Iris recognition
Towards More Accurate Iris Recognition Using Deeply Learned Spatially Corresponding Features

Zijing Zhao, Ajay Kumar
Department of Computing, The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong

swkim@dongguk.edu
sunwoo KIm

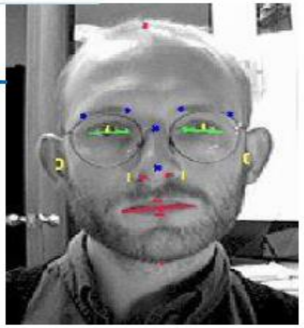
What is Biometry?

Methodology of how to identify people based on all characteristics which can be known from the organism

Dynamic signature

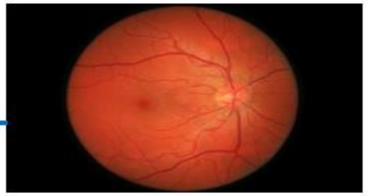
- Trajectory over time, pressure, tilt...

Face



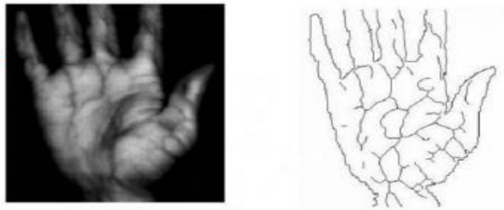
- Facial geometry or nonparam.

Retina



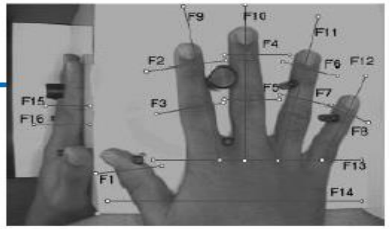
- Retinal vein patters

Vein recognition



- Vein patterns on hand

Hand geometry

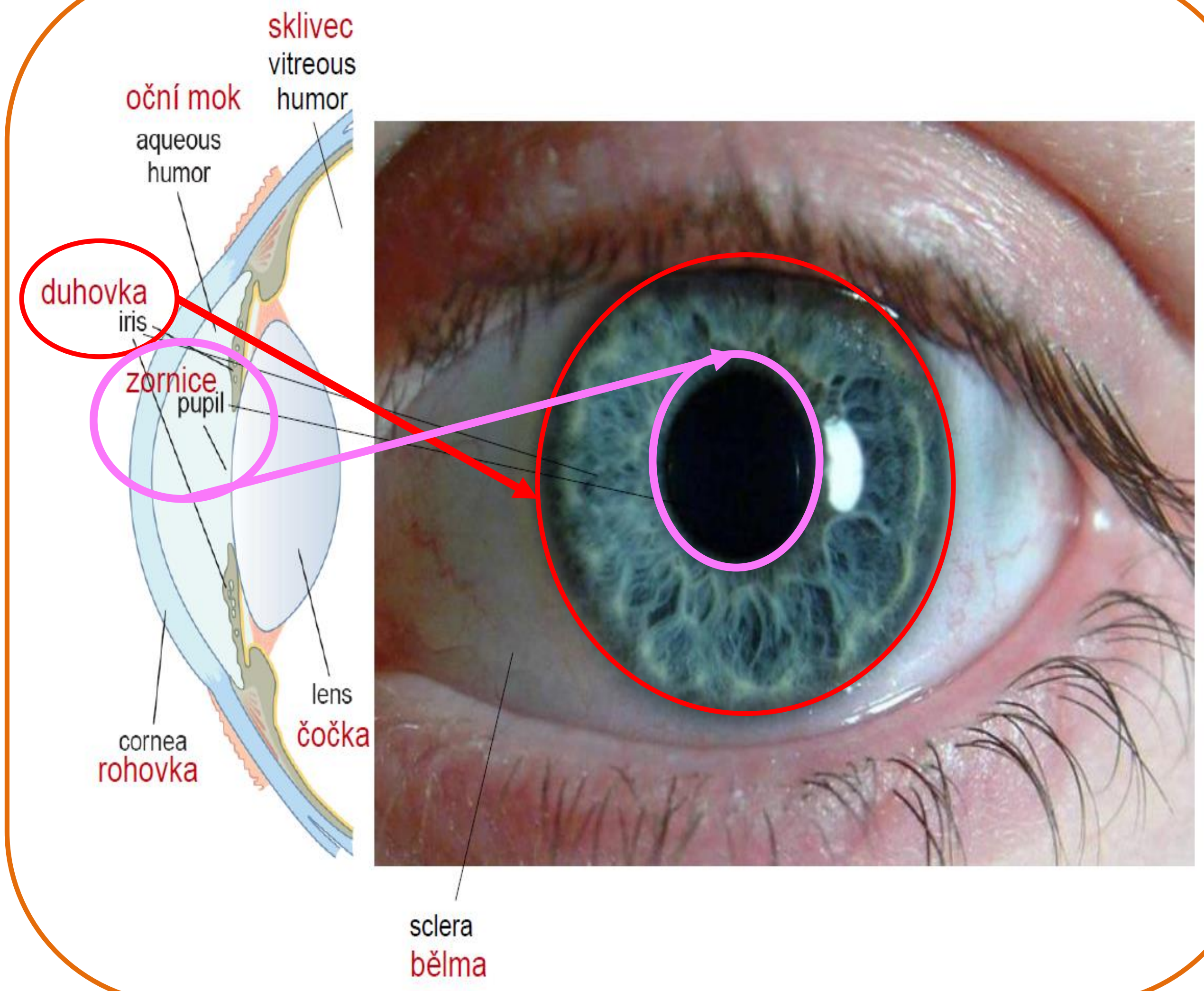


- Hand measurements at given characteristic points

Gait, voice, keystroke dynamics, lip motion...

What is the iris?

What is iris?



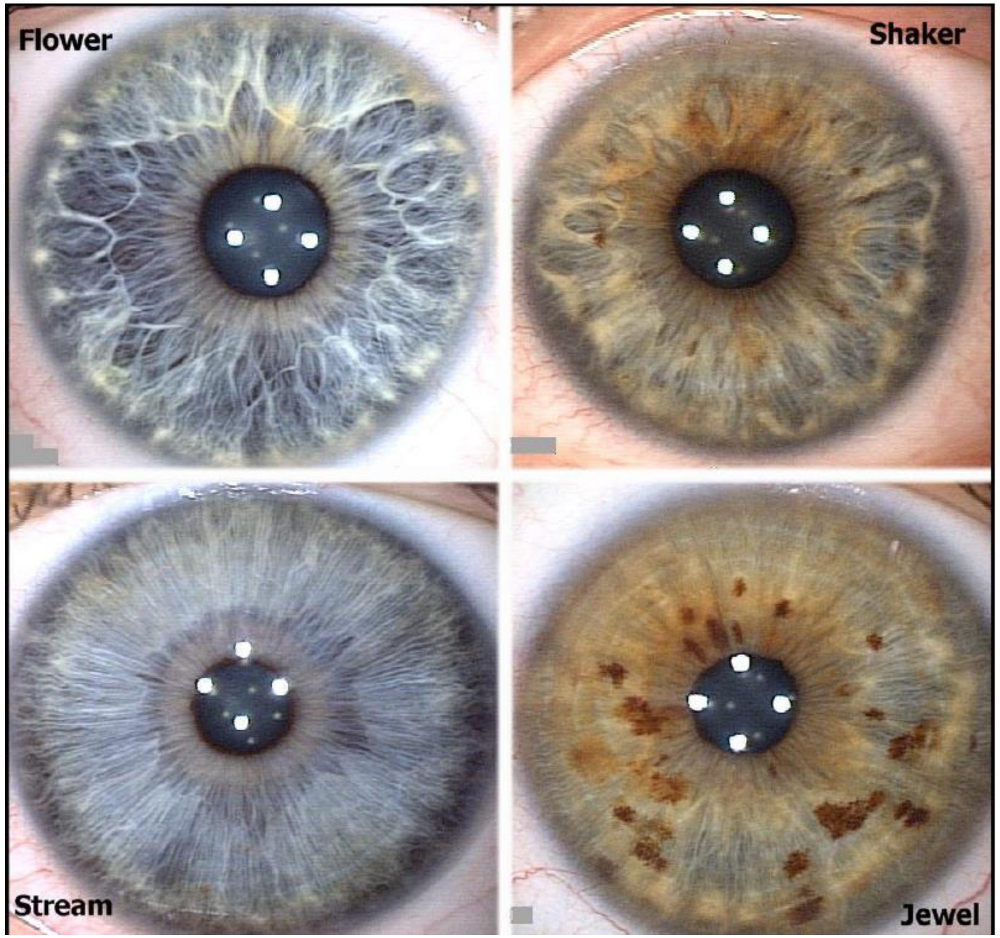
- **Pupil**

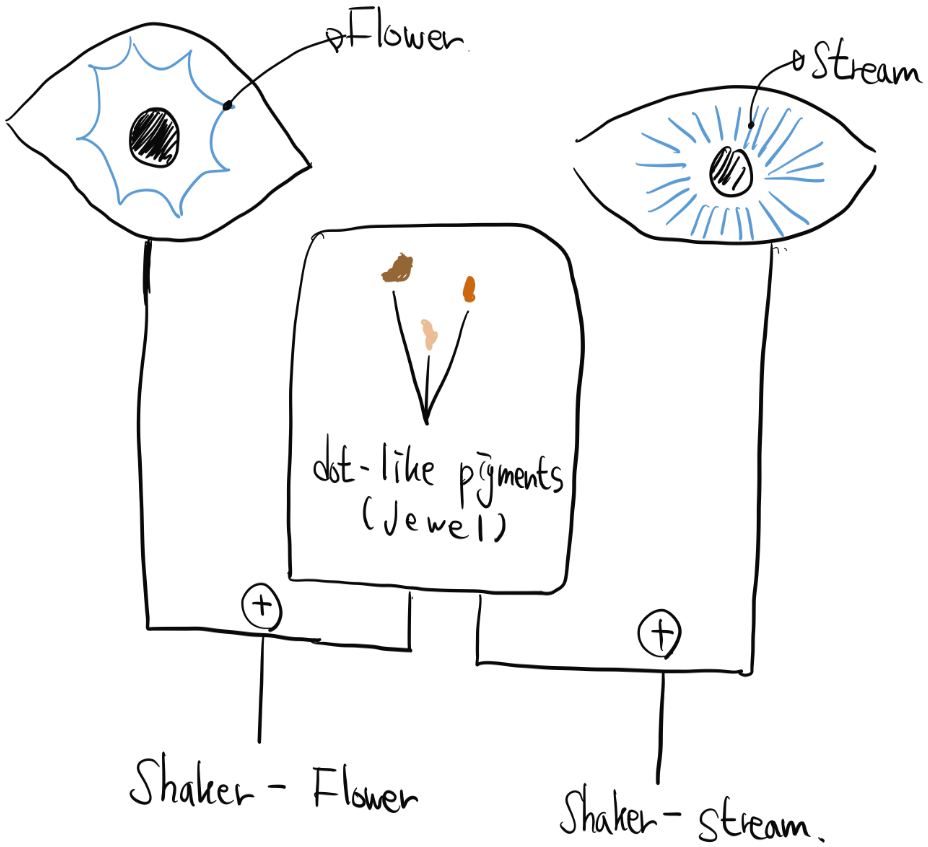
  Area which is inside of the pink boundary

- **Iris**

  Inside of the red boundary except for pupil area

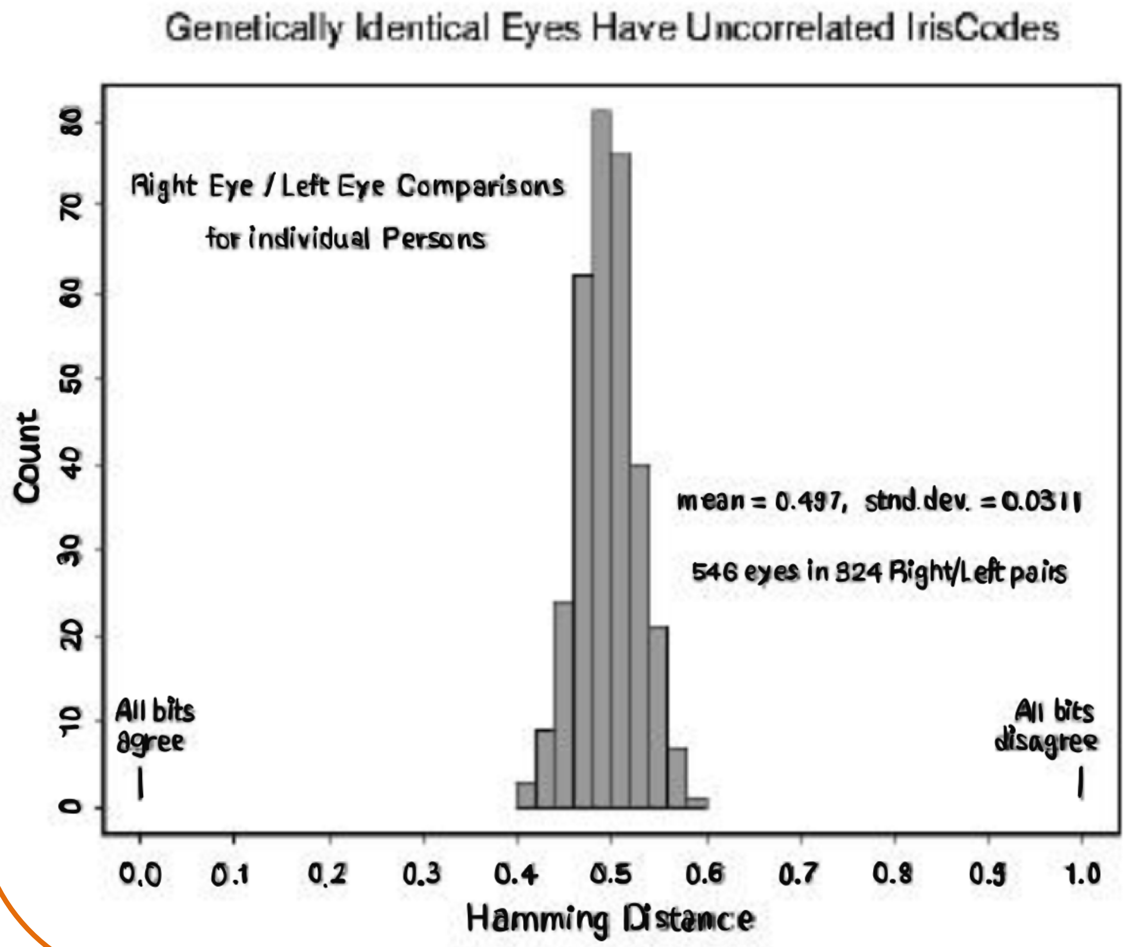
What is iris? (Patterns)





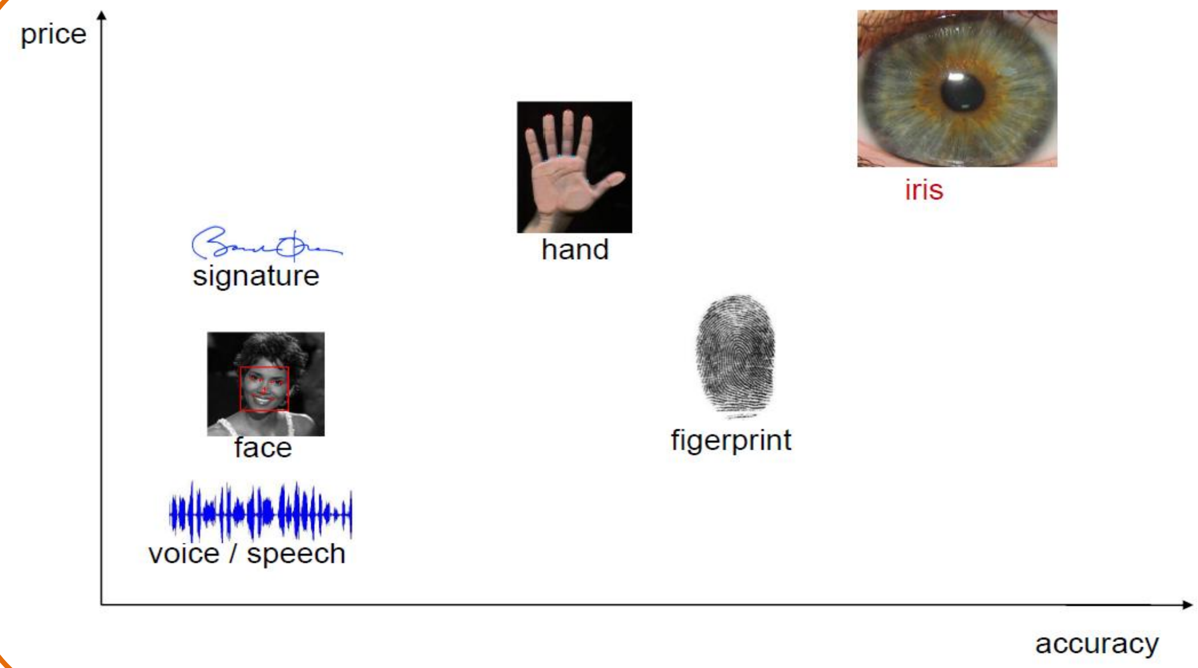
How can the iris be one of the biometry?

<Monozygotic twins>

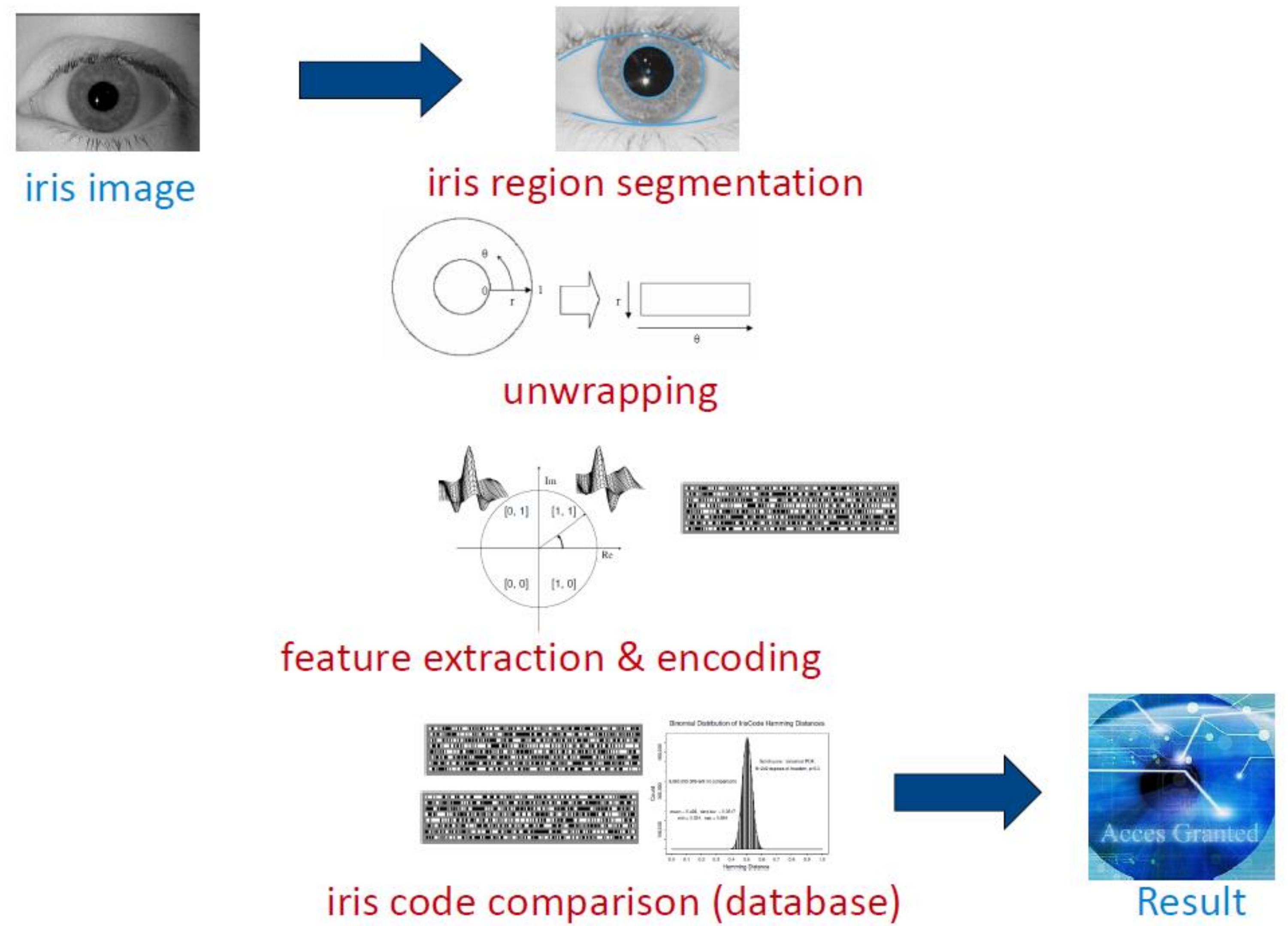


How can the iris be one of the biometry?

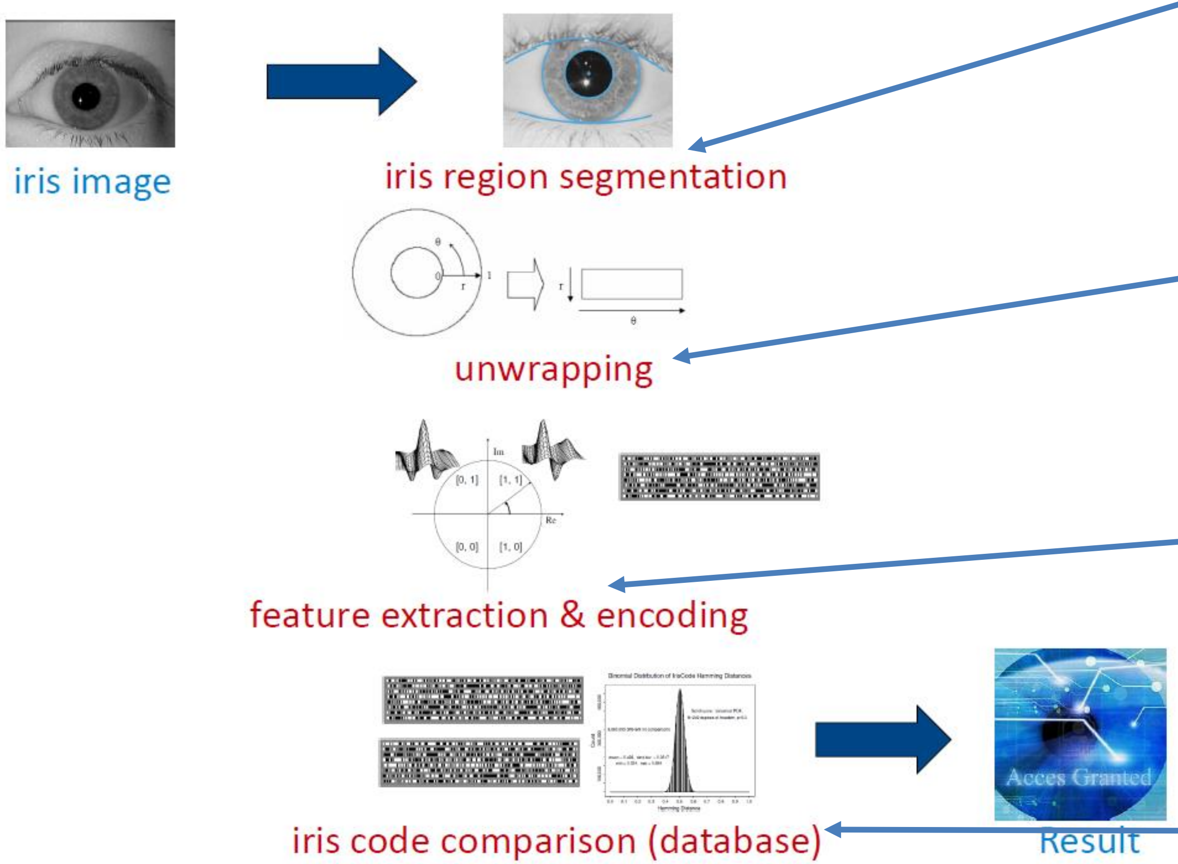
1. Everyone has Iris
2. Everyone has different iris pattern that can't be easily imitated
3. High accuracy, high speed



Overview of iris recognition process



Traditional computer vision techniques used



- Daugman Algorithm
- Hough Transform

- (x, y) plane to (r, theta) plane

- Feature extraction by Gabor Filter

- Hamming distance used

This Paper

Iris recognition
Towards More Accurate Iris Recognition Using
Deeply Learned Spatially Corresponding Features

Network Architecture (Configuration)

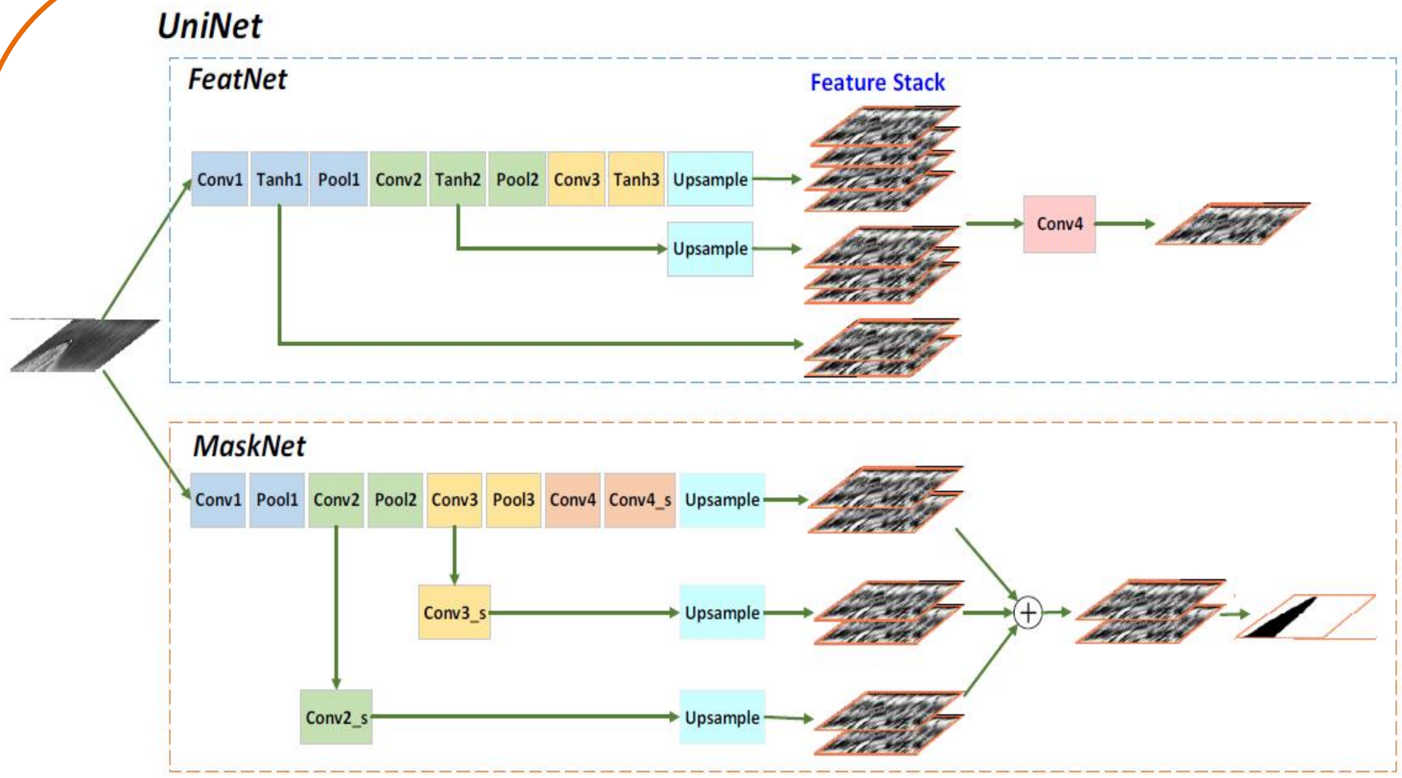


Figure 2: Detailed structures for *FeatNet* (top) and *MaskNet* (bottom) respectively. The *FeatNet* generates a single-channel feature map for each sample for matching. The *MaskNet* outputs a two-channel map, on which the values for each pixel along two channels represent the probabilities of belonging to iris and non-iris regions, respectively.

Table 1: Layer configurations for *MaskNet* and *FeatNet*.

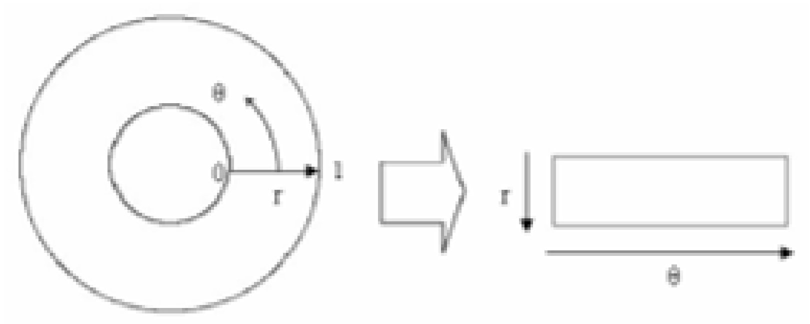
FeatNet

Layer	Type	Kernel size	Stride	# Output channels
Conv1	Convolution	3×7	1	16
Conv2	Convolution	3×5	1	24
Conv3	Convolution	3×3	1	32
Conv4	Convolution	3×3	1	1
Tanh1, 2, 3	TanH activation	/	/	/
Pool1, 2, 3	Average pooling	2×2	2	/

MaskNet

Layer	Type	Kernel size	Stride	# Output channels
Conv1	Convolution	3×3	1	16
Conv2	Convolution	3×3	1	32
Conv2_s	Convolution	1×1	1	2
Conv3	Convolution	3×3	1	64
Conv3_s	Convolution	1×1	1	2
Conv4	Convolution	3×3	1	128
Conv4_s	Convolution	1×1	1	2
Pool1, 2	Max pooling	2×2	2	/
Pool3	Max pooling	4×4	4	/

Network Architecture (Image processing)



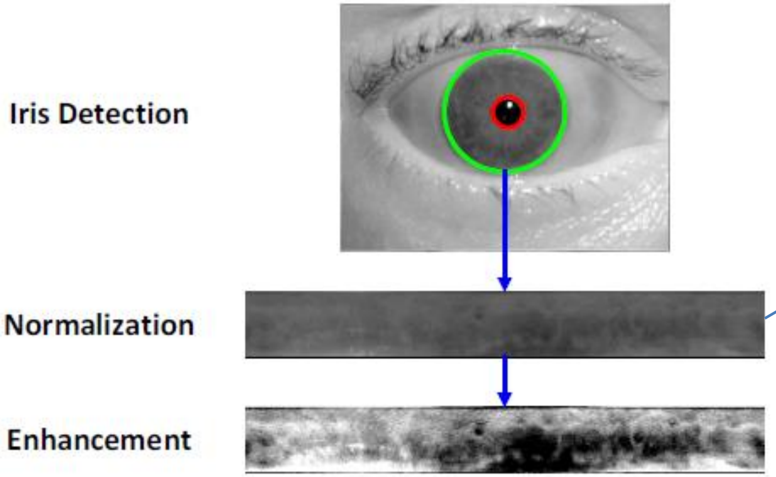
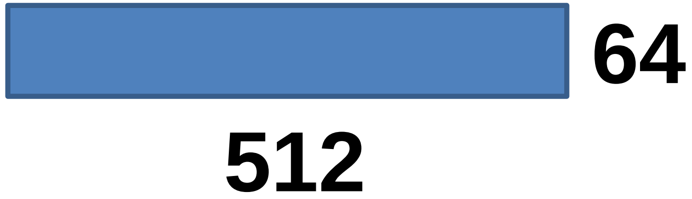
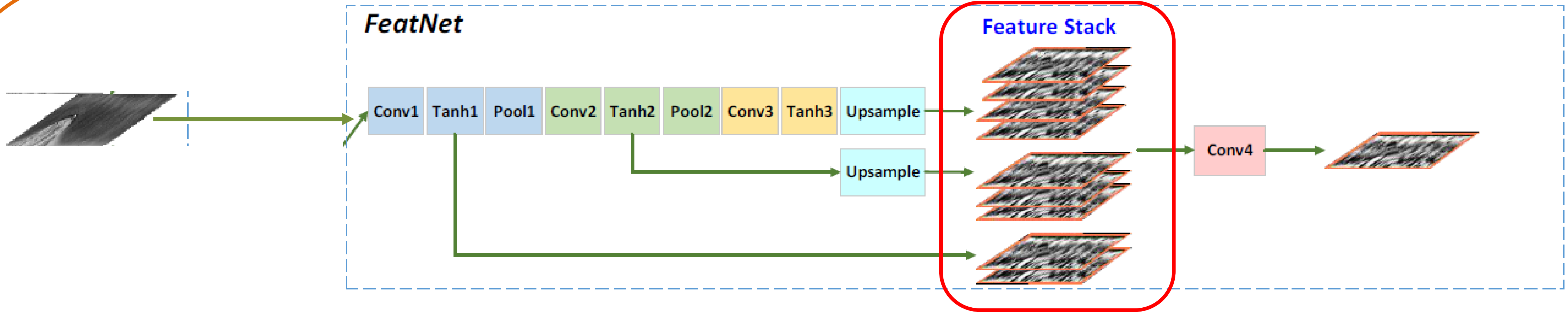


Figure 1: Illustration of key steps for iris image preprocessing.

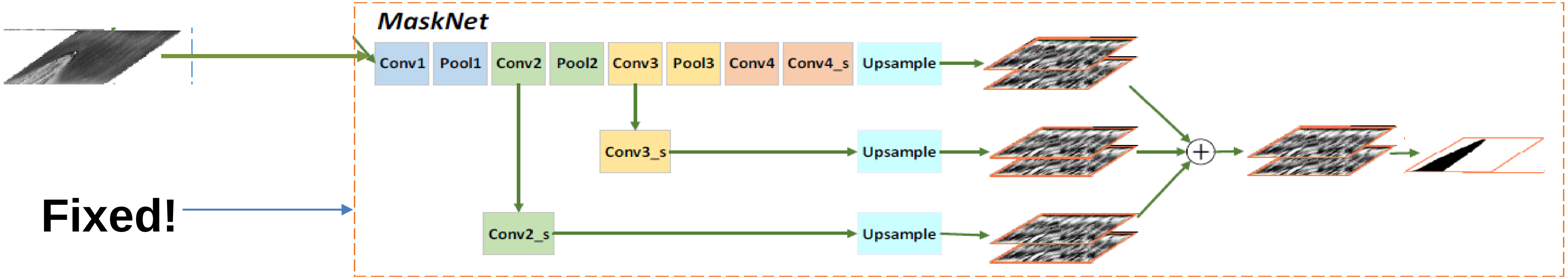


Network Architecture (FeatNet)



The reason for selecting FCN instead of CNN for iris feature extraction primarily lies in the previous analysis on iris patterns in Section 1.2, i.e., the most discriminative information of an iris probably comes from small and local patterns. FCN is able to maintain local pixel-to-pixel correspondence between input and output, and therefore is a better candidate for the iris feature extraction

Network Architecture (MaskNet)



- *MaskNet* is developed to provide adequate and immediate information for masking non-iris regions
- *MaskNet* is supervised by a pixel-wise softmax loss, where each pixel is classified into one of two classes, *i.e.*, iris or non-iris

Original Triplet Loss Function

$$L = \frac{1}{N}\sum_{i=1}^{N}\left[\left\| f^{A}_{i} - f^{P}_{i} \right\|^{2} - \left\| f^{A}_{i} - f^{N}_{i} \right\|^{2} + \alpha\right]_{+}$$

< Original triplet loss >

N : Number of triplet samples in a mini-batch

f^{A}_{i} : Feature map of anchor

f^{P}_{i} : Feature map of positive

f^{N}_{i} : Feature map of negative

α : margin

$[\bullet]_{+} = \max(\bullet, 0)$

Original Triplet Loss Function

$$L = \frac{1}{N}\sum_{i=1}^{N}\left[\left\|f^{A}_{i} - f^{P}_{i}\right\|^{2} - \left\|f^{A}_{i} - f^{N}_{i}\right\|^{2} + \alpha\right]_{+}$$

In our case, however, using Euclidean distance as the dissimilarity metric is far from sufficient. As discussed earlier, we propose using spatial features which have the same resolution with the input, the matching process has to deal with non-iris region masking and horizontal shifting

Extended Triplet Loss Function

It can be summarized

$$ETL = \frac{1}{N}\sum_{i=1}^{N}\left[D(f^{A}_{i}, f^{P}_{i}) - D(f^{A}_{i}, f^{N}_{i}) + \alpha\right]_{+}$$

Triplet Loss Function with sum of valid pixels' intensity of among A,P,N

$$L = \frac{1}{N}\sum_{i=1}^{N}\left[\left\| f^{A}_{i} - f^{P}_{i}\right\|^{2} - \left\| f^{A}_{i} - f^{N}_{i}\right\|^{2} + \alpha\right]_{+}$$

Triplet Loss Function with Euclidean distance of among A,P,N

Extended Triplet Loss Function

$$ETL = \frac{1}{N}\sum_{i=1}^{N}\left[D(f^{A}{}_{i}, f^{P}{}_{i}) - D(f^{A}{}_{i}, f^{N}{}_{i}) + \alpha\right]_{+}$$

$D(f^{1}, f^{2}) = \min_{-B \le b \le B}\{FD(f_b^{1}, f^{2})\}$ **: Minimum Shifted and Masked Distance (MMSD)**

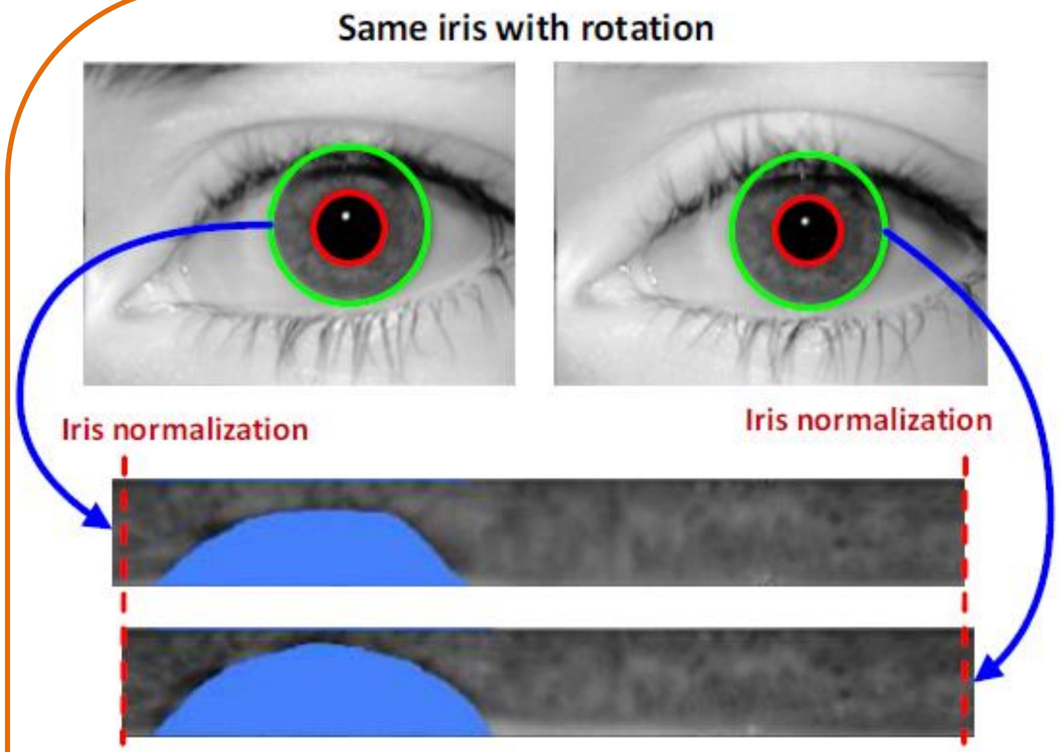
$FD(f^{1}, f^{2}) = \frac{1}{|M|}\sum_{(x,y)\in M}(f^{1}{}_{x,y} - f^{2}{}_{x,y})^2$ **: Fractional distance**

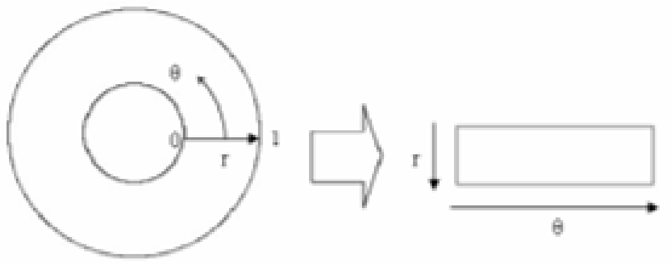
$M = \{(x,y) \mid m^{1}{}_{x,y} \neq 0 \text{ and } m^{2}{}_{x,y} \neq 0\}$ **: Set of the iris pixels' coordinates (not non-iris)**

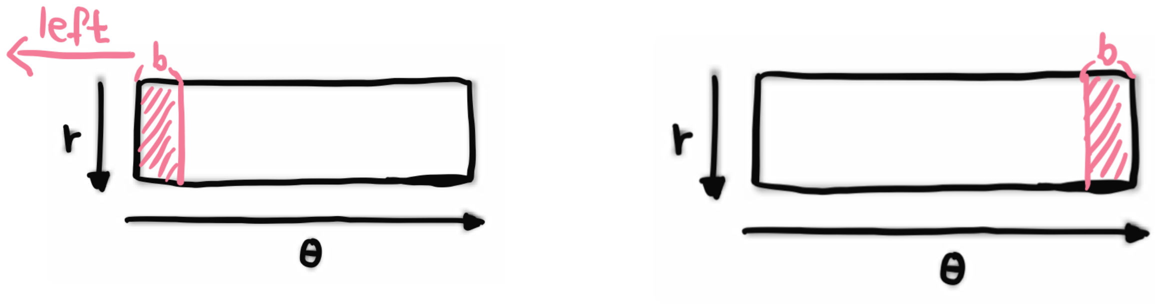
b **: number of pixels that feature map has been shifted left horizontally (It is for rotated image)**

$$ETL = \frac{1}{N}\sum_{i=0}^{N}\left[\min_{-B\le b\le B}\left\{\frac{1}{|M|}\sum_{(x,y)\in M}\left(f_{b\,x,y}^{A^i} - f_{x,y}^{P^i}\right)^2\right\} - \min_{-B\le b\le B}\left\{\sum_{(x,y)\in M}\left(f_{b\,x,y}^{A^i} - f_{x,y}^{N^i}\right)^2\right\}\right]_{+}$$

About ‘b’ pixel shifting







When input image is rotated with some degree of ‘theta’, coordinate in (r, theta) plane is changed. Let’s suppose, a point’s coordinate is (3, 30º). And rotate the image 50º counterclockwise. It’s coordinate will be (3, -20º)=(3, 340º)

Minimizing FD scope of –B<=b<=B can be inferred that it will be for remove the error caused by input image rotation.

Calculating gradient

dongguk UNIVERSITY

$$ETL = \frac{1}{N}\sum_{i=1}^{N}\left[D(f^A{}_i, f^P{}_i) - D(f^A{}_i, f^N{}_i) + \alpha\right]_+$$

$$\frac{\partial ETL}{\partial f^P} = \begin{cases} 0, & \text{if } ETL = 0 \\ \frac{1}{N}\frac{\partial ETL}{\partial D(f^A, f^P)}\frac{\partial D(f^A, f^P)}{\partial f^P}, & \text{otherwise} \end{cases}, \quad \frac{\partial ETL}{\partial D(f^A, f^P)} = 1$$

$$b_{AP} = \underset{-B \le b \le B}{\operatorname{argmin}}\left\{FD(f^A{}_b, f^P)\right\}$$

$$b_{AN} = \underset{-B \le b \le B}{\operatorname{argmin}}\left\{FD(f^A{}_b, f^N)\right\}$$

: value ‘b’ that minimize FD

$$D(f^1, f^2) = \min_{-B \le b \le B}\left\{\frac{1}{|M|}\sum_{(x,y)\in M}\left(f^1_{x,y} - f^2_{x,y}\right)^2\right\}$$

$$\frac{\partial D(f^A, f^P)}{\partial f^P[x,y]} = \frac{\partial FD(f^A{}_{b_{AP}}, f^P)}{\partial f^P[x,y]} = \begin{cases} 0, & \text{if } (x,y) \notin M_{AP} \text{ or } ETL = 0 \\ \frac{-2}{|M_{AP}|}(f^A[x_{b_{AP}}, y] - f^P[x,y]), & \text{otherwise} \end{cases}$$

$$\frac{\partial ETL}{\partial f^P[x,y]} = \begin{cases} 0, & \text{if } (x,y) \notin M_{AP} \text{ or } ETL = 0 \\ \frac{-2(f^A[x_{b_{AP}}, y] - f^P[x,y])}{N|M_{AP}|}, & \text{otherwise} \end{cases}$$

$$\frac{\partial ETL}{\partial f^N[x,y]} = \begin{cases} 0, & \text{if } (x,y) \notin M_{AN} \text{ or } ETL = 0 \\ \frac{2(f^A[x_{b_{AN}}, y] - f^N[x,y])}{N|M_{AN}|}, & \text{otherwise} \end{cases}$$

$$FD(f^A{}_{b_{AP}}, f^P) = FD(f^A, f^P{}_{-b_{AP}})$$

$$\frac{\partial ETL}{\partial f^A[x,y]} = -\frac{\partial ETL}{\partial f^P[x_{-b_{AP}}, y]} + \frac{\partial ETL}{\partial f^N[x_{-b_{AN}}, y]}$$

Feature encoding & matching

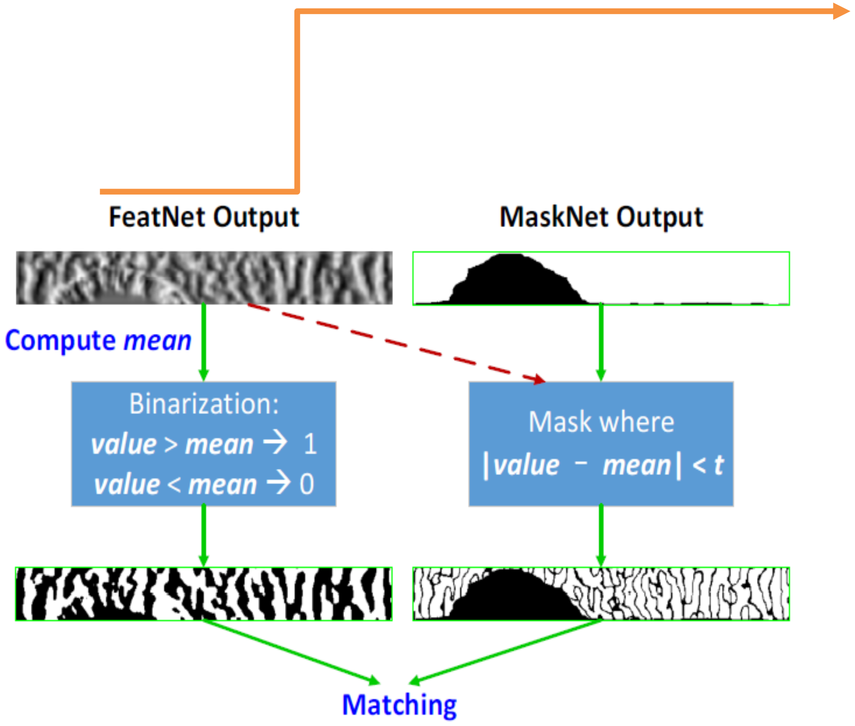
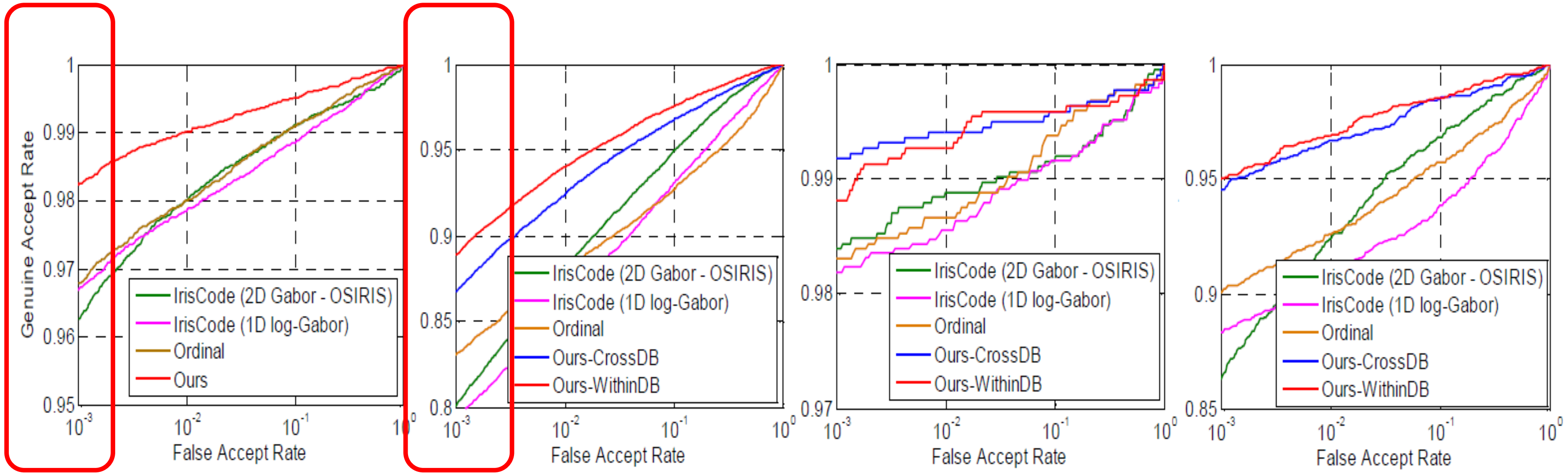


Figure 5: Illustration of feature binarization process.

It is widely accepted by the community that binary features are more resistant to illumination change, blurring and other underlying noise. Besides, binary features consume smaller storage and enable faster matching

The mean value of the elements within the non-masked iris regions is firstly computed as *m*. This mean value is then used as the threshold to binarize the original feature map. In order to avoid marginal errors, elements with feature values *v* close to *m* (i.e., $|v-m|<t$) are regarded as less reliable and will be masked(as non-iris) together with the original mask output by *MaskNet*

Results



(a) ND-IRIS-0405 (b) CASIA.v4-distance (c) IITD (d) WVU Non-ideal

Figure 7: ROCs for comparison with other state-of-the-art methods on for employed databases. *Best viewed in color.*

Table 2: Summary of false reject rates (FRR) at **0.1%** false accept rate (FAR) and equal error rates (EER) for the comparison.

	ND-IRIS-0405		CASIA.v4-distance		IITD		WVU Non-ideal	
	FRR	EER	FRR	EER	FRR	EER	FRR	EER
IrisCode (OSIRIS)	3.73%	1.70%	19.93%	6.39%	1.61%	1.11%	13.70%	4.43%
IrisCode (log-Gabor)	3.31%	1.88%	20.72%	7.71%	1.81%	1.38%	11.63%	6.82%
Ordinal	3.22%	1.74%	16.93%	7.89%	1.70%	1.25%	9.89%	5.19%
Ours-CrossDB	/	/	13.27%	4.54%	**0.82%**	**0.64%**	5.46%	2.83%
Ours-WithinDB	**1.78%**	**0.99%**	**11.15%**	**3.85%**	1.19%	0.73%	**5.00%**	**2.28%**

- **FAR(False Accept Rate)**
  : Rate how falsely accept non-authorized iris rate
- **GAR(Genuine Accept Rate)**
  : Rate how correctly accept authorized iris rate
- **FRR(False reject rate)**
  : Rate how falsely reject authorized iris r
- **EER(Equal error rate)**
  : Rate when FRR equals FAR

Result(comparison with other DL model)

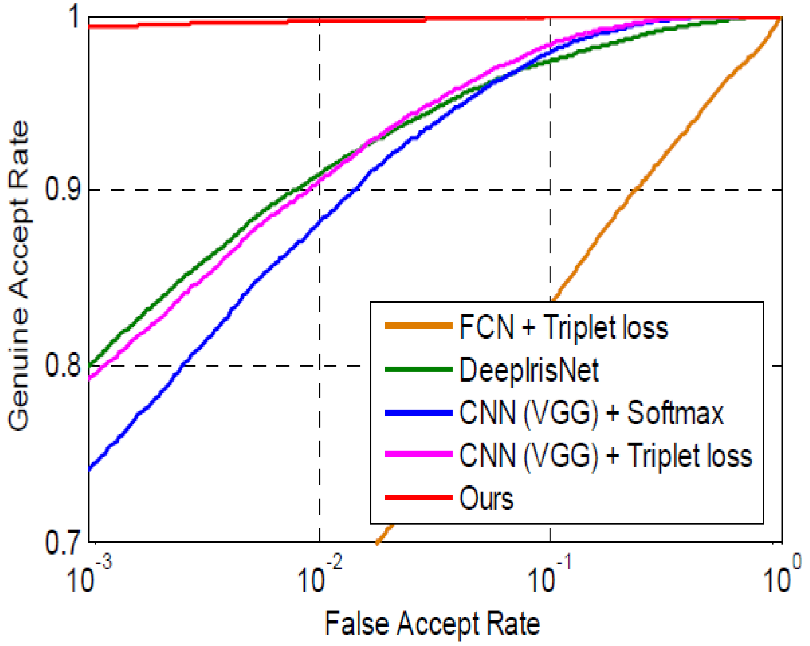


Figure 8: ROC curves for typical deep learning architectures available in the literature and our method on ND-IRIS-0405.

- **Poorness CNN**
  : It can't represent iris' detailed, rare, local features

- **Poorness FCN+TPL**
  : bit-shifting & non-iris masking is necessary

Thank you!