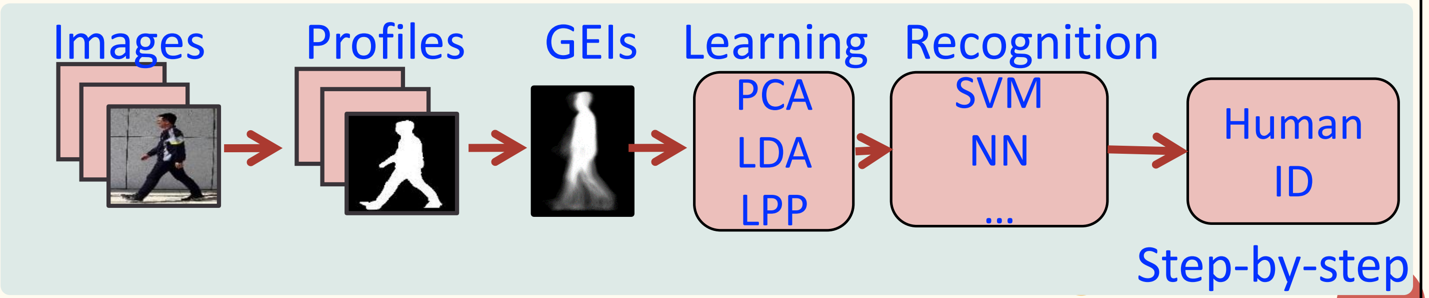


CASIA-HT database (expected to be released early next year)

Details	
Indoor/Outdoor	outdoor
# of subjects	1000
# of carrying conditions	3
# of walking conditions	2
# of viewpoints	13 horizontal, 2 vertical
# of backgrounds/scenarios	2
# of sequences	>760,000

Another super large database for gait recognition [C. Song, Y. Huang, et al.]

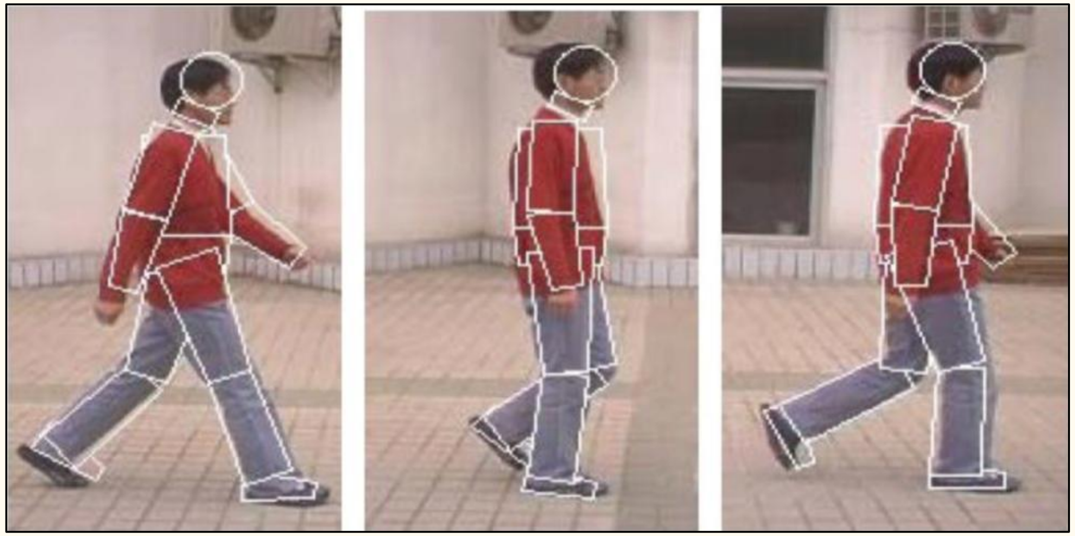
Categories of learning methods for gait recognition

Model-based: use the human body structure	Model free (appearance-based): use the whole motion pattern of the human body
	

Categories of learning methods for gait recognition

Model-based: use the human body structure	Model free (appearance-based): use the whole motion pattern of the human body
• Greater invariant properties and better at handling occlusion, noise, scale and rotation. • Require a high resolution and are not yet very suitable for outdoor surveillance	• Computational efficiency and simplicity • Can handle low-resolution case • Suitable for outdoor surveillance

Model-based approaches: an example



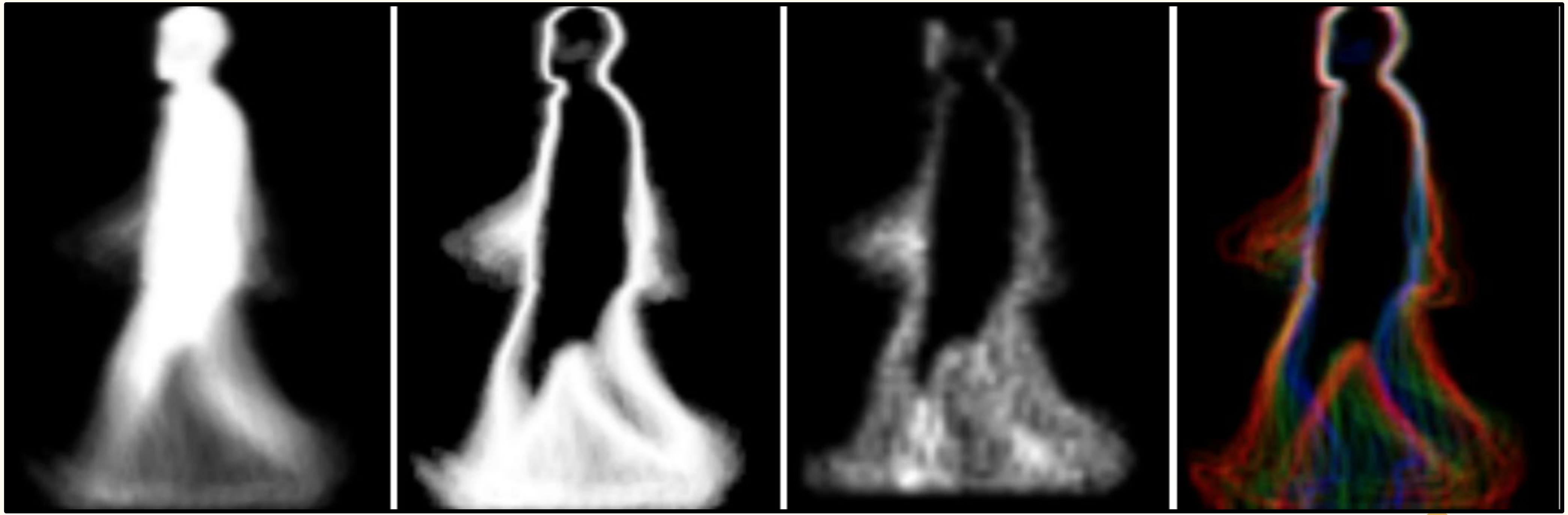
- Fusion of static and dynamic body information.
- The static body information is in a form of a compact representation obtained by Procrustes shape analysis.
- The dynamic information is obtained by a model based approach which tracks the subject and recover joint-angle trajectories of lower limbs.
- Fusion at the decision level used to improve recognition results.

Fusion of Static and Dynamic Body Biometrics for Gait Recognition, Liang Wang, Huazhong Ning, Tieniu Tan, Weiming Hu , ICCV 2003

Model-free approaches: examples

- SVR: “Support vector regression for multi-view gait recognition based on local motion feature selection,” in *CVPR*, 2010.
- TSVD: Multiple views gait recognition using view transformation model based on optimized gait energy image,” in *Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS)*, 2009.
- CMCC: “Cross-view gait recognition using correlation strength,” in *BMVC*, 2010.
- ViDP: “View-invariant discriminative projection for multi-view gait-based human identification,” TIFS 2013

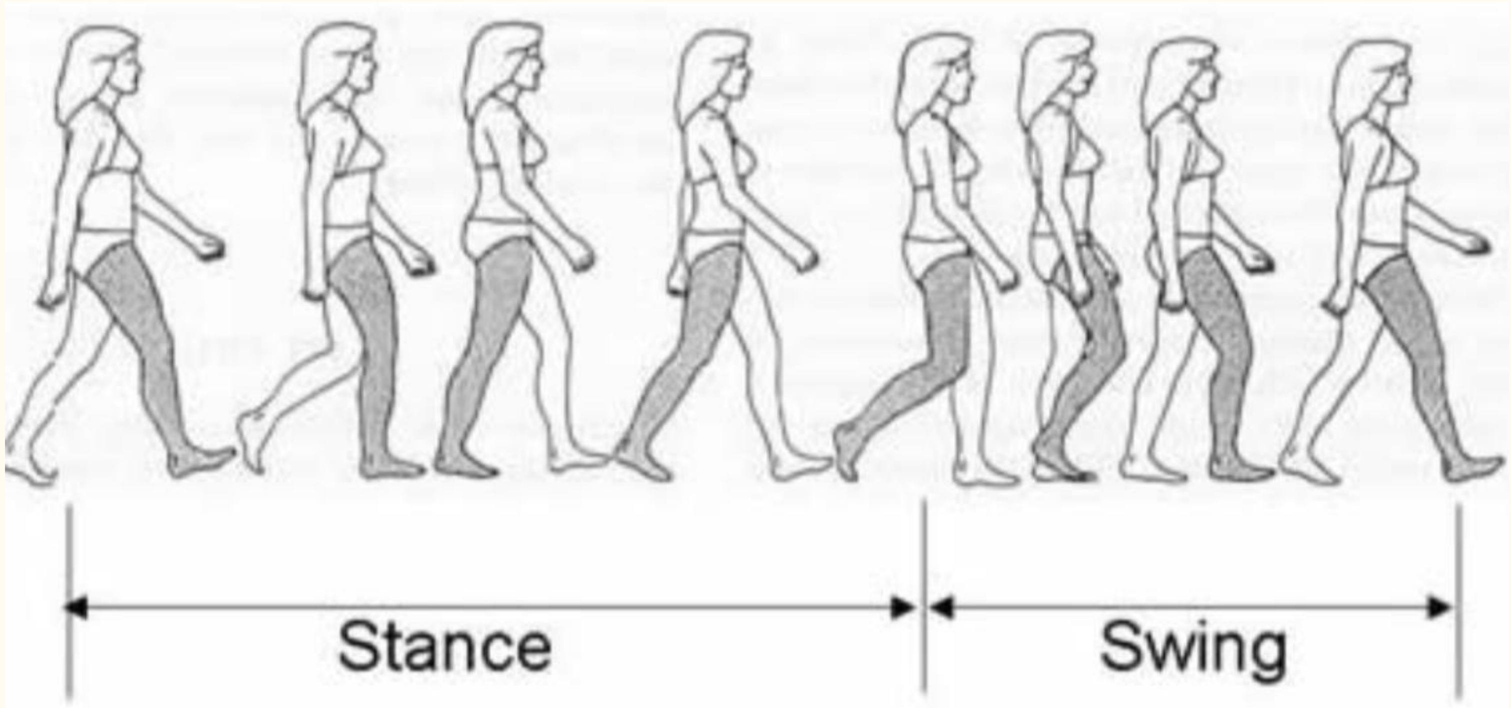
(Intermediate) Gait Representation



GEI (Gait Energy Image) | GEnI (Gait Entropy Image) | GFI (Gait Flow Image) | CGI (Chrono Gait Image)

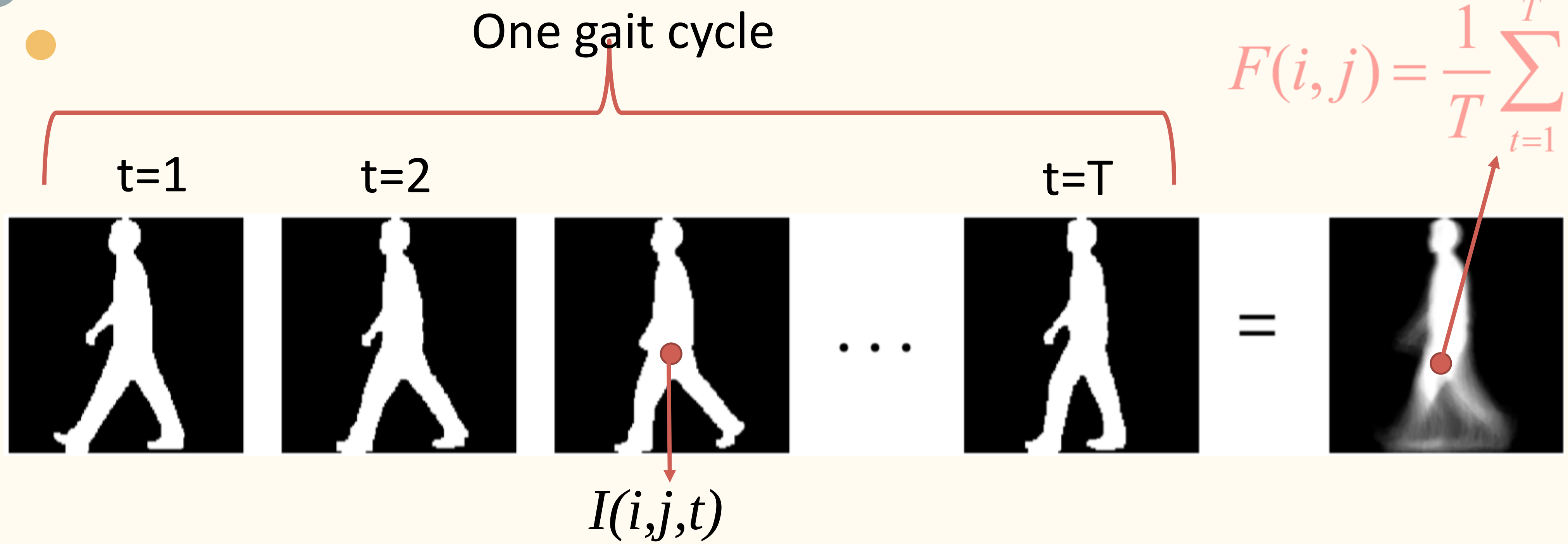
Most widely used

One key concept: gait cycle



- Between where the same foot touches the ground for the first and second time.
- For the purpose of normalization of silhouettes and computing gait templates such as GEI

Gait Energy Image (GEI)



$$F(i,j)=\frac{1}{T}\sum_{t=1}^{T} I(i,j,t)$$

- Spatially well-aligned, temporally averaged gait frames within one gait cycle
- Empirically 30 frames/whole sequence of frames enough to cover a complete gait cycle.
- $F(i,j)$ indicates how likely there appears part of a human body in the position (i,j)
- GEI is robust to the silhouette noise, but may have a high dimensionality

J. Han & B. Bhanu, "Individual recognition using gait energy image," *TPAMI*, 2006.

Gait Entropy Image (GEnI)

$$H(x,y) = -\sum_{k=1}^{K} p_k(x,y)\log_2 p_k(x,y)$$

$$G(x,y) = \frac{(H(x,y) - H_{min}) * 255}{(H_{max} - H_{min})}$$

- Calculate Shannon entropy for each pixel in the silhouette images.
- The dynamic area of human body (legs and arms) are represented by higher intensity values in the GEnIs. In contrast, the static areas such as torso give rise to low intensity values.
- Silhouette pixel values in the dynamic areas are more uncertain and thus more informative leading to higher entropy values.

K.Bashir,T.Xiang,andS.Gong. Gait recognition using gait entropy image. In *Proc. of the 3rd Int. Conf. on Imaging for Crime Detection and Prevention*, pages 1–6, Dec. 2009.

Gait Flow Image (GFI)

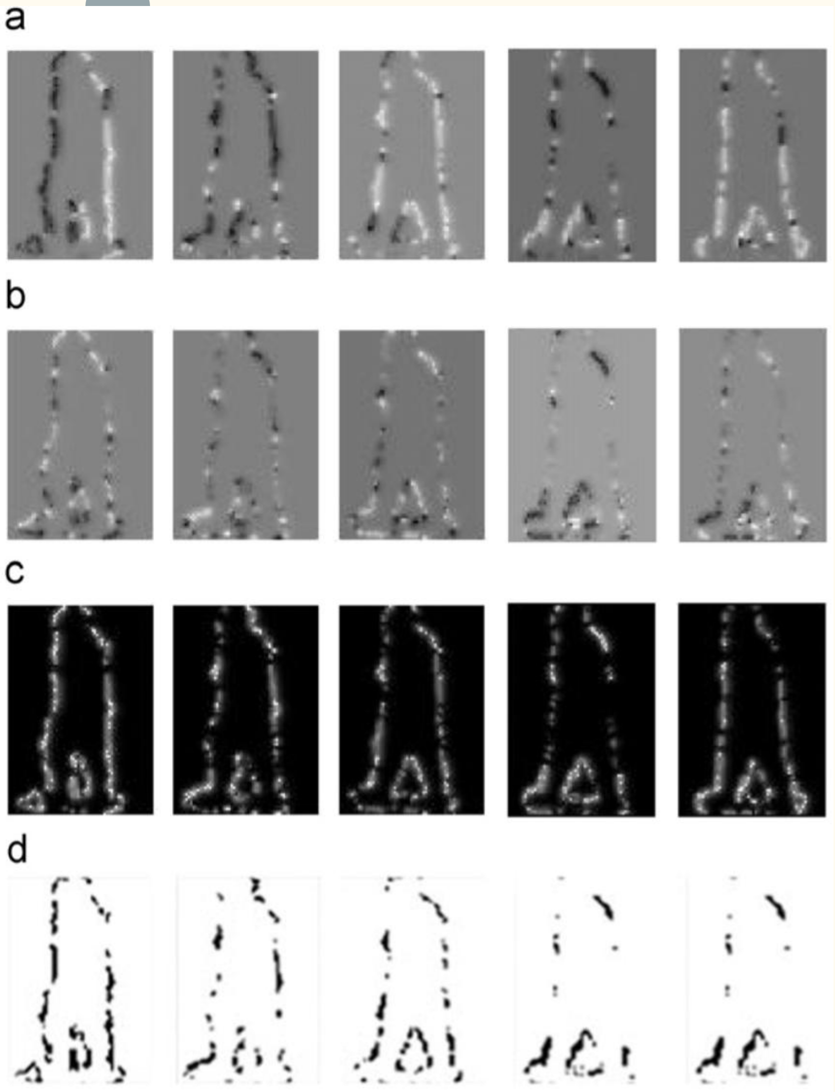


Fig. 4. Optical flow silhouette images: (a) horizontal optical flow field images, (b) vertical optical flow field images, (c) the magnitude of optical flow fields' images and (d) binary flow images.

GFI contains the motion information of the human gait. GFIs are generated by determining the optical flow field from the binary silhouettes of each cycle.

$$(uF_{t,\,i}(x, y), vF_{t,\,i}(x, y)) = OpticalFlow(SI_{t,\,i}(x, y), SI_{t+1,\,i}(x, y))$$

$$\begin{aligned} MagF_{t,\,i}(x, y) &= \left|\left|(uF_{t,\,i}(x, y), vF_{t,\,i}(x, y))\right|\right| \\ &= \sqrt{(uF_{t,\,i}(x, y))^2 + (vF_{t,\,i}(x, y))^2} \end{aligned}$$

$$BF_{t,\,i}(x, y) = \begin{cases} 0 & \text{if } MagF_{t,\,i}(x, y) \geq 1 \\ 1 & \text{otherwise} \end{cases}$$

$$GFI_i(x, y) = \frac{\sum_{t=1}^{N-1} BF_{t,\,i}(x, y)}{N}$$

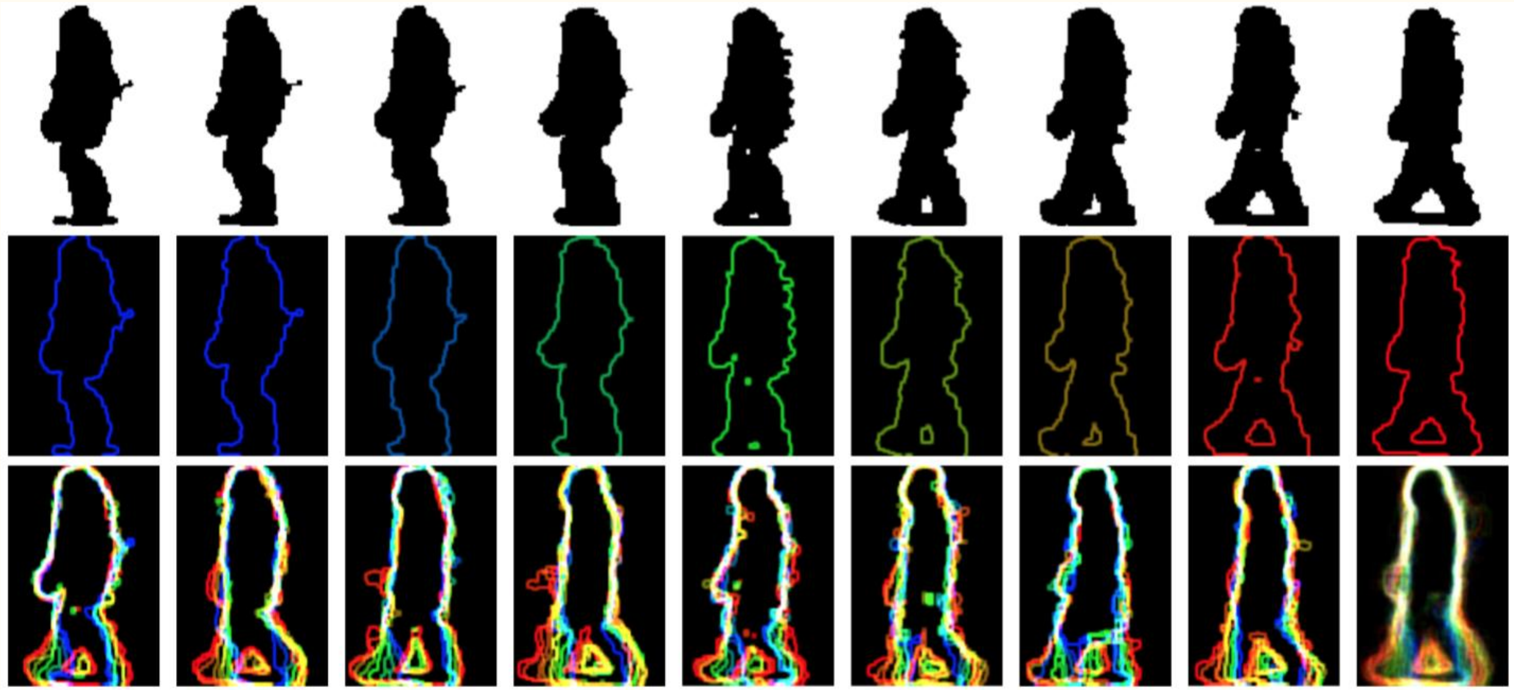
Gait Flow Image (GFI)

- A great advantage of using GFI is that the number of GFIs is smaller than the number of silhouette images. In other words, GFI is more computationally efficient.

- However, if the silhouettes are extracted at a low quality, a GFI may be embedded with irrelevant information, which affects the recognition rate.

T. Lam et al "Gait flow image: A silhouette-based gait representation for human identification," *Pattern Recognition* 2010.

Chrono Gait Image (CGI)



- We encode temporal information in the silhouette images with additional colors to generate a chrono-gait image.
- The goal of CGIs is to compress the silhouette images into a single image without losing too much temporal relationship between the images

C. Wang et al, “Chrono-gait image: A novel temporal template for gait recognition,” in *ECCV*, 2010.
C. Wang et al, Human Identification Using Temporal Information Preserving Gait Template, TPAMI, 2012.

Performance of different gait representations

Performance comparison of six gait features in terms of the rank-1 and rank-5 identification rates

Dataset	#Subjects	Rank-1 identification rate [%]						Rank-5 identification rate [%]					
		GEI	FDF	GEnI	CGI	GFI	MGEI	GEI	FDF	GEnI	CGI	GFI	MGEI
A-55	3,706	**84.70**	83.89	76.42	75.58	75.15	68.35	**92.39**	91.53	86.67	86.02	85.83	80.09
A-65	3,770	**86.63**	85.49	78.65	78.97	77.11	68.91	**92.84**	92.81	88.14	88.06	87.32	79.71
A-75	3,751	**86.91**	86.59	79.95	81.58	76.54	67.10	92.78	**92.88**	89.23	89.28	85.84	78.41
A-85	3,249	85.72	**85.90**	80.95	83.35	74.92	61.19	**93.01**	92.83	89.60	90.80	84.73	73.19
A-ALL	3,141	**94.24**	94.17	90.93	91.60	87.46	84.18	**97.13**	97.10	95.35	95.32	92.84	90.58

A recent empirical study by Iwama et al. shows that GEI, despite of its simplicity, is the most stable and effective kind of features for gait recognition on their proposed dataset with 4,007 subjects.

H.Iwama, et al,“The OU-ISIRgait database: Comprising the large population dataset and performance evaluation of gait recognition,” IEEE Trans. Inf. Forensics Security,. 2012.

Outline

1. Introduction and overview
2. Traditional approaches for gait-based human identification • History and databases • Gait representation and learning algorithms
3. Deep networks for gait-based human identification • Cross-view gait based human identification with deep CNNs
4. How to build a practical gait-based human identification system? • End-to-end deep network for gait segmentation & recognition • System demo
5. Open questions and discussion

The pipeline of a typical GEI-based gait recognition method.

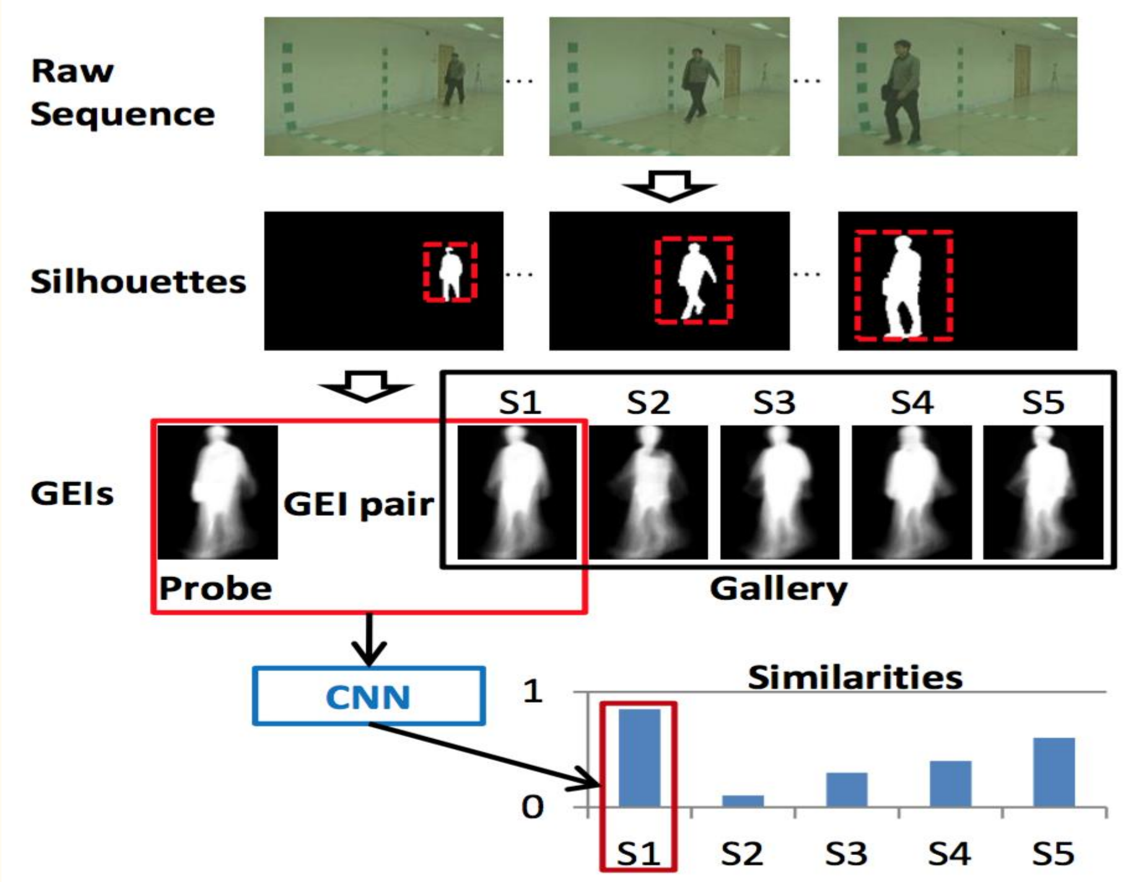
1. Extract human silhouettes from video sequences

2. Align and average the silhouettes along the temporal dimension to get a GEI.

3. Given a probe GEIs and those in the gallery, evaluate the similarities between each pair of probe and gallery GEIs.

4. Assign the identity of the probe GEI, usually with the nearest neighbor classifier.

Different from previous methods, here the third step above is realized with deep convolutional neural networks (CNN).



Z. Wu, Y. Huang, L. Wang, X. Wang, T. Tan, A comprehensive study on cross-view gait based human identification with deep CNNs, IEEE TPAMI, 2016

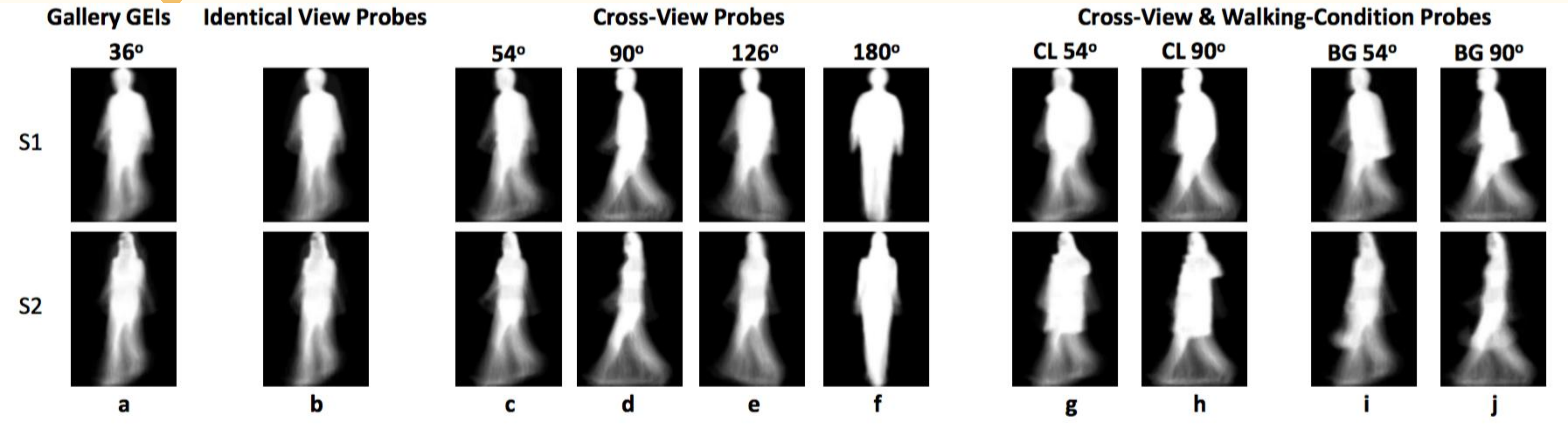
Robustness of gait recognition system

One of the biggest challenges is to disentangle the identity-unrelated factors

- subject-related ones : walking speed, dressing and carrying conditions,
- device-related ones : different frame rates and filming resolutions,
- environment-related ones : illumination conditions and camera viewpoints.

Among these, the change of viewpoints would be one of the most tricky factors.

Cross-view examples in the CASIA-B database

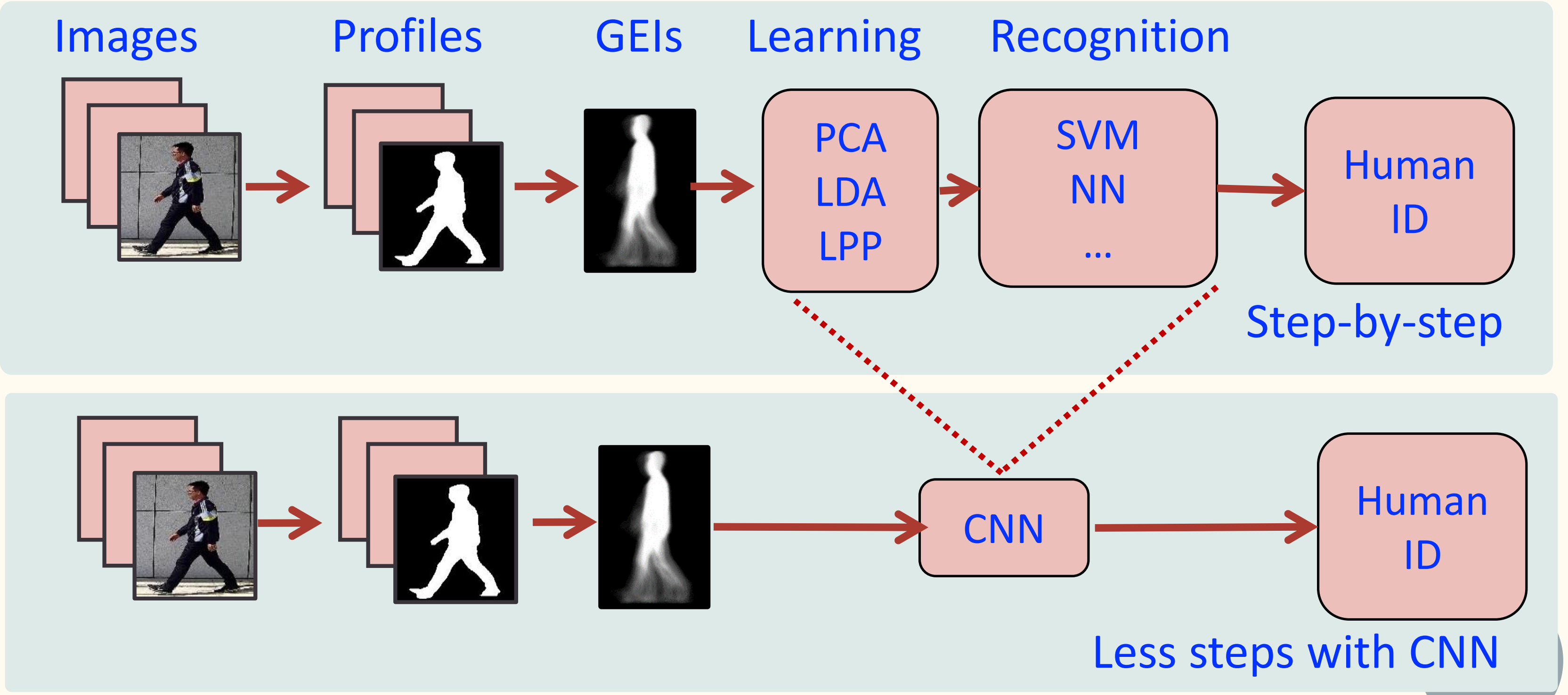


- The performance of an approach ignoring cross-view variations would drop drastically when the viewpoint changes.
- Because the appearances of objects can be substantially altered, leading to intra-class variations larger than inter-class variations.

Feature learning for gait recognition

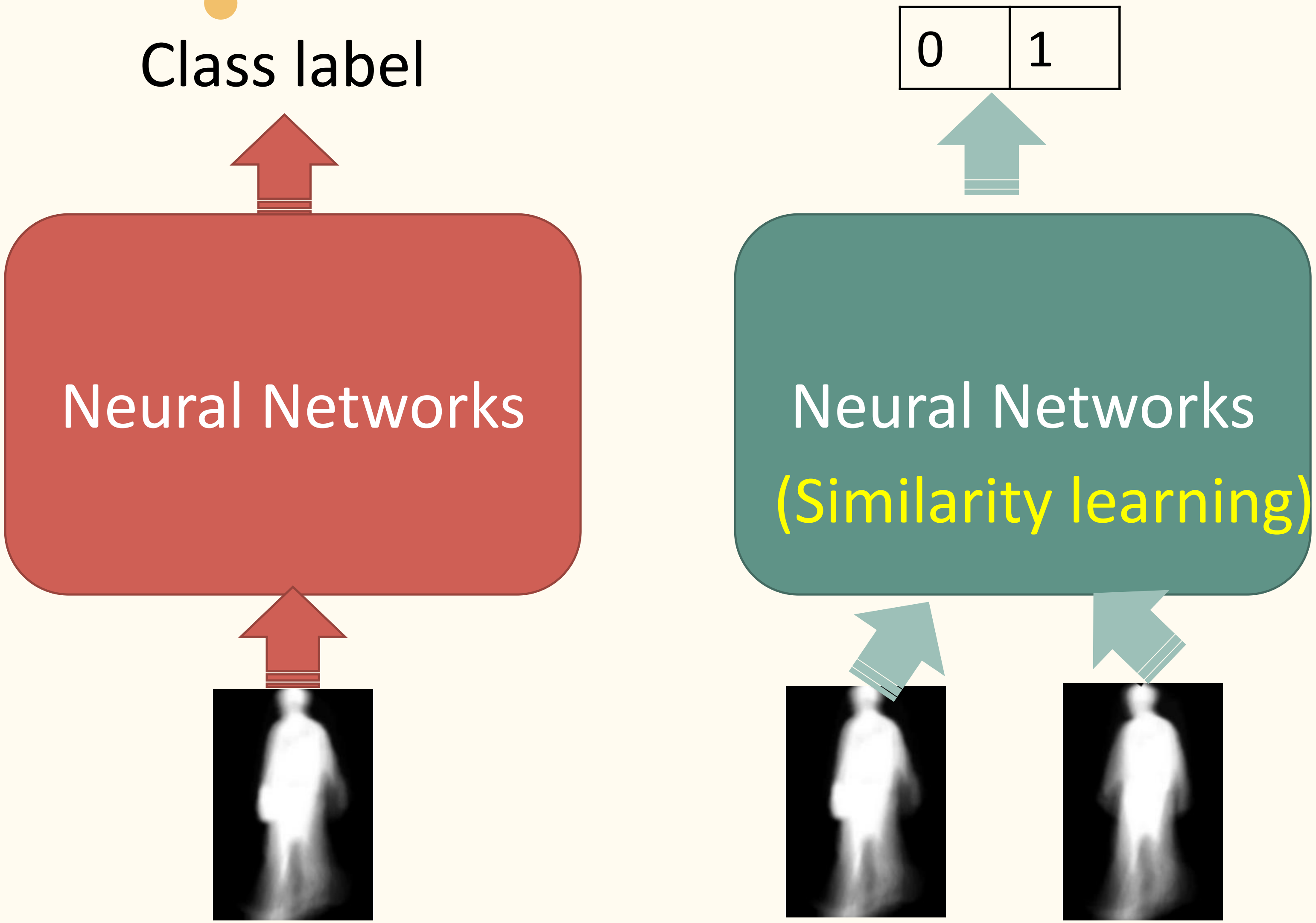
Traditional learning pipeline

Deep learning pipeline



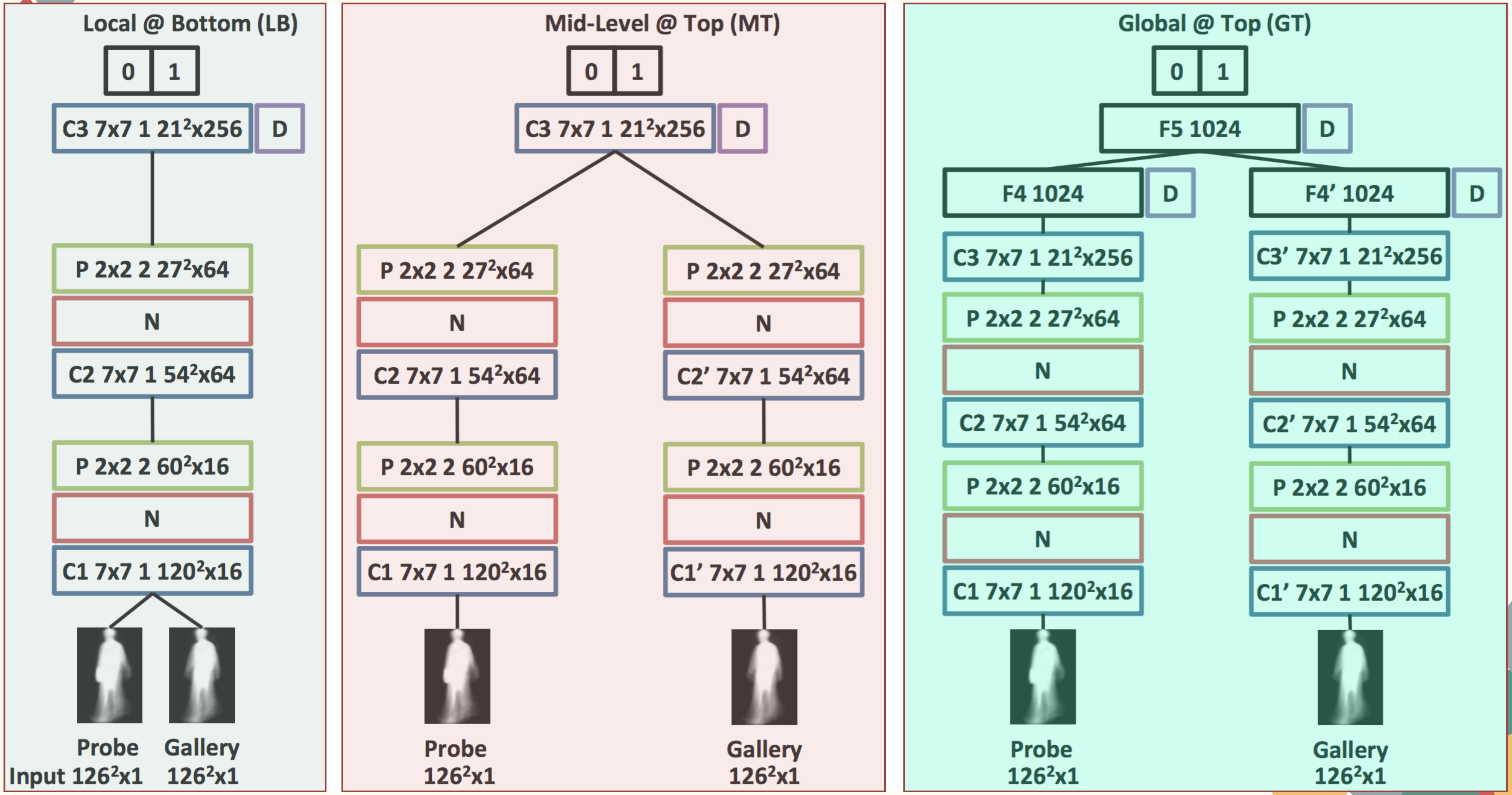
It is difficult to manually design view-invariant feature representations for gait recognition

A similarity learning approach



- Few labeled multi-view human walking videos, many labeled pairs
- Train deep networks to recognize the most discriminative changes of gait patterns which suggest the change of human identity

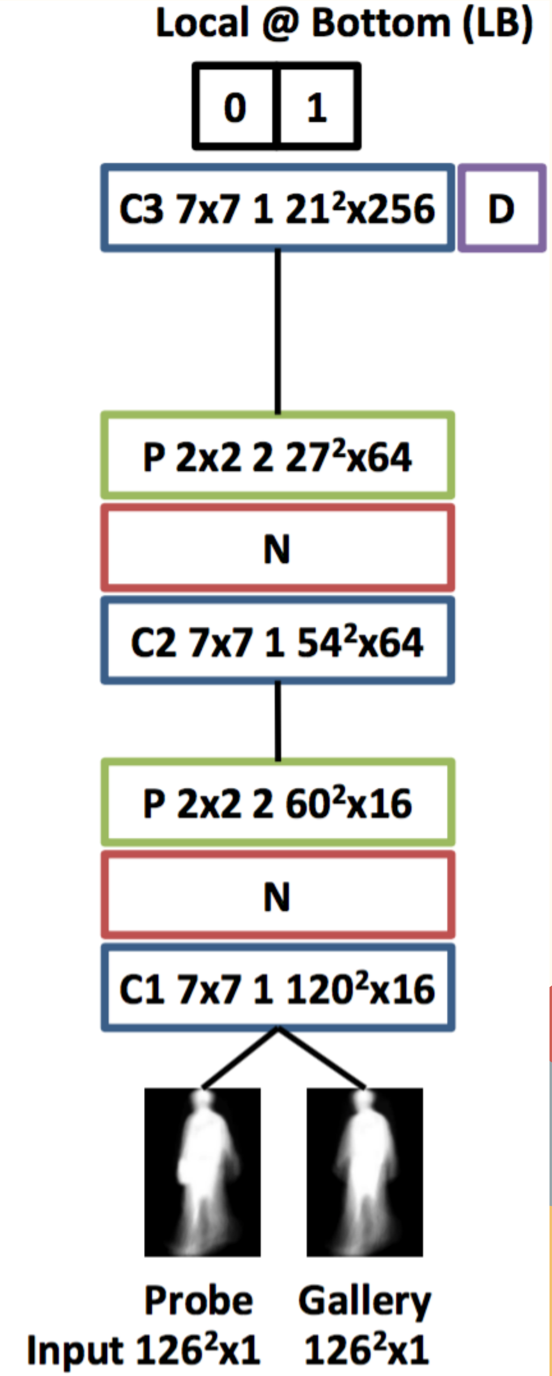
Three network architectures to be investigated.



Network architectures

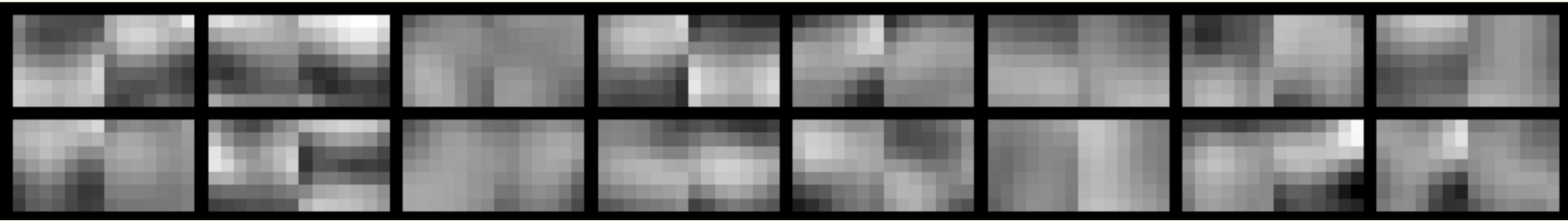
1) Matching Local Features at the Bottom Layer (LB)

- Pairs of GEIs are compared within local regions
- Only linear projection is applied before computing the differences between pairs of GEIs, which is realized by the sixteen pair-filters in the bottom-most convolution stage.
- A pair-filter takes two inputs and can be seen as a weighted comparator.
- At each spatial location, it will first re-weight the local regions of its two inputs respectively, and then render the sum of these weighted entries to simulate the *subtraction*.



The "Subtraction" pair-filters

Two horizontally adjacent filters constitute a pair-filter.

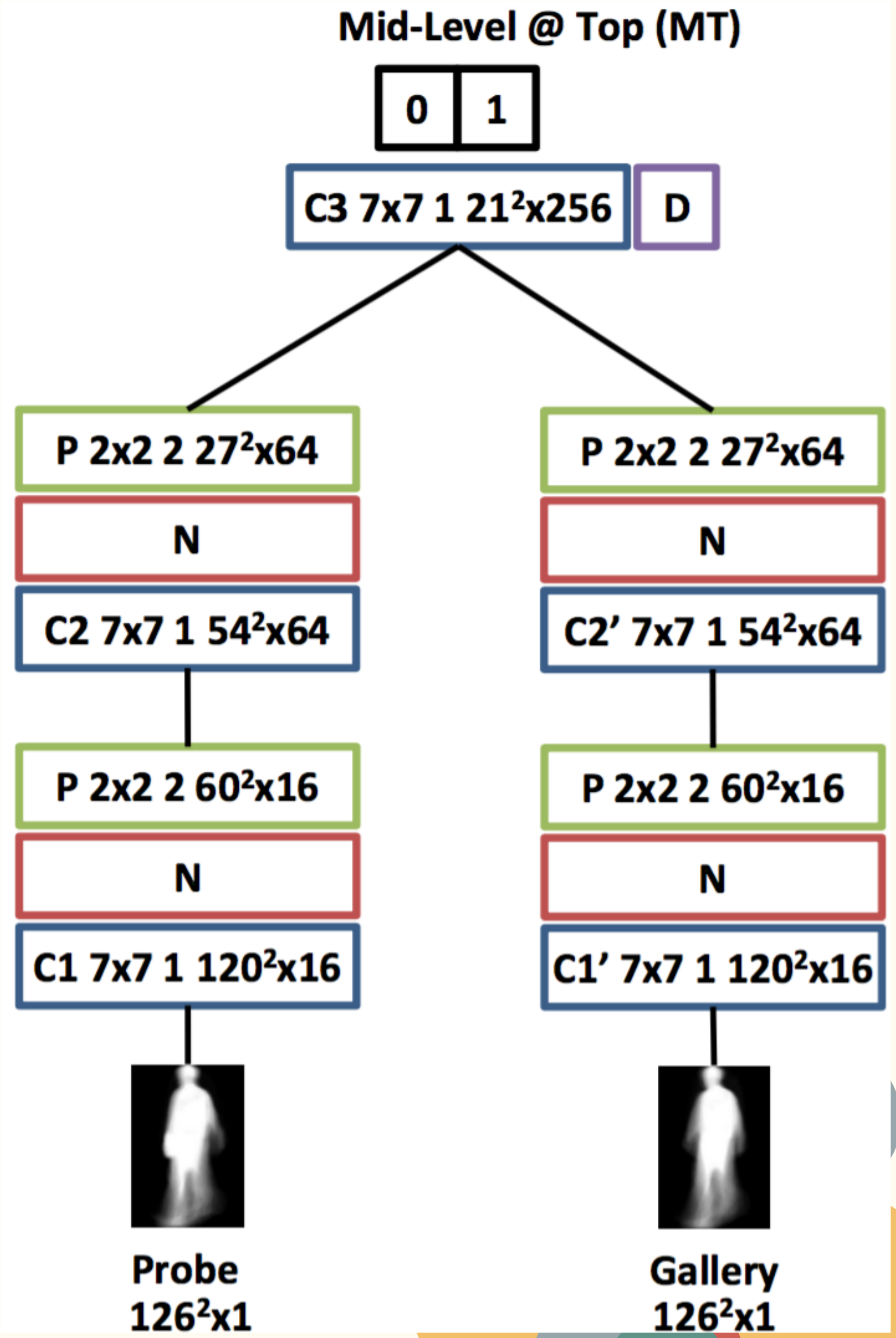


- Some of the learned pair-filters are subtracting gallery GEIs from probe GEIs.
- Project GEIs of different views into a common space where the GEIs become more comparable.
- There are two more convolution stages above the matching layer, whose nonlinearity is supposed to be beneficial to learning complex patterns from the differences between GEI pairs.

Network architectures

2) Matching Mid-Level Features at the Top Layer (MT)

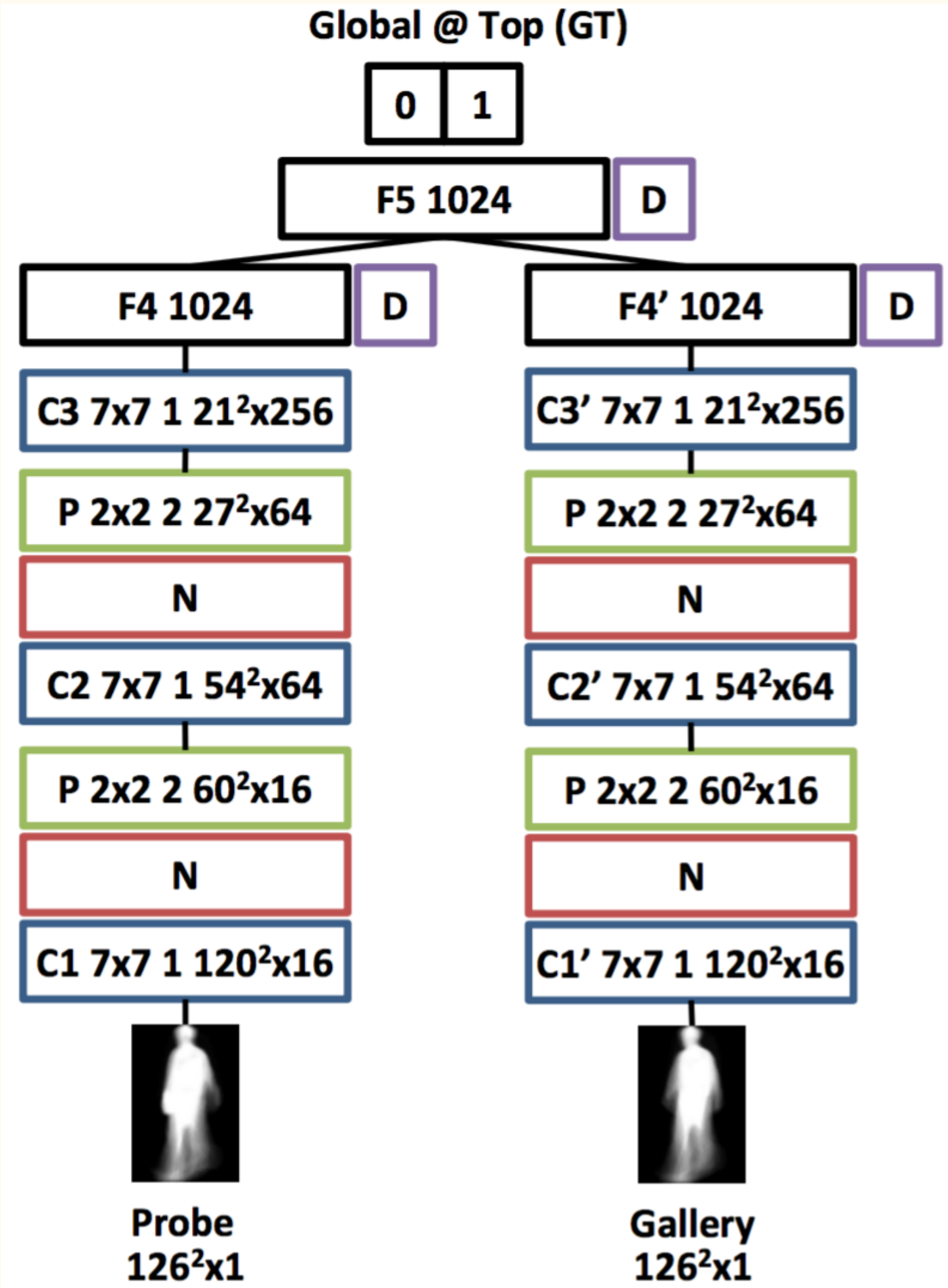
- Two extra non-linear projections are applied
- The motivation is to apply deep non-linear normalization to GEIs instead of the shallow linear one in LB.
- LB directly computes the weighted differences at the bottom layer (with local features), and then learns to recognize the patterns in the obtained differences with the rest two convolution layers.
- In contrast, MT learns mid-level features first, and then computes the weighted differences.
- Model complexities of LB and MT are consistent



Network architectures

3) Matching Global Features at the Top Layer (GT)

- Pairs of GEIs are compared with each other by learned global features.
- Two more fully-connected layers compared with Network MT.
- The weighted differences are computed from global features at Layers F4 and F4'. Each of them is the descriptions of a whole GEI, with only 1,024 entries, which is much more compact than those of Networks LB and MT.



Model complexity

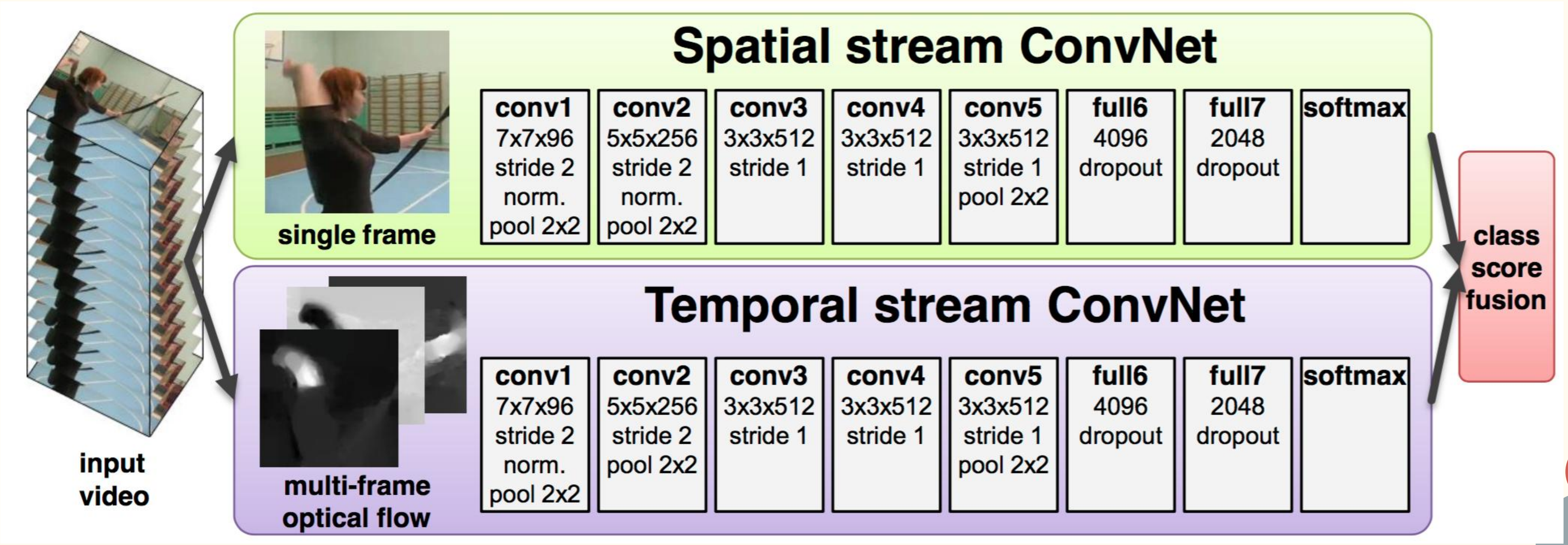
- The model complexity of Network GT is higher than the previous two due to the use of fully-connected layers, which can lead to over-fitting depending on the size of training data.

- However, the advantage of this network is its compactness, which can lead to computational efficiency.
  - First, we can store in advance the output of Layer F4' for all gallery GEIs.
  - Second, feed a probe GEI to the network once and obtain the output of Layer F4.
  - Finally, compare the two 1,024-dimensional features using Layer F5 and the two-way classifier.

A substitute model: Compact Mid-Level & Top (CMT)

- We do NOT compare Global @ Top (GT) in detail considering its less satisfactory performance.
- In our experiments, Network GT suffers from severe over-fitting, probably due to the small training dataset. However, we sometimes do favor its computational efficiency.
- As a compromise, we modify Network MT to obtain more compact features, which is the very third network compared here, i.e., Compact Mid-Level & Top (CMT).
- It amounts to use a larger stride in the third convolution stage. For example, when we use a stride of five, the resulting feature map will be in size $3 \times 5 \times 256$, with only 3,840 entries.

Inspiration: Two-Stream Convolutional Networks for Action Recognition in Videos [NIPS 2015]

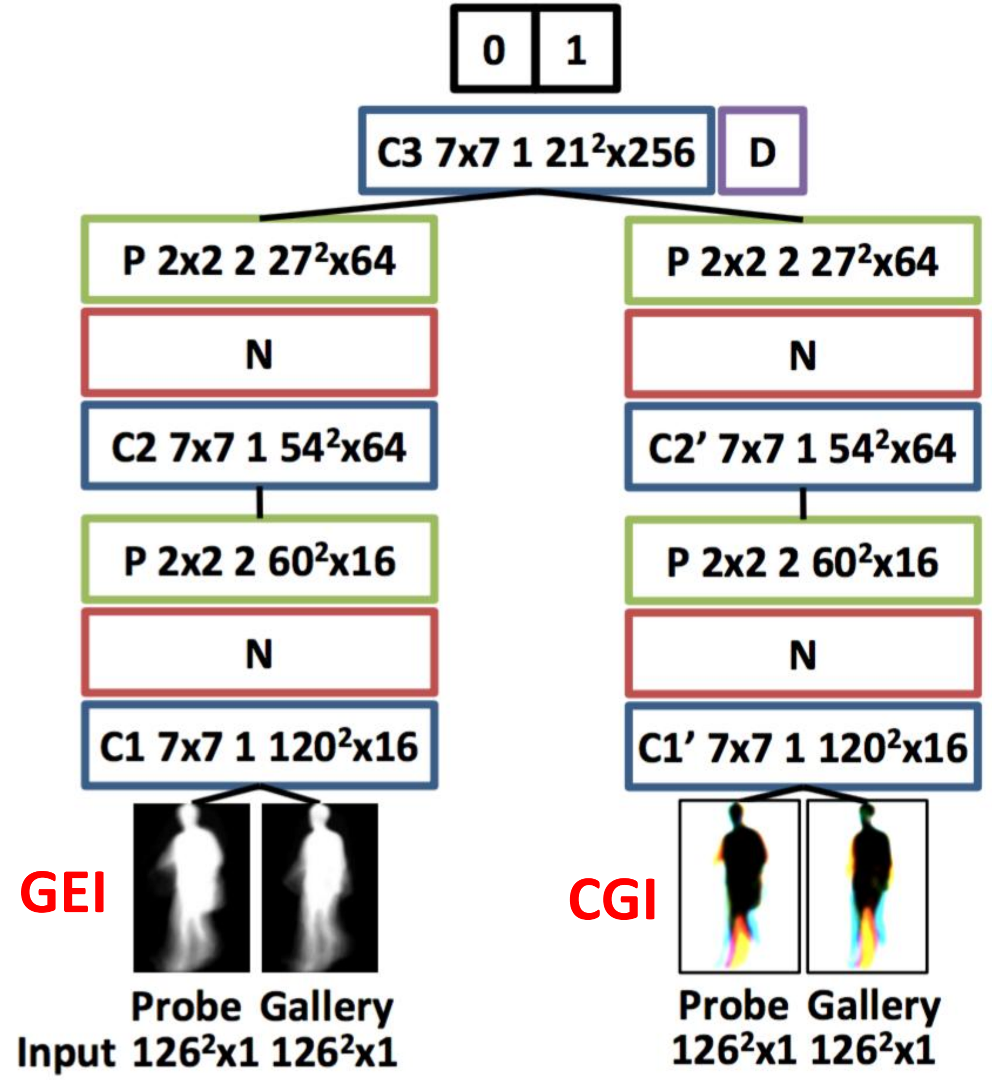


Two-stream architecture for video classification: capture the complementary information on appearance from still frames and motion between frames.

Network architectures

4) Two-stream network

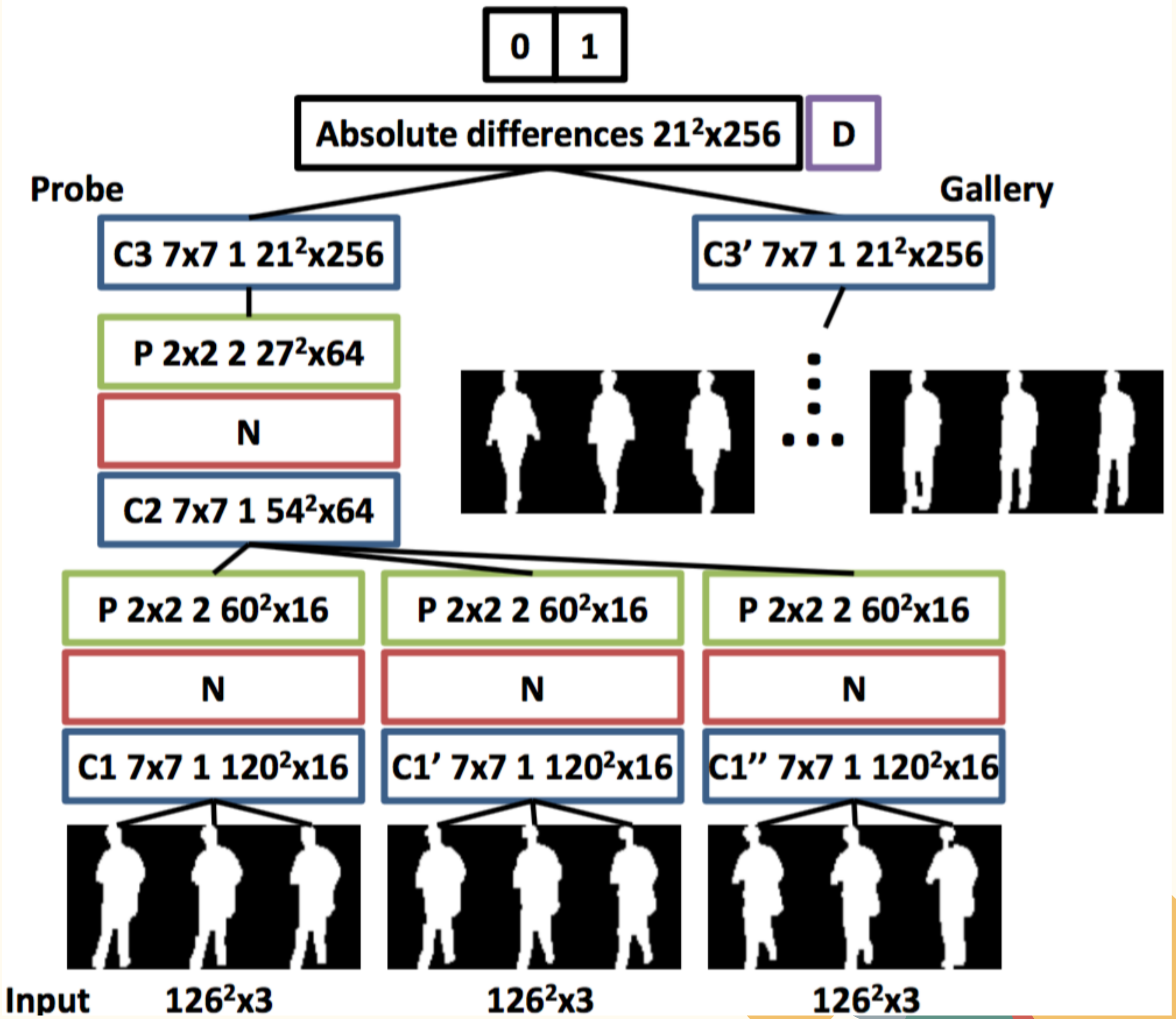
- Composed of two LB networks.
- The left stream takes a pair of GEIs as the input, which is the counter part of the stream processing still images in Zisserman's network.
- The right stream takes a pair of chrono-gait images (CGIs) as the input, which is the counter part of the stream processing optical flow features.



Network architectures

4) Two-stream 3D CNN network

- Train a network with 3D convolutions in its first & second layers.
- Training: each time we feed it with a pair of sequence slices, each of which contains nine adjacent frames sampled from a gait sequence.
- Testing: we feed it with all frames of a sequence (nine by nine to fit the network input), and average the output.

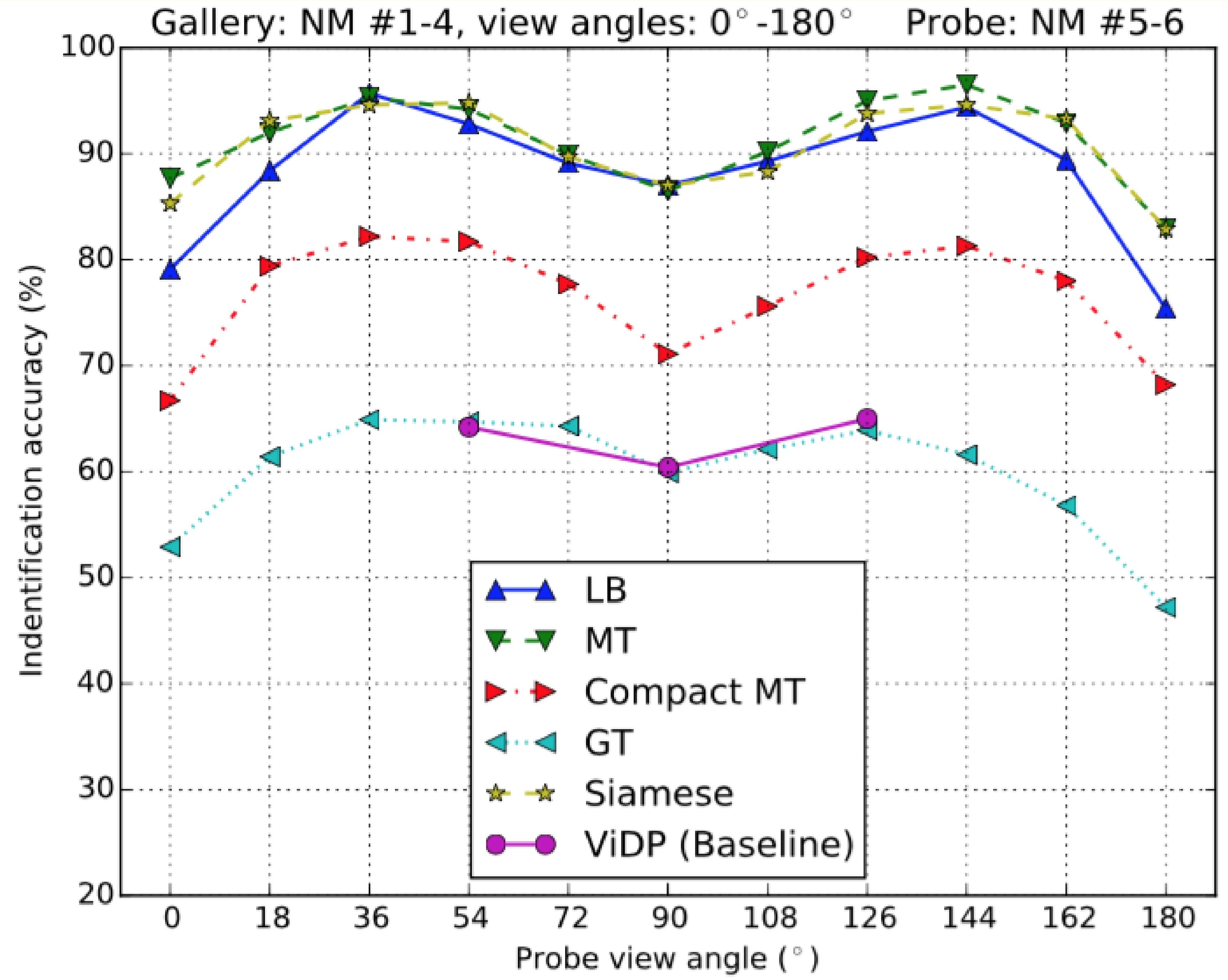


Experimental results

Gallery NM #1-4	0°-180°				36°-144°		
Probe NM #5-6	0°	54°	90°	126°	54°	90°	126°
SVR [30]	–	28	29	34	35	44	45
TSVD [29]	–	39	33	42	49	50	54
CMCC [12]	46.3	52.4	48.3	56.9	-	-	-
ViDP [23]	–	59.1	50.2	57.5	83.5	76.7	80.7
Ours	**54.8**	**77.8**	**64.9**	**76.1**	**90.8**	**85.8**	**90.4**

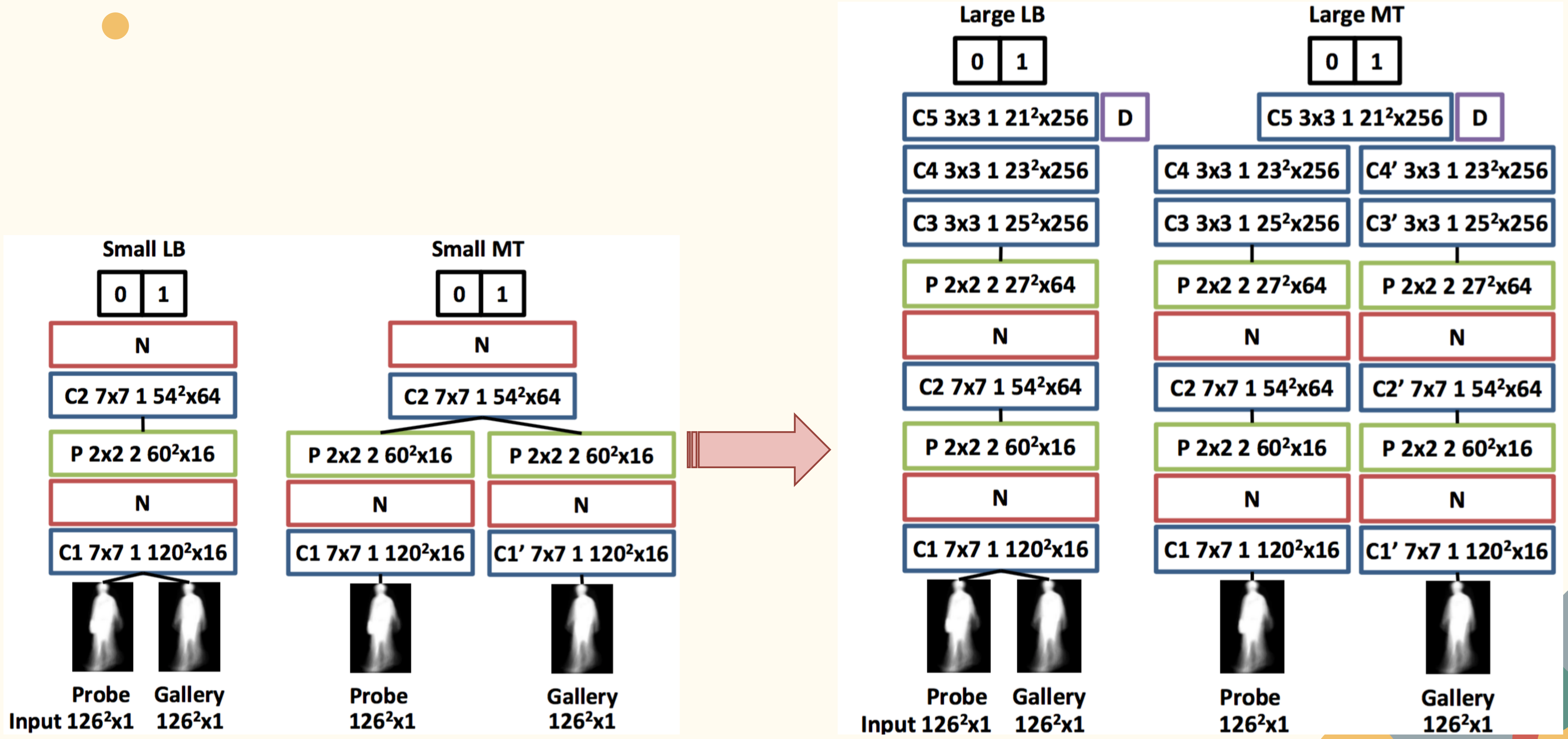
Comparison of our method with previous ones on CASIA-B by average accuracies. Models are trained with GEIs of the first 24 subjects

Impact of network architectures

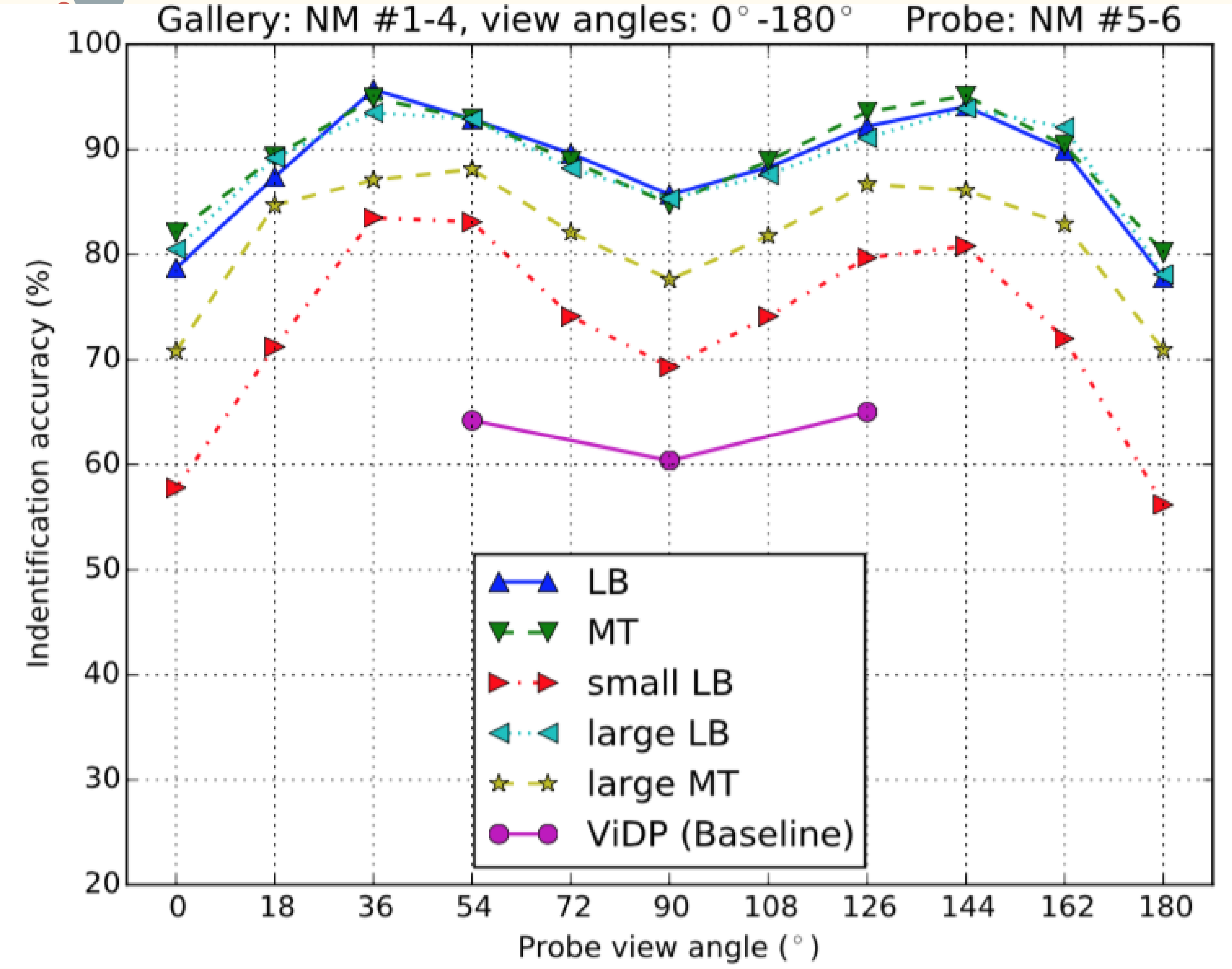


1) *LB ≈ MT ≫ GT:* There are no significant gaps between the performances of LB and MT, and they both outperform GT with a clear margin.
2) *LB vs. MT:* The most notable difference between the two is that MT performs better for view angles around 0◦ or 180◦.
3) *MT vs. CMT:* There is a moderate drop in performance for CMT compared with MT
4) *MT vs. Siamese:* The Siamese network can approximately be seen as a special case of MT .
5) *0◦ ≈180◦ >90◦ >··· >36◦ ≈144◦*

Influence of network depth

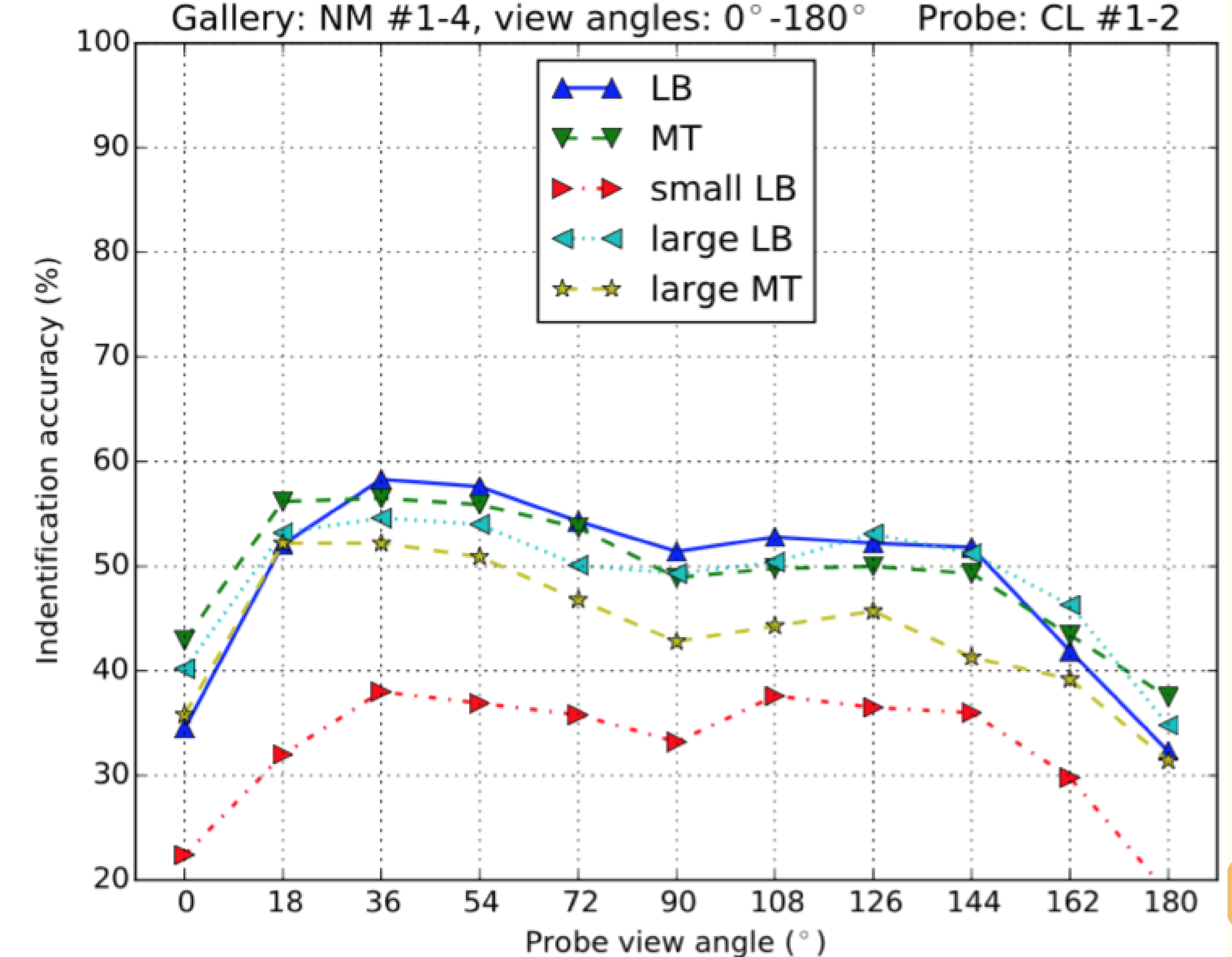


Influence of network depth

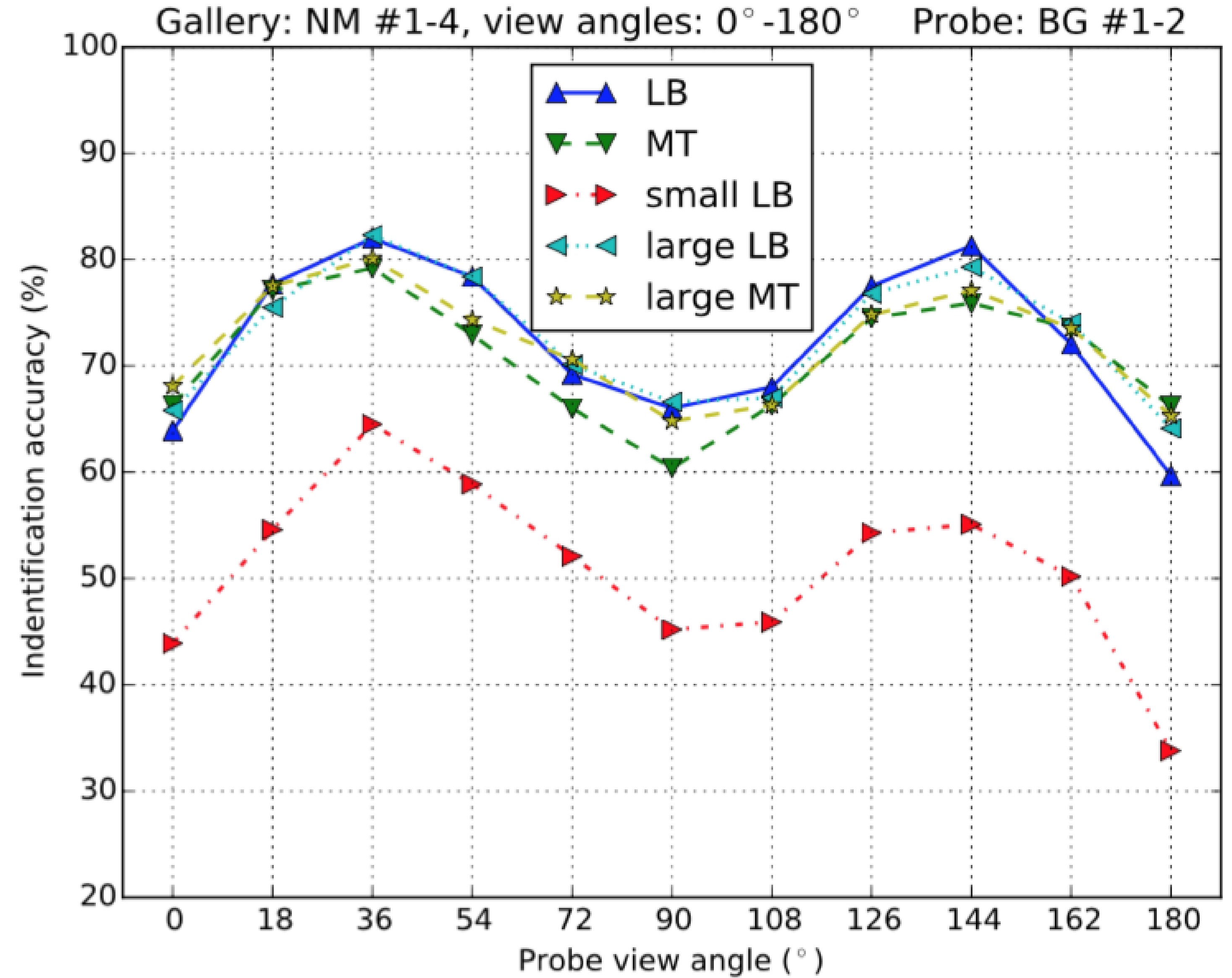


MT>LB~large LB>large MT>small LB.

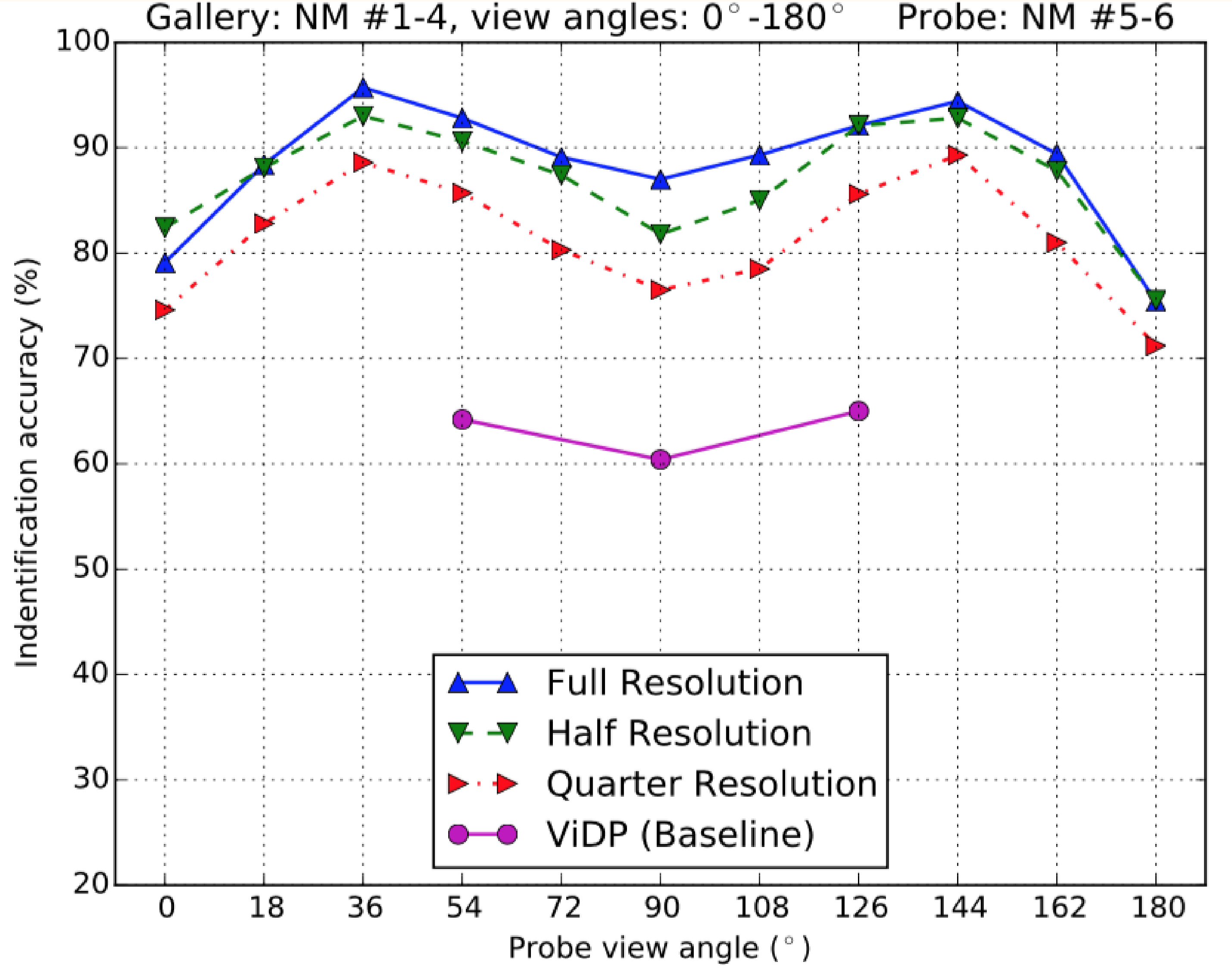
Influence of network depth



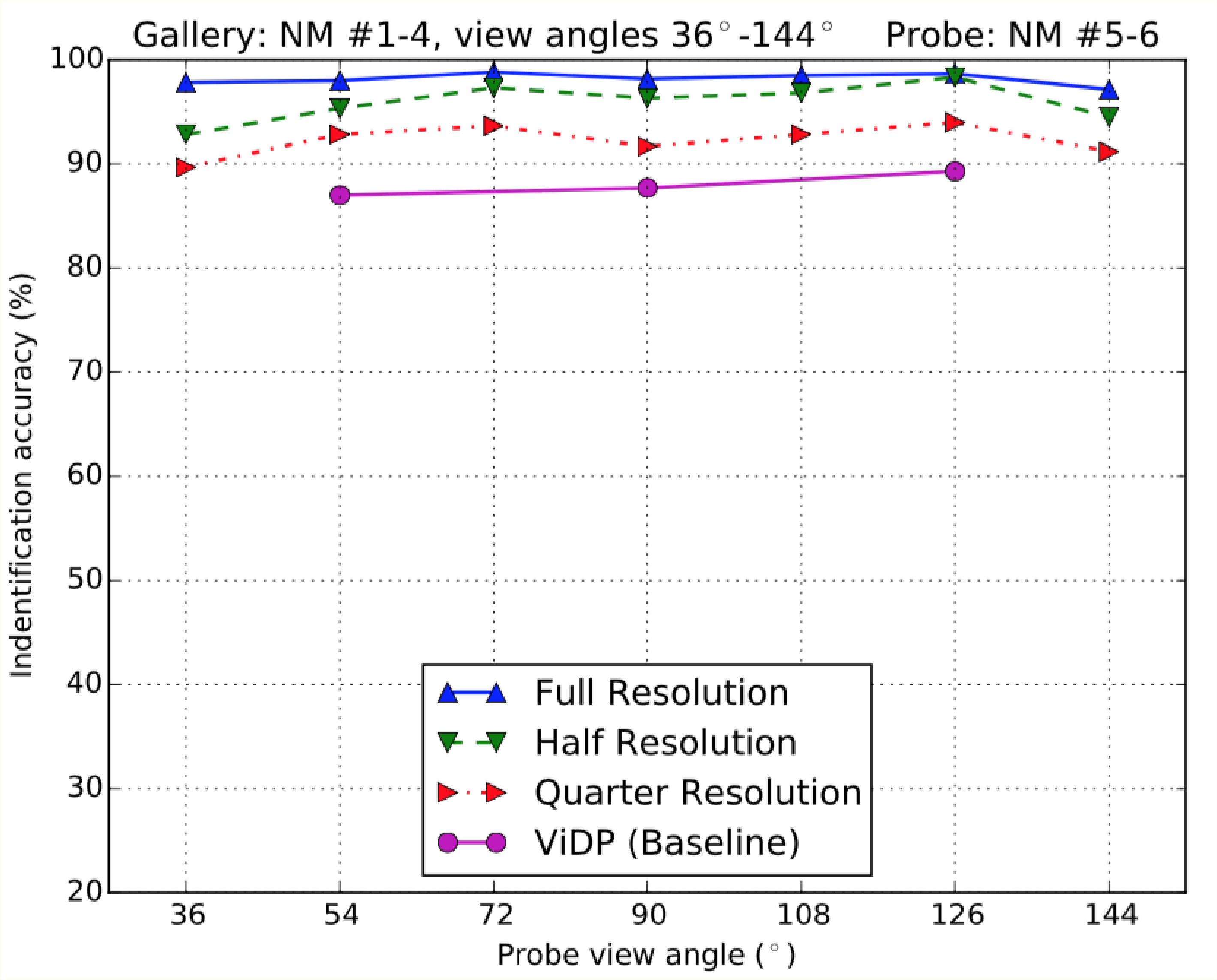
Influence of network depth



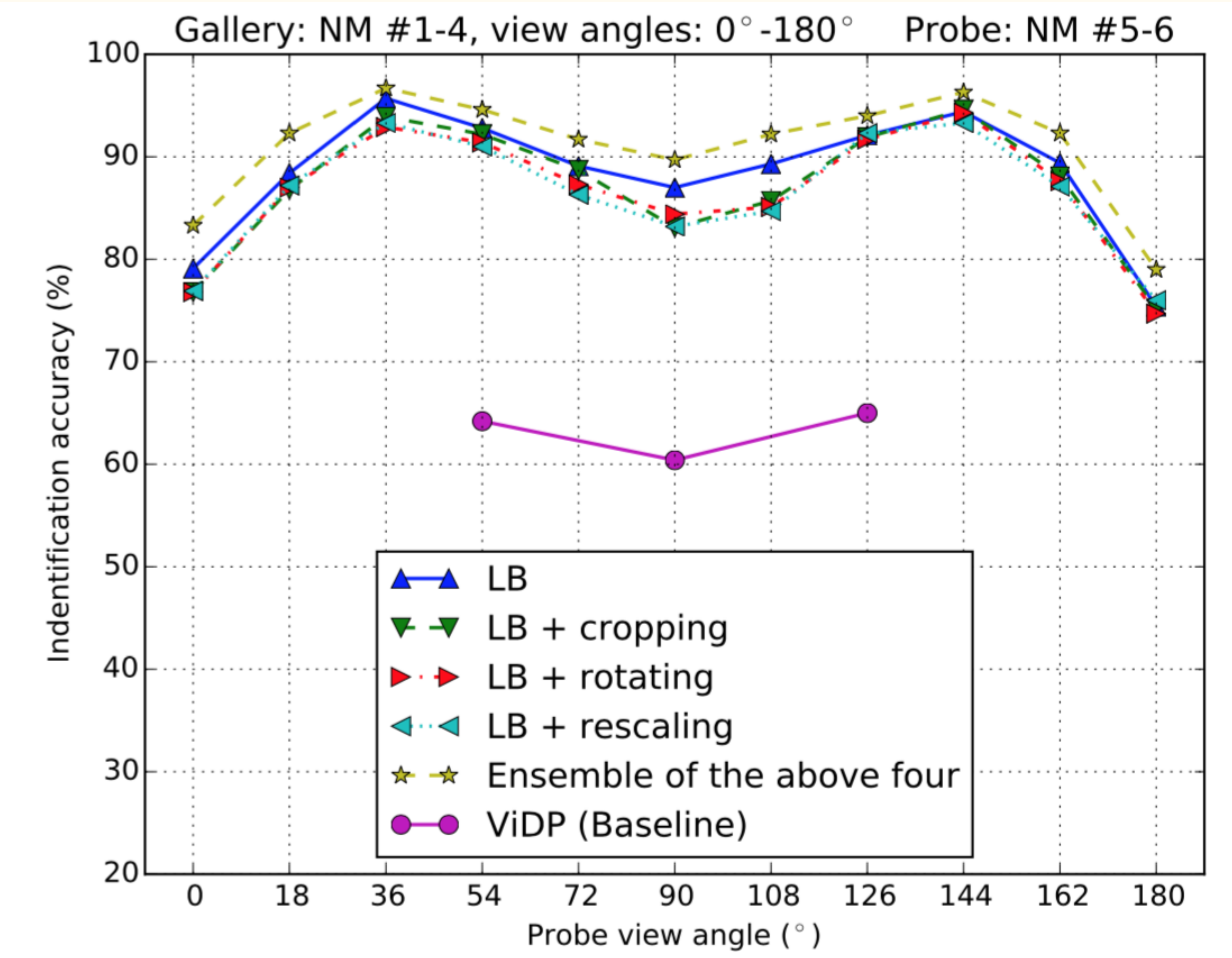
Influence of input resolutions



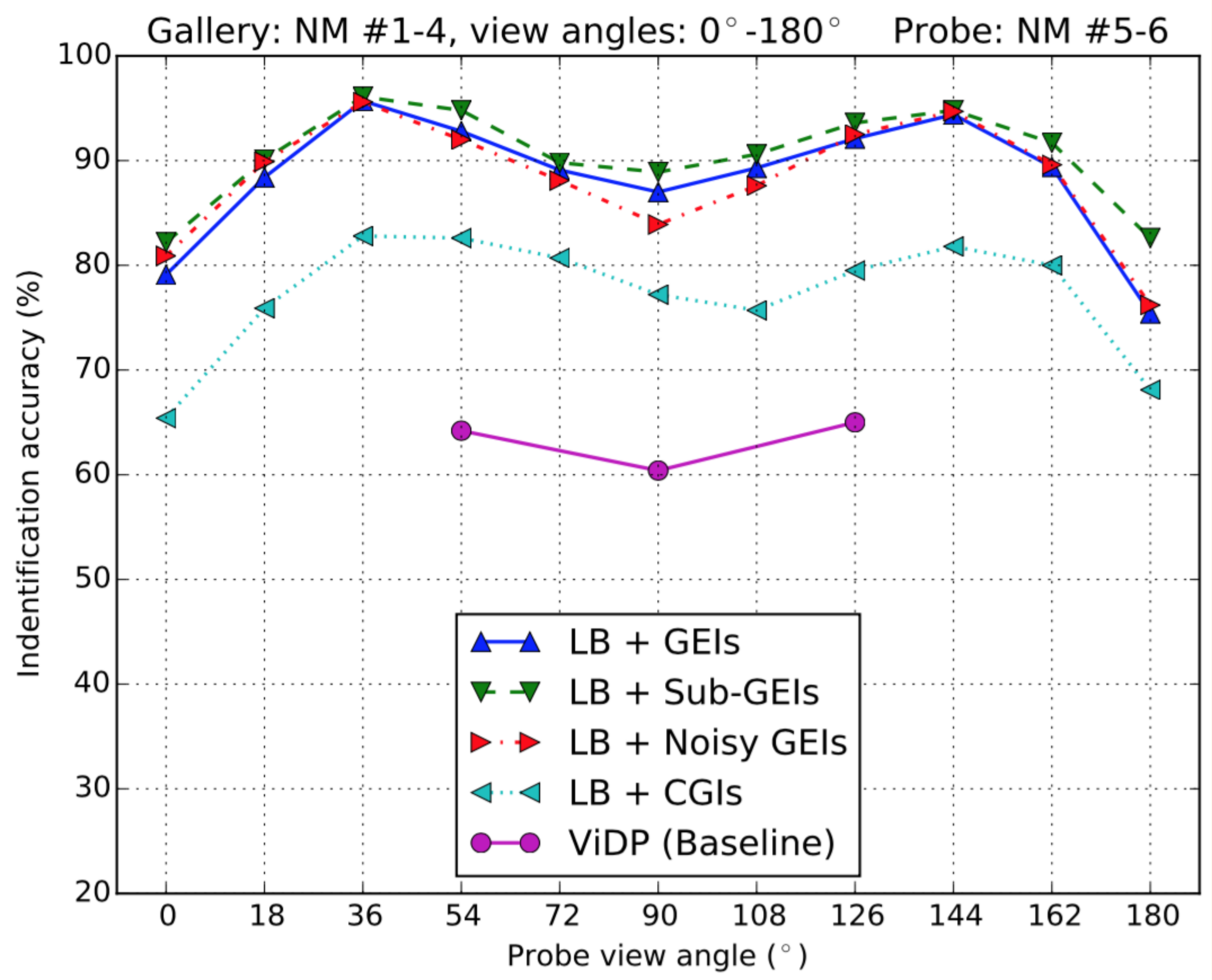
Influence of input resolutions



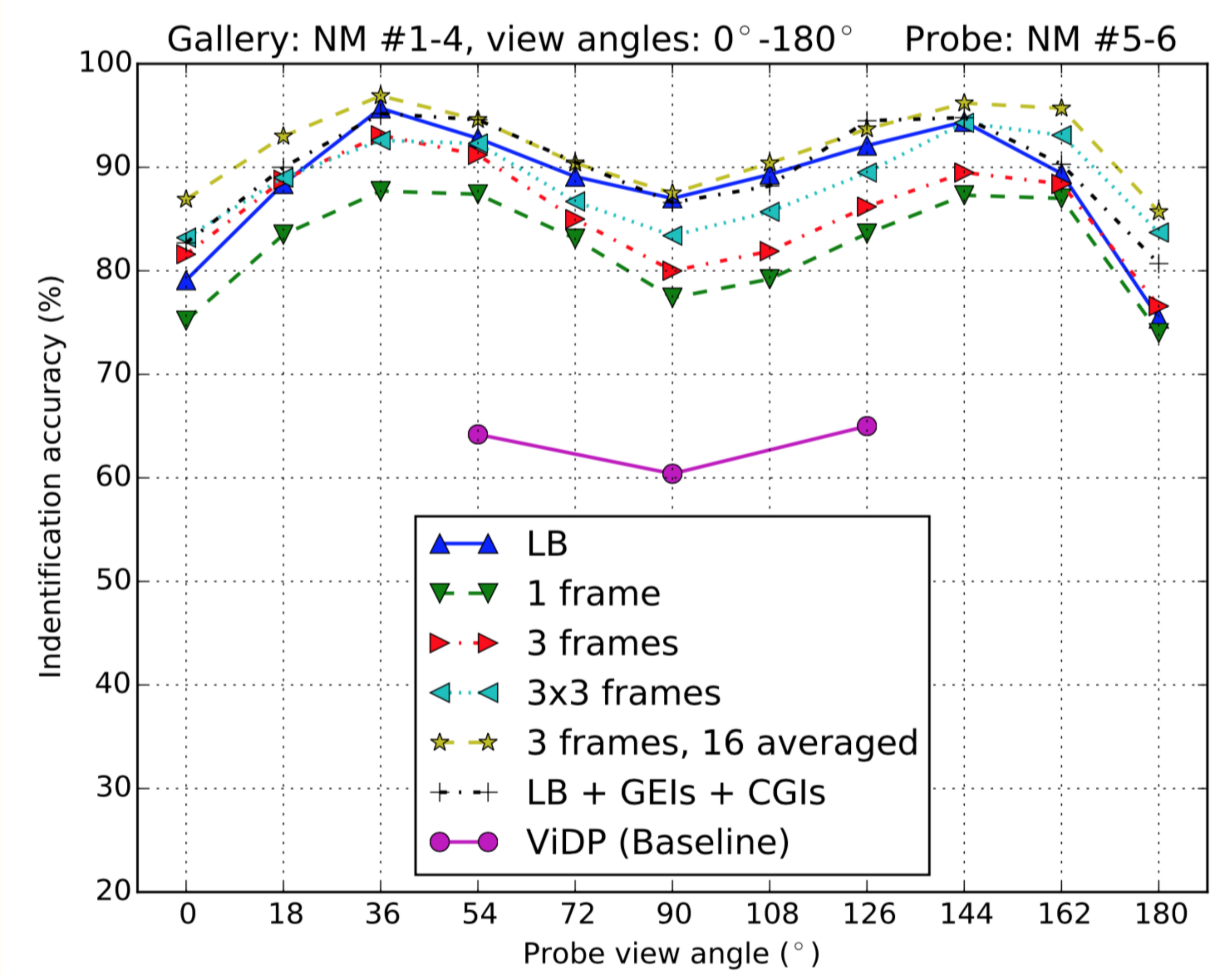
Influence of data augmentation



Influence of input features



Influence of temporal information



Summary

Lack of Datasets for uncooperative gait recognition:

- A subject may halt, or turn around, so his/her gait sequence is not consecutive.
- There may be multiple subjects at the same time, and moving objects in the background, so it is harder to extract silhouettes.
- The cameras may be above the subjects, so more viewpoints should be considered.

It would be very hard to train cross- view gait recognition models on so small a dataset due to severe over-fitting.

Besides, considering the above mentioned factors, to re-identify a person in unscripted surveillance videos only relying on gait recognition, there still seems a long way to go. Probably, such a dataset with enough number of training data can push us forward to this goal.

Summary

Less heuristic preprocessing:

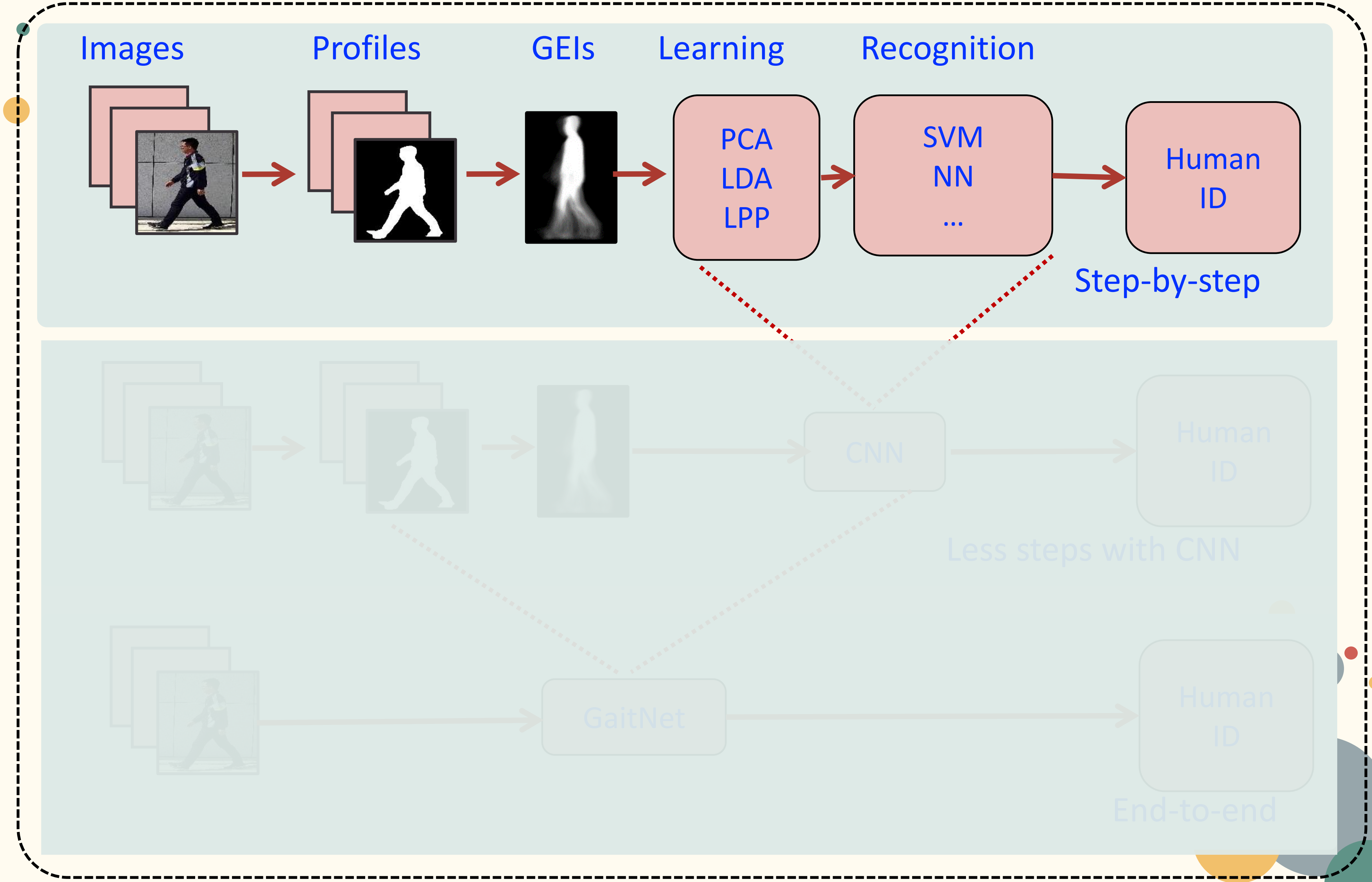
- There are many methods which can be used to improve our preprocessing. For example, pedestrian detection methods can locate a subject from complex backgrounds, pixel-wise labeling methods can extract silhouettes from raw images, and pose estimation methods can provide auxiliary information or help refining the silhouettes.

- Without these comprehensive methods, it would be intractable to deal with the above discussed kind of datasets for uncooperative gait recognition.

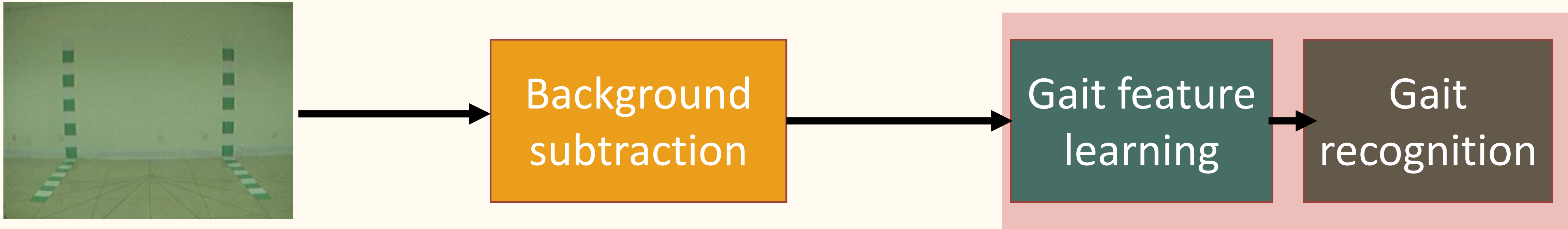
- But in this work, preprocessing is not our main concern, so we keep it as our future work.

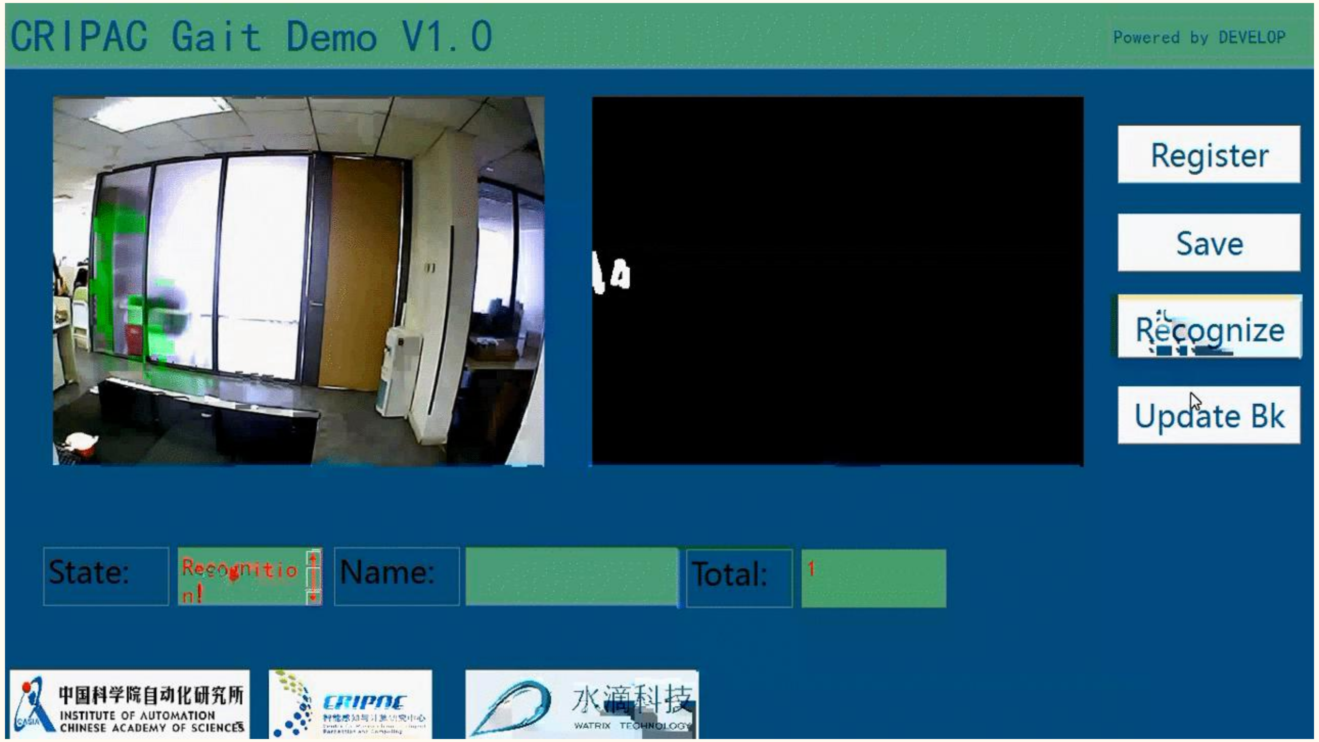
Outline

1. Introduction and overview
2. Traditional approaches for gait-based human identification • History and databases • Gait representation and learning algorithms
3. Deep networks for gait-based human identification • Cross-view gait based human identification with deep CNNs
4. How to build a practical gait-based human identification system? • End-to-end deep network for gait segmentation & recognition • System demo
5. Open questions and discussion


Images
Profiles
GEIs
Learning
Recognition
PCA
LDA
LPP
SVM
NN
...
Human
ID
Step-by-step
CNN
Human
ID
Less steps with CNN
GaitNet
Human
ID
End-to-end

A simplest system





Method

- Background subtraction
- GEI template matching

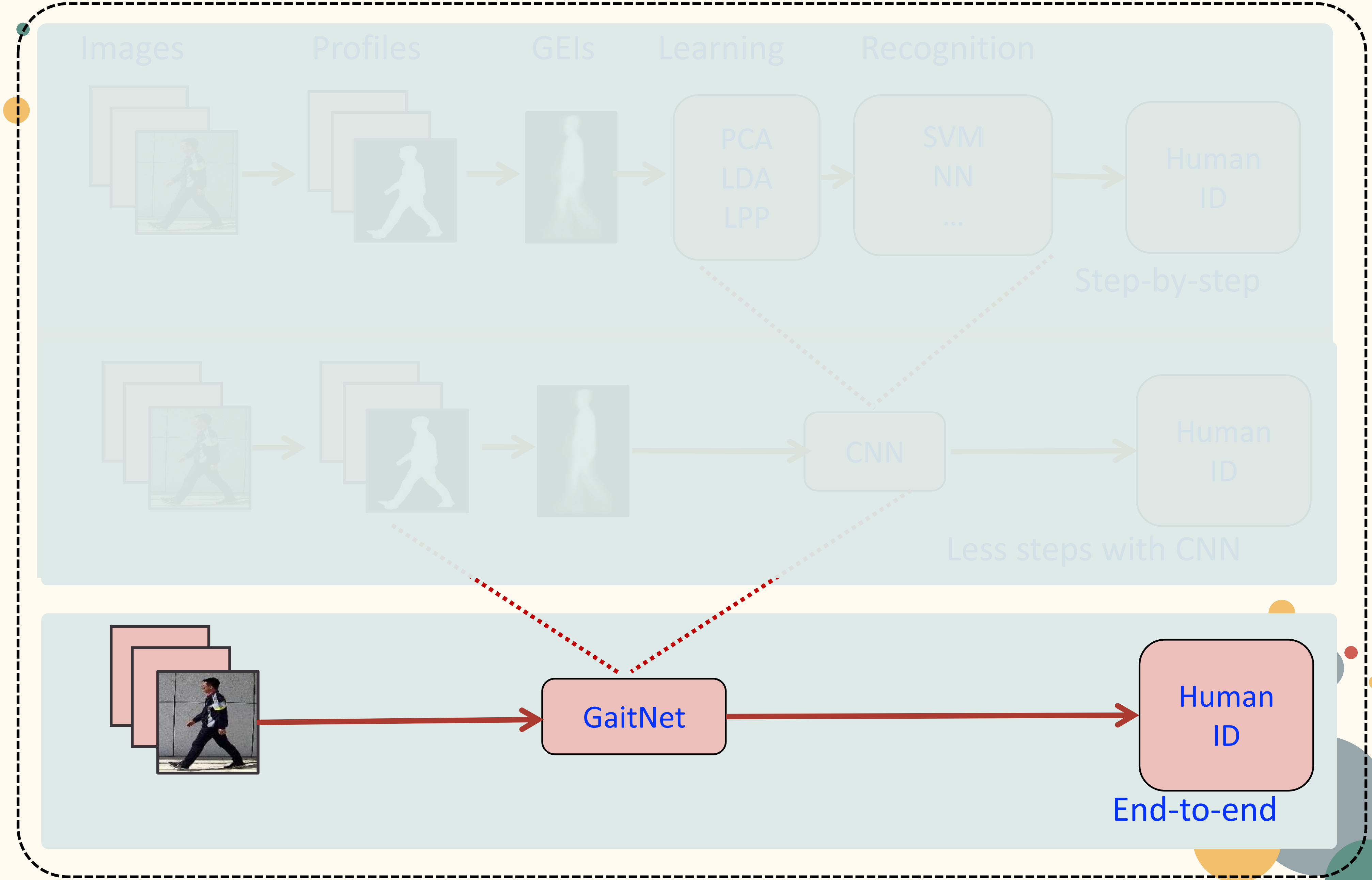
Requirement

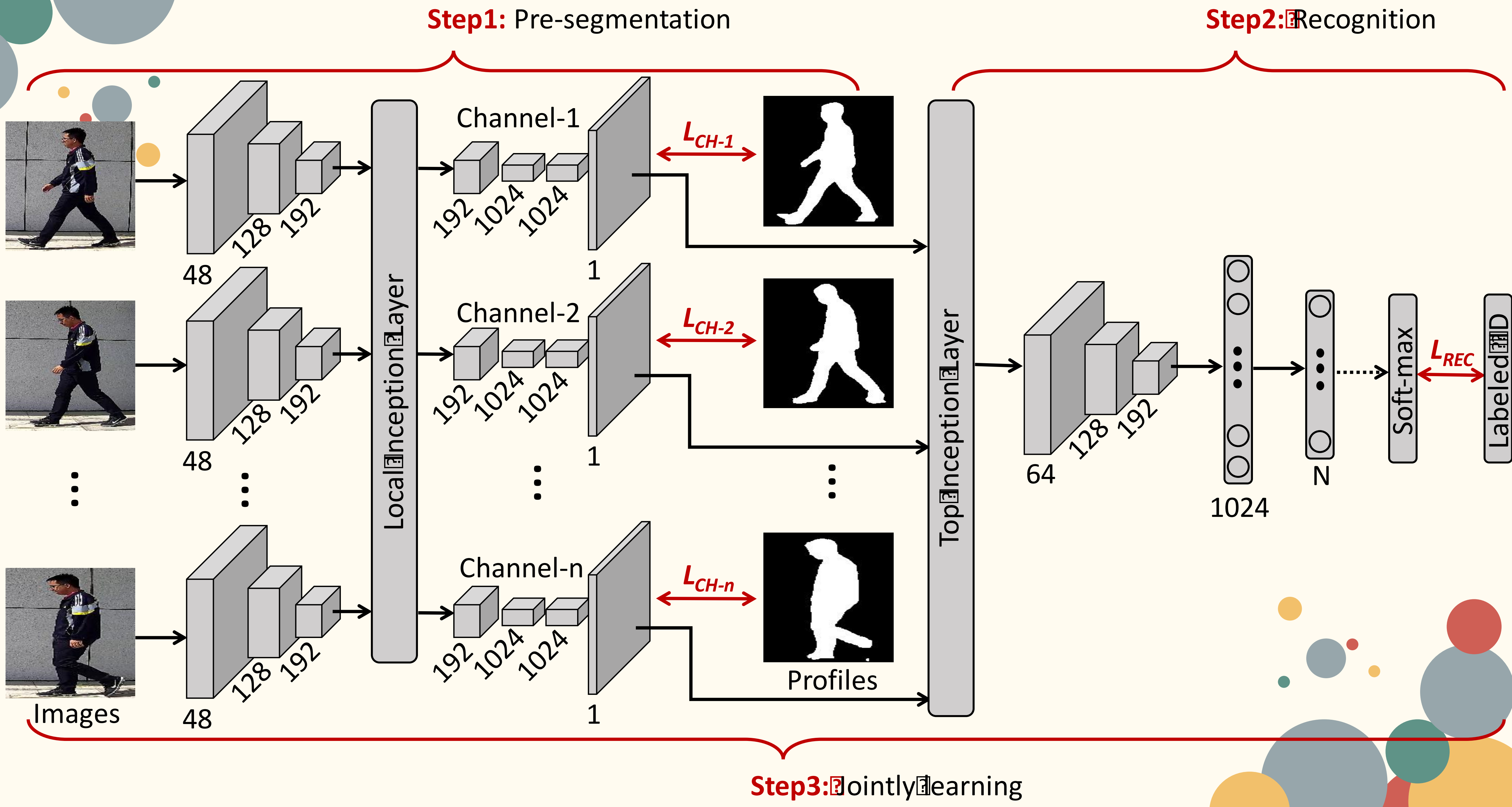
- Indoor
- Simple background and texture

Code

https://github.com/developfeng/GaitRecognition

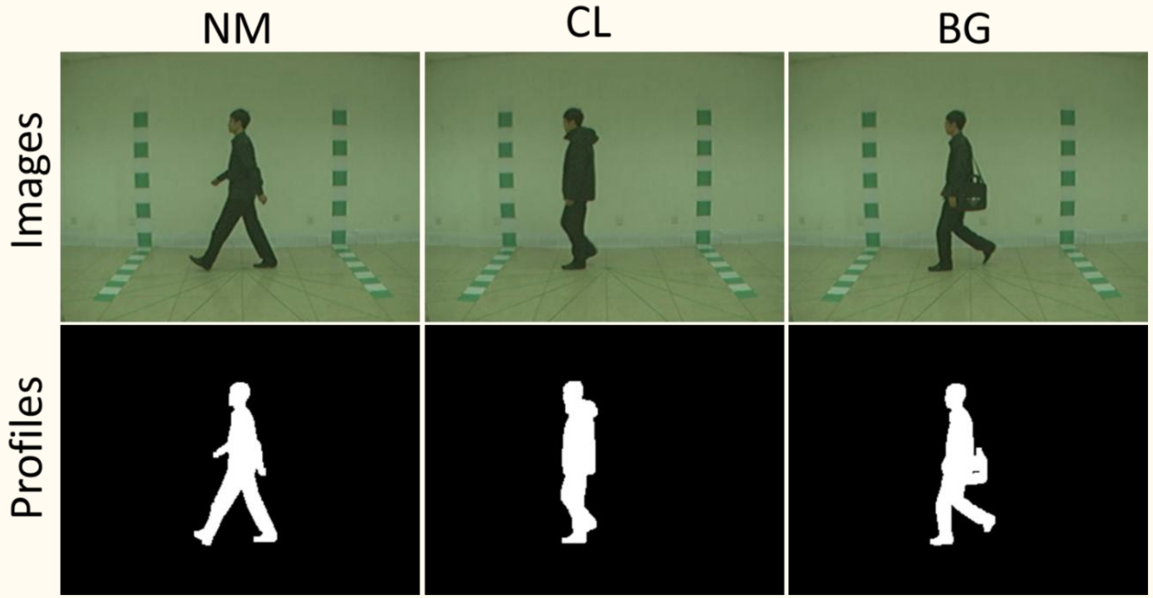
An end to end gait recognition system





C. Song, Y. Huang L. Wang, et al, A jointly learning end-to-end deep network for gait segmentation and recognition, submitted to CVPR2017.

Experimental analysis



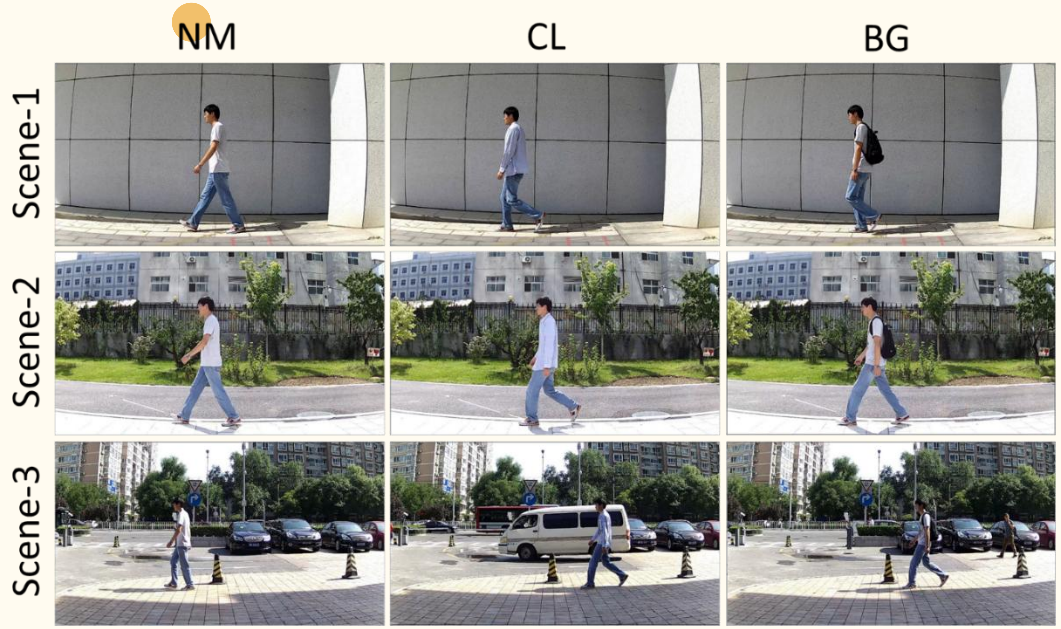
CASIA-B

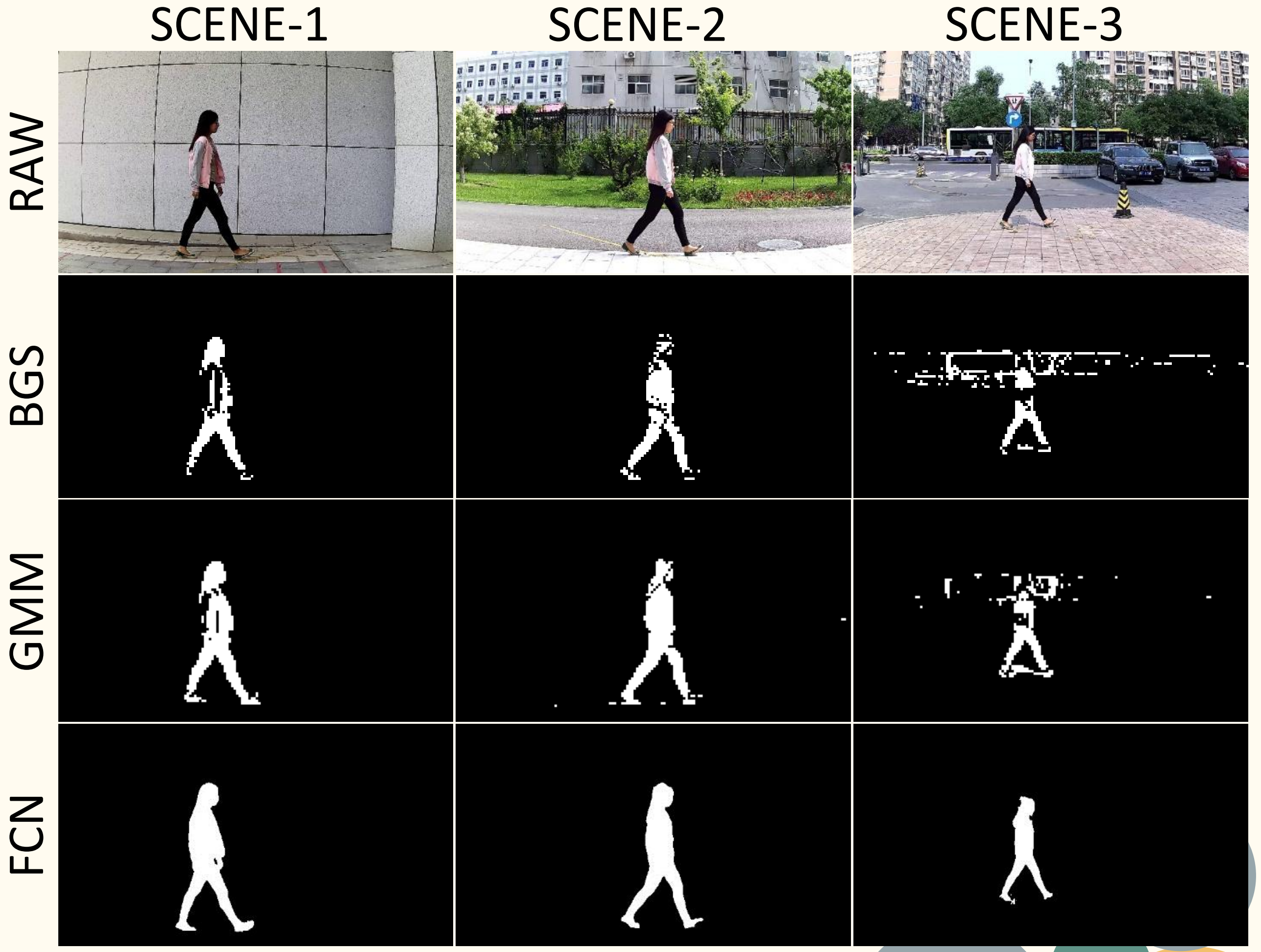
- ✧ Simple background;
- ✧ Indoor;
- ✧ Fine profiles.

Experiments-Results on CASIA-B

Method		NM	CL	BG	Mean
GEI [1]	PCA	0.9593	0.9355	0.8862	0.9270
	LDA	1.0000	0.9839	0.9837	0.9892
	LPP	1.0000	0.9758	0.9350	0.9703
GEnI [2]	PCA	0.9675	0.9597	0.8943	0.9405
	LDA	1.0000	0.9839	0.9675	0.9838
	LPP	1.0000	0.9839	0.9431	0.9757
GFI [3]	PCA	0.9675	0.9516	0.9024	0.9405
	LDA	0.9837	0.9113	0.9024	0.9325
	LPP	0.8618	0.8065	0.7642	0.8108
CGI [4]	PCA	0.9512	0.9435	0.8943	0.9297
	LDA	1.0000	1.0000	0.9675	0.9892
	LPP	1.0000	1.0000	0.9512	0.9837
GEI-CNN [5]		0.9756	0.9194	0.9024	0.9325
GaitNet	No-Joint	0.9677	0.9194	0.9113	0.9328
	Joint	1.0000	0.9919	0.9839	0.9919

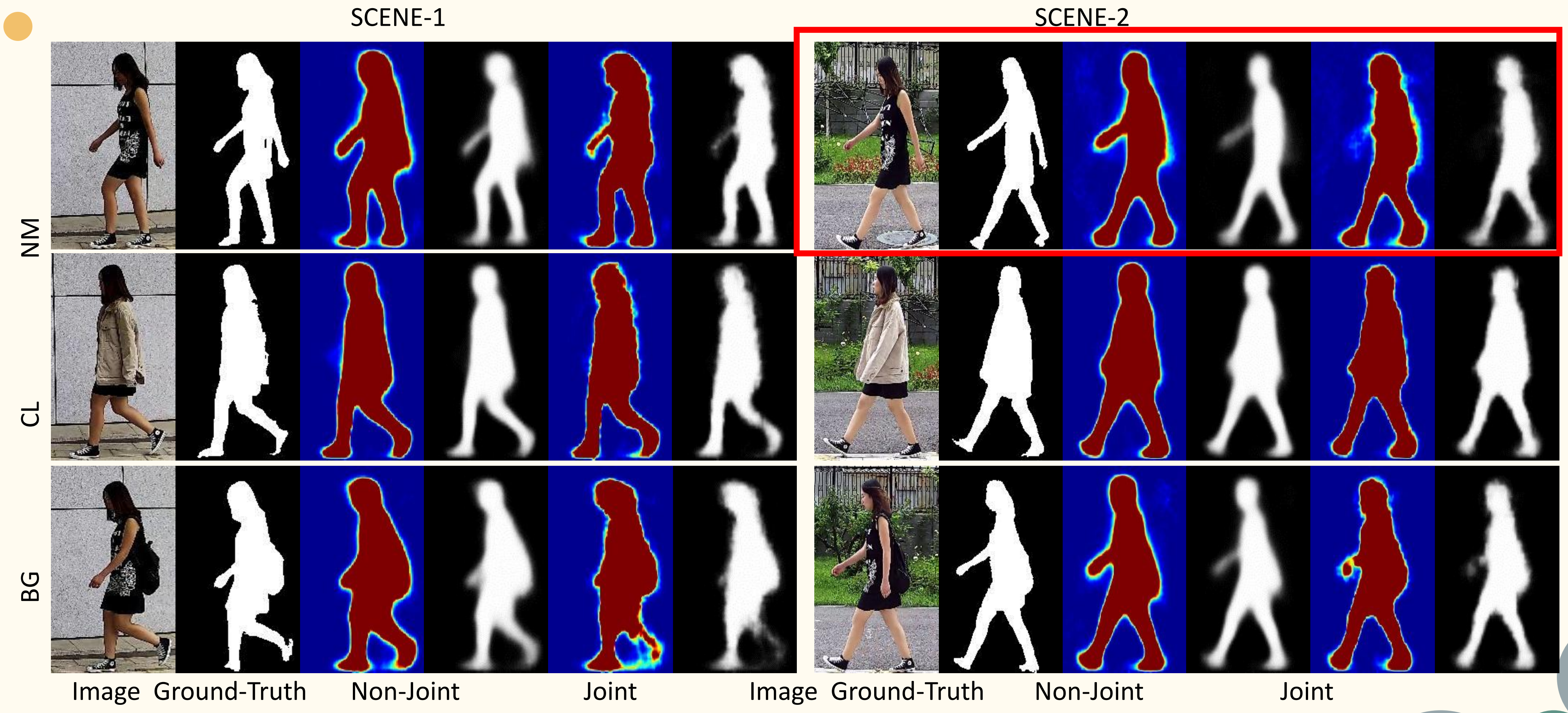
Outdoor-Gait database





- ✧ Complex background;
- ✧ Outdoor;
- ✧ Hard to get profiles.

Experiments-Joint Learning

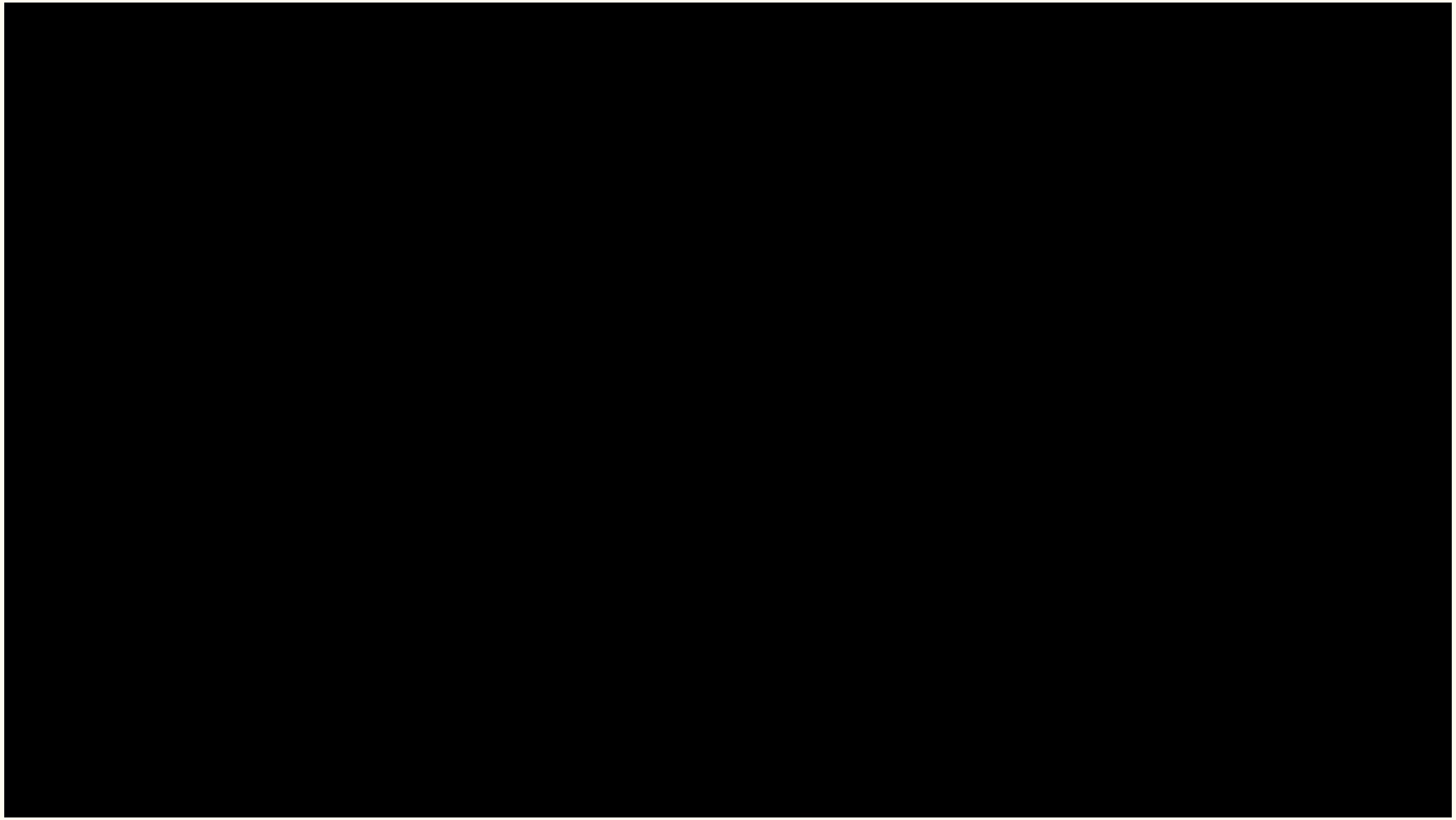


Visualization of Segmentation Network

Experiments-Results on Outdoor-Gait

Methods		S-1			S-2			S-3			Mean
		NM	CL	BG	NM	CL	BG	NM	CL	BG	
GEI [1]	PCA	0.7971	0.8456	0.8623	0.9783	0.9348	0.9638	0.6522	0.6642	0.7226	0.8245
	LDA	0.8841	0.8750	0.8623	0.9710	0.9493	0.9710	0.6087	0.6194	0.7153	0.8285
	LPP	0.8696	0.8750	0.8913	0.9348	0.9203	0.9710	0.6087	0.5970	0.7664	0.8260
GEnI [2]	PCA	0.7971	0.7868	0.7826	0.9855	0.9275	0.9638	0.5725	0.5149	0.6569	0.7764
	LDA	0.8261	0.8603	0.8478	0.9710	0.9275	0.9565	0.5870	0.5746	0.6934	0.8049
	LPP	0.8623	0.8603	0.8551	0.9348	0.9565	0.9565	0.5580	0.5821	0.7153	0.8090
GFI [3]	PCA	0.8116	0.8382	0.8768	0.9565	0.9130	0.9493	0.6667	0.5896	0.7226	0.8138
	LDA	0.7971	0.6838	0.8188	0.8841	0.8696	0.9130	0.4638	0.4328	0.5766	0.7155
	LPP	0.6667	0.6985	0.7826	0.8188	0.8623	0.8696	0.4493	0.5075	0.5329	0.6876
CGI [4]	PCA	0.7101	0.7299	0.8044	0.8696	0.8913	0.9130	0.3986	0.4105	0.5183	0.6940
	LDA	0.7101	0.6861	0.7899	0.8478	0.8841	0.9058	0.3188	0.3955	0.5037	0.6713
	LPP	0.7101	0.6861	0.7464	0.8406	0.8406	0.8696	0.3841	0.4478	0.4891	0.6683
GEI-CNN [5]		0.8623	0.9055	0.9348	0.9601	0.9565	0.9674	0.7065	0.7055	0.7681	0.8630
GaitNet	No-Joint	1.0000	0.9779	0.9816	0.9963	0.9926	0.9890	0.9779	0.9559	0.9706	0.9824
	Joint	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9963	0.9963	0.9992

Demo of near-commercial gait recognition



Outline

1. Introduction and overview

2. Traditional approaches for gait-based human identification
- History and databases
- Gait representation and learning algorithms

3. Deep networks for gait-based human identification
- Cross-view gait based human identification with deep CNNs

4. How to build a practical gait-based human identification system?
- End-to-end deep network for gait segmentation & recognition
- System demo

5. Open questions and discussion

Future directions and open questions

- Multiple overlapping persons
- Soft biometrics: attributes based gait recognition: fast query retrieval
- Speedup of deep networks/ model learning
- Super large-scale gait databases: >10,000 subjects, real world scenarios
- Multi-modal human identification: face recognition + gait recognition

References

- Z. Wu, **Y. Huang**, **L. Wang**, X. Wang, and **T. Tan**, “A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs” IEEE Trans on Pattern Analysis and Machine Intelligence (TPAMI), 2016.
- C. Wang, J. Zhang, **L. Wang**, J. Pu, X. Yuan, “Human identification using temporal information preserving gait templates”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 34(11), pp 2164-2176, 2012.
- **L. Wang**, **T. Tan**, H. Ning and W. Hu, “Silhouette analysis based gait recognition for human identification”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2003, 25(12): 1505-1518.
- **L. Wang** (**Lead Guest Editor**), G. Y. Zhao, N. Rajpoot, and M. Nixon, Special issue on new advances in video-based gait analysis and applications: challenges and solutions, IEEE Transactions on Systems, Man and Cybernetics, Part-B (TSMC-B), 2010, 40(4).
- W. Kusakunniran, Q. Wu, J. Zhang, H. Li, **L. Wang**, “Recognizing gaits across views through correlated motion co-clustering”, IEEE Transactions on Image Processing (TIP), 23(2), pp 696-709, 2014.
- **L. Wang**, **T. Tan**, W. Hu and H. Ning, “Automatic gait recognition based on statistical shape analysis”, IEEE Transactions on Image Processing (TIP), 2003, 12(9): 1120-1131.
- P. Larsen, E. Simonsen, and N. Lynnerup, “Gait analysis in forensic medicine,” Journal of Forensic Sciences, vol. 53, pp. 1149–1153, 2008.
- I. Bouchrika, M. Goffredo, J. Carter, and M. S. Nixson, “On using gait in forensic biometrics,” Journal of Forensic Sciences, vol. 56(4), pp. 882–889, 2011.
- D. Weinland, R. Ronfard, and E. Boyer, “Free viewpoint action recognition using motion history volumes,” Computer Vision and Image Understanding, vol. 104(2-3), pp. 249–257, 2006.
- A. Farhadi and M. K. Tabrizi, “Learning to recognize activities from the wrong view point,” in ECCV, 2008.
- J. Han and B. Bhanu, “Individual recognition using gait energy image,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28(2), pp. 316–322, 2006.
- S. Yu, D. Tan, and **T. Tan**, “A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition,” in ICPR, 2006.
- G. Zhao, G. Liu, H. Li, and M. Pietikainen, “3D gait recognition using multiple cameras,” in Int. Conf. Automatic Face and Gesture Recognition, 2006.
- G. Ariyanto and M. Nixon, “Model-based 3D gait biometrics,” in Int. Joint Conf. Biometrics, 2011.
- M. Goffredo, I. Bouchrika, J. Carter, and M. Nixon, “Self-calibrating view-invariant gait biometrics,” IEEE Trans. Systems, Man, and Cybernetics, Part B, vol. 40(4), pp. 997–1008, 2010.
- W. Kusakunniran, Q. Wu, J. Zhang, Y. Ma, and H. Li, “A new view invariant feature for cross-view gait recognition,” IEEE Trans. Information Forensics and Security, vol. 8(10), pp. 1642–1653, 2013.
- Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi, “Gait recognition using a view transformation model in the frequency domain,” in ECCV, 2006.
- Y. LeCun, K. Kavukvuoglu, and C. Farabet, “Convolutional networks and applications in vision,” International Symposium on Circuits and Systems, 2010.
- Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “DeepFace: Closing the gap to human-level performance in face verification,” in CVPR, 2014.

References

- A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks,” in NIPS, 2012.
- P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, “OverFeat: Integrated recognition, localization and detection using convolutional networks,” arXiv:1312.6229, 2013.
- C. Farabet, C. Couprie, L. Najman, and Y. LeCun, “Learning hierarchical features for scene labeling,” IEEE Trans. Pattern Analysis and Machine Intelligence, 2013.
- A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. sukthankar, and F. Li, “Large-scale video classification with convolutional neural networks,” in CVPR, 2014.
- R. H. S. Chopra and Y. LeCun, “Learning a similarity metric discriminatively, with application to face verification,” in CVPR, 2005.
- M. Hu, Y. Wang, Z. Zhang, J. Little, and D. Huang, “View-invariant discriminative projection for multi-view gait-based human identification,” IEEE Trans. Information Forensics and Security, vol. 8(12), pp. 2034–2045, 2013.
- H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi, “The OU-ISIR gait database: Comprising the large population dataset and performance evaluation of gait recognition,” IEEE Trans. Information Forensics and Security, vol. 7(5), pp. 1511–1521, 2012.
- C. Wang, J. Zhang, J. Pu, X. Yuan, and **L. Wang**, “Chrono-gait image: A novel temporal template for gait recognition,” in ECCV, 2010.
- T. Lam, K. Cheung, and J. Liu, “Gait flow image: A silhouette-based gait representation for human identification,” Pattern Recognition, vol. 44(4), pp. 973–987, Apr. 2010.
- W. Kusakunniran, Q. Wu, H. Li, and J. Zhang, “Multiple views gait recognition using view transformation model based on optimized gait energy image,” in Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS), 2009.
- W. kusakunniran, Q. Wu, J. Zhang, and H. Li, “Support vector regression for multi-view gait recognition based on local motion feature selection,” in CVPR, 2010.
- ——， “Gait recognition under various viewing angles based on correlated motion regression,” IEEE Trans. Circits and Systems for Video Technology, vol. 22(6), pp. 966–980, 2012.
- K. Bashir, T. Xiang, and S. Gong, “Cross-view gait recognition using correlation strength,” in BMVC, 2010.
- H. Hu, “Enhanced gabor feature based classification using a regularized locally tensor discriminant model for multiview gait recognition,” IEEE Trans. Circits and Systems for Video Technology, vol. 23(7), pp. 1274– 1286, 2013.

Acknowledgements

- Special thanks to Chunfeng Song and Zifeng Wu
- Members of Multi-Modal Computing Group & CRIPAC

THANK YOU!