Biometric Quality Workshop
National Institute of Standards and Technology, March 8-9, 2006

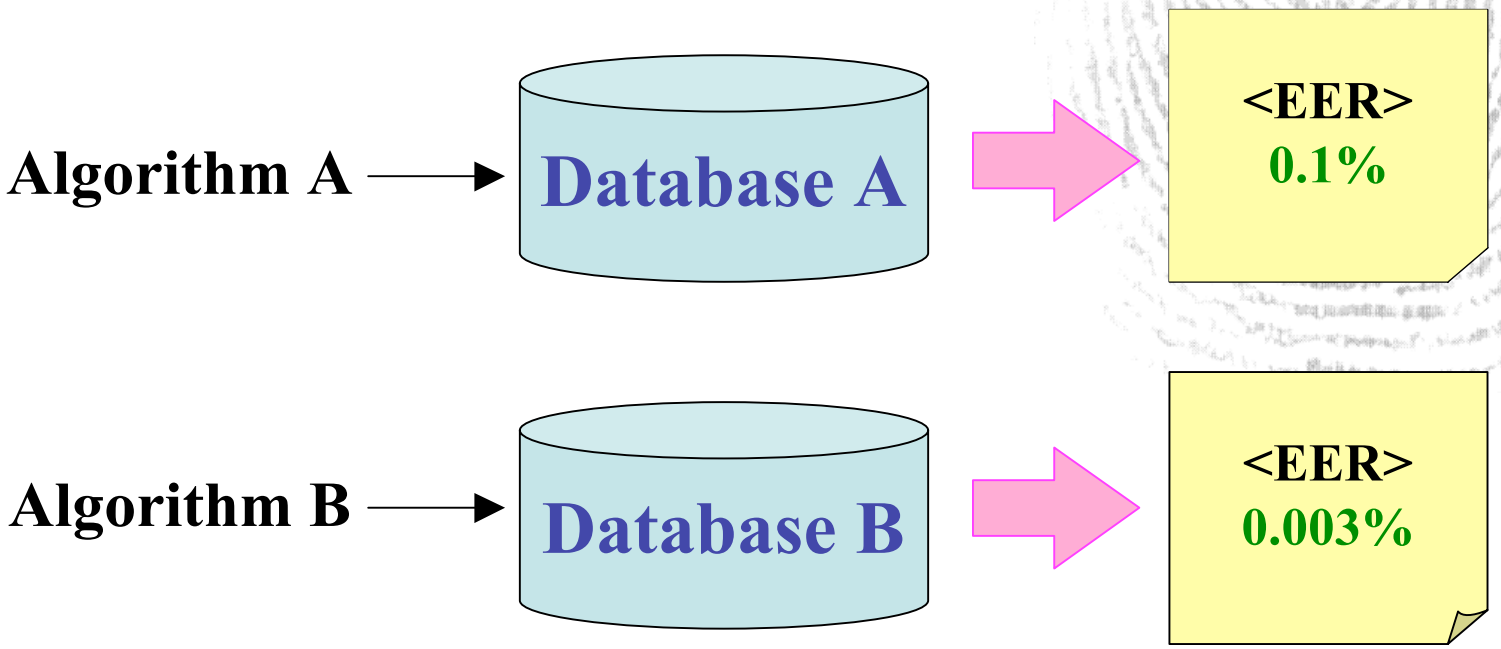
Measuring the Quality of Biometric Database

Hale Kim and Jihyun Moon

Inha University
&
Biometrics Engineering Research Center
Korea

Introduction

❖ Motivation 1

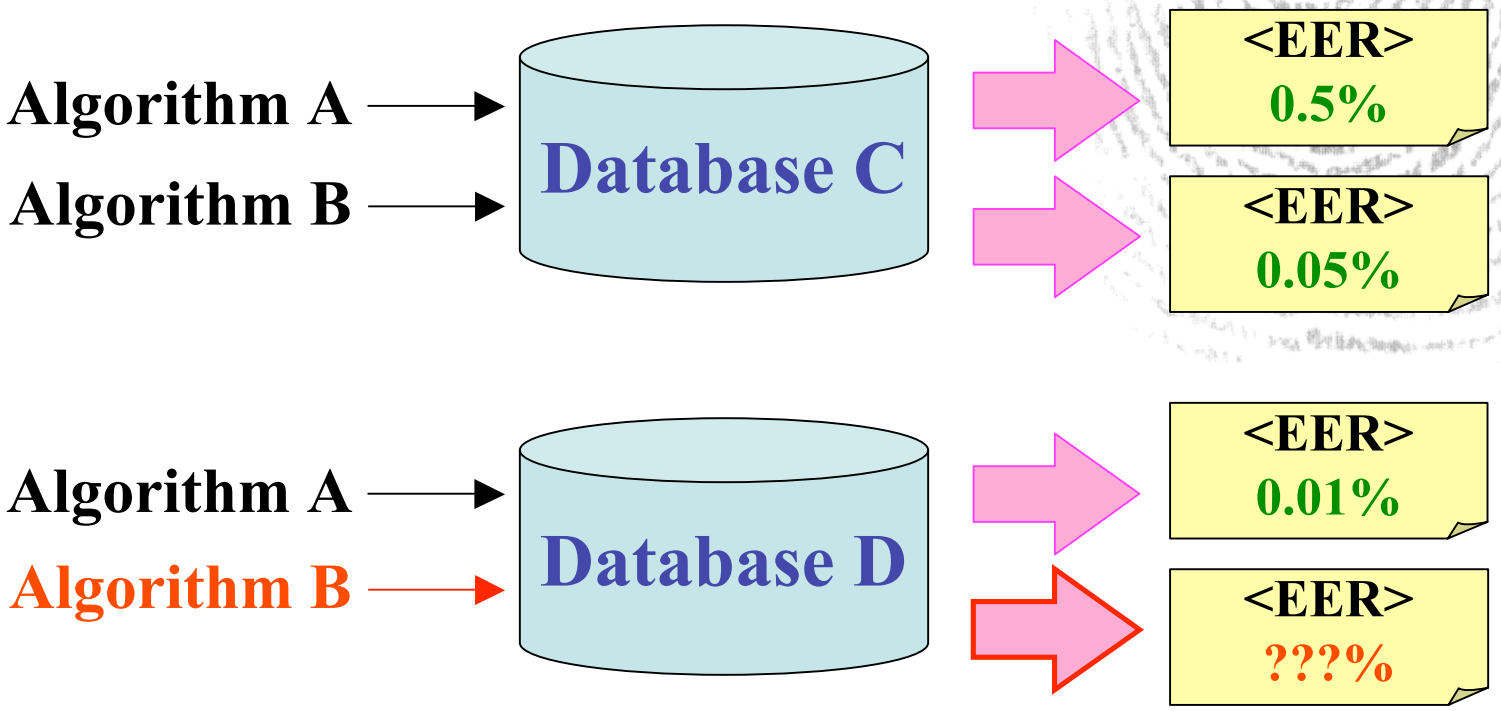


✓ Is Algorithm B better than Algorithm A?

CVLab

Introduction

❖ Motivation 2



✓ Can I predict the performance of Algorithm B without actual testing over Database D?

CVLab

Introduction

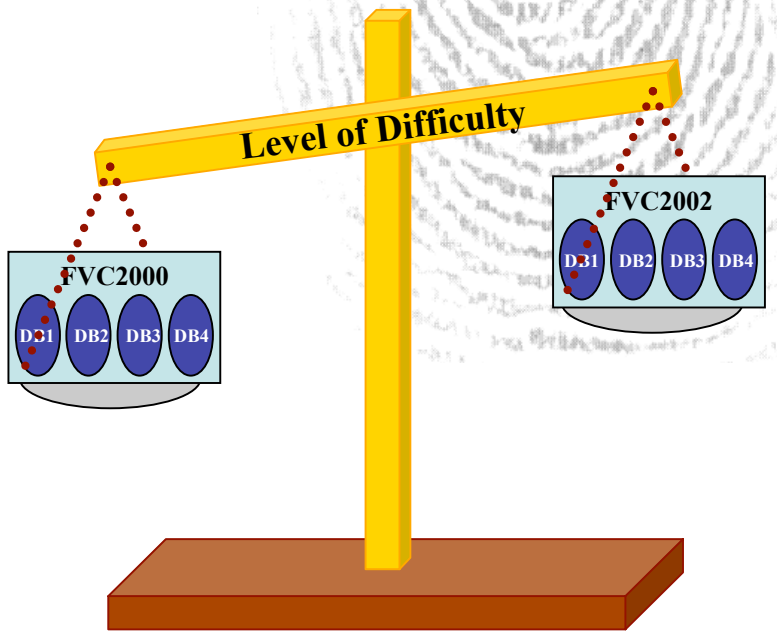
- Purpose
  - To develop testing and evaluation methodologies for quantifying and comparing "Level of Difficulty (LoD)" of biometric databases t hat are collected for performance evaluation of biometric recognit ion algorithms

- Scope
  - Developing measures for evaluating LoD
    - Defining objective measures representing LoD
    - Quantifying the measures
  - Developing methods for testing and evaluating LoD
    - Defining procedures for testing and reporting
    - Predicting the performance of recognition algorithms on differ ent databases

CVLab

Definitions

- Level of Difficulty for biometric databases
  - Grades or scores quantifying the overall characteristics of biometric databases that influence the performance of biometric recognition algorithms
  - Integration of measures of various influencing factors which degrade the performance of genuine matching

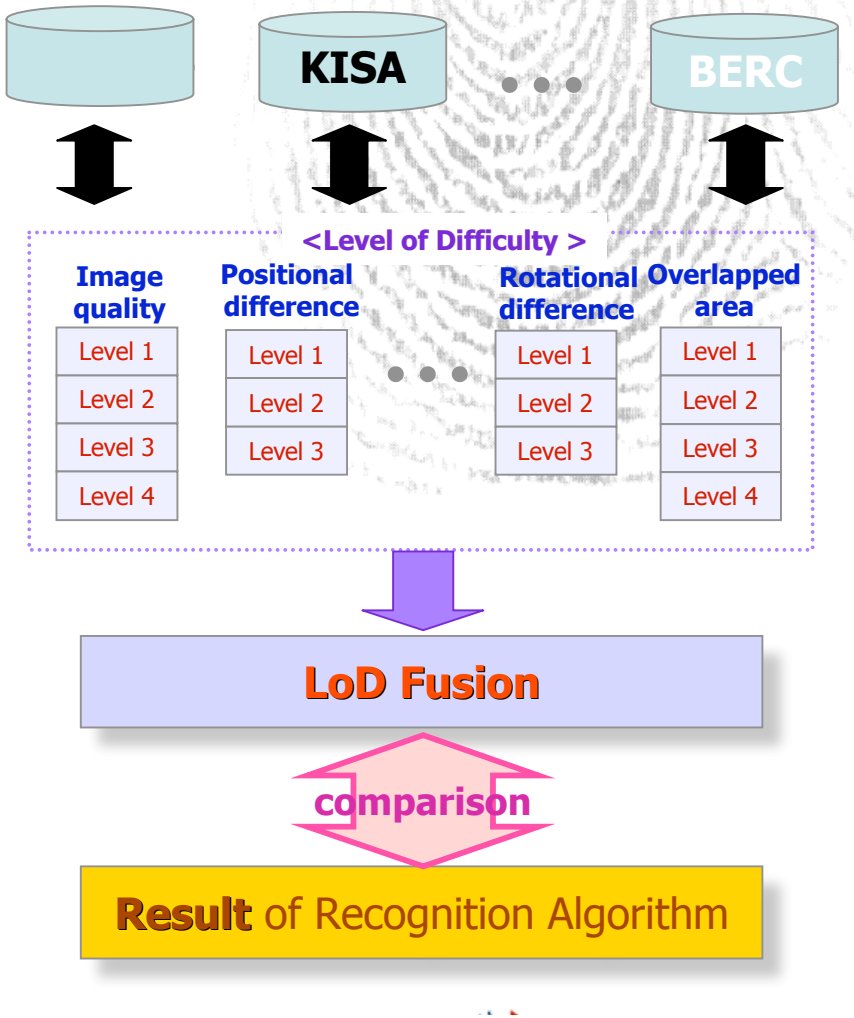


- ✓ How to objectively measure the similarity of impostor pairs?
- ✓ Uniqueness is one of underlying hypothesis for biometrics.

CVLab

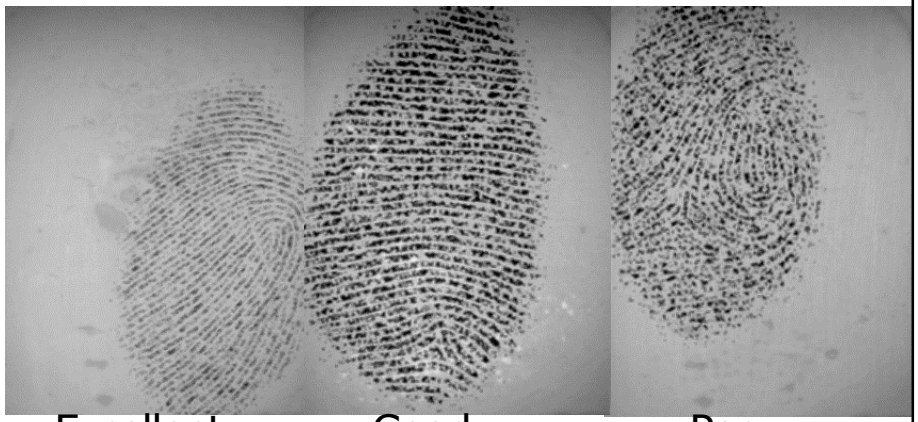
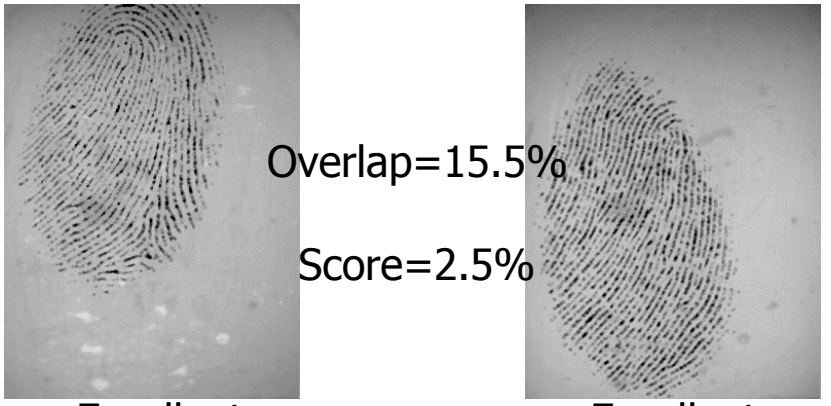
Definitions

- Components of LoD
  - Attributes of a biometric database that are to be measured in order to quantify its LoD
  - A subset of influencing factors to the performance of a recognition algorithm, which represent differences between a pair of genuine biometric samples
  - LoD can be obtained by combining the measures of the attributes



Process of measuring and evaluating LoD

Sample Quality vs. DB Quality

Sample Quality	Database Quality (LoD)
❖ Quality of a single image sample ❖ Directly affects the performance of feature extractors, and indirectly affects the matching performance	❖ Aggregation of various differences between genuine pairs ❖ Directly affects the matching performance with no influence to feature extractor ❖ Quality of genuine sample pairs is an important aspect of DB quality
 Excellent Good Poor	 Overlap=15.5% Score=2.5% Excellent Excellent

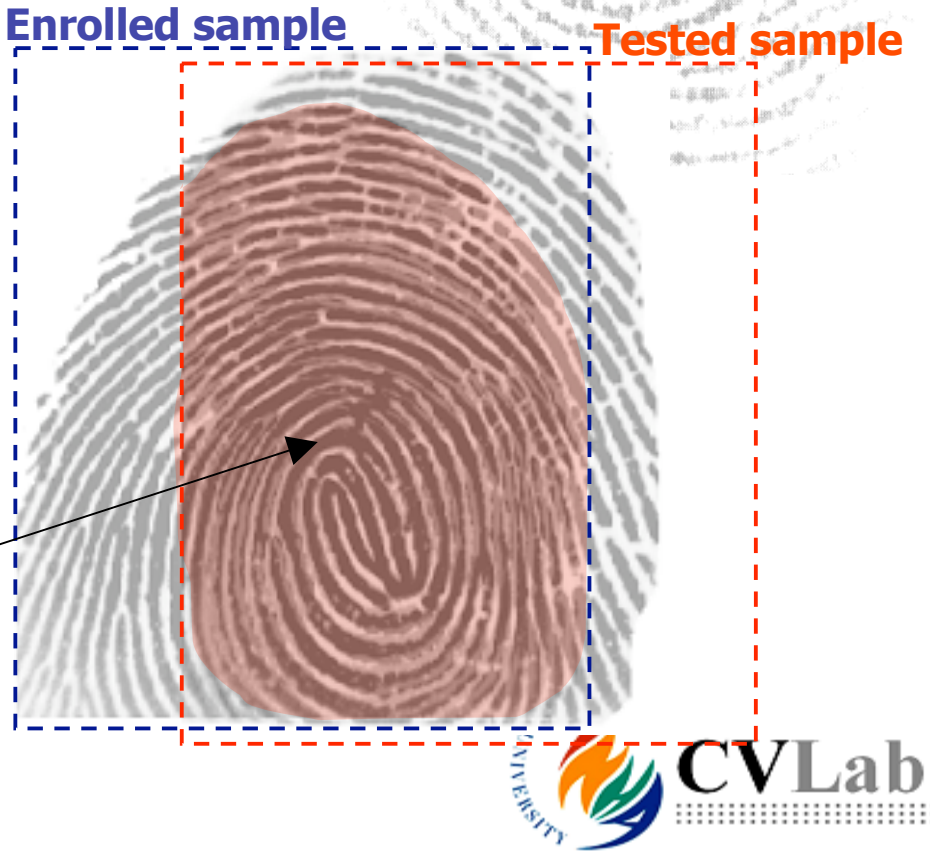
Definitions

- Examples for Component of LoD
  - Fingerprint Database
    - Distribution of sample qu ality
    - Co-occurrence of sample quality
    - Ratio of overlapped area
      - Translational difference
      - Rotational difference
    - etc.
  - Face Database
    - Pose
    - Illumination
    - Facial expression
    - etc.
  - Iris Database
    - Occlusion
    - Illumination
    - Focusing
    - etc.

Components of LoD for Fingerprint

- Ratio of Overlapped area (R_o)
  - Step 1: Shift and rotate tested sample to find the same region of both enrolled sample and tested sample, which is overlapped are a
  - Step 2: Calculate the ratio of the overlapped area for target sam ple pair

$$R_o = \frac{P_{overlapped}}{P_{enrolled}} \times 100(\%)$$



Components of LoD for Fingerprint

- Image quality of sample pairs
  - Ratio of Poor Pairing (RPP)

$$RPP = \frac{P_{Total} - P_{Good}}{P_{Total}} \times 100(\%)$$

  - 3 Levels of Poor Pairing

Quality co-occurrence matrix of genuine sample pairs

Enrolled \ Test	Excellent	Very Good	Good	Fair	Poor
Excellent	1171	331	134	16	7
Very Good	472	305	120	8	1
Good	123	47	48	6	6
Fair	5	0	0	0	0
Poor	0	0	0	0	0

Level 3

CVLab

Components of LoD for Face

- Pose difference (P)
  - Feature information can be changed according to positions

Pose variation in the FERET database.

CVLab

Components of LoD for Face

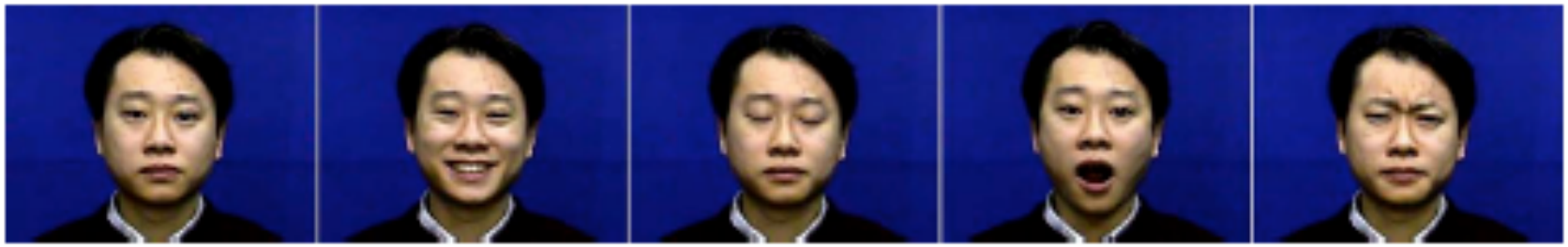
- Illumination difference (I)
  - Change of feature information according to the brightness, color and location of the light source.

Illumination variation in the CAS-PEAL database.

CVLab

Components of LoD for Face

- Expression difference (E)
  - Change of feature information according to expression.



(a) Normal (b) Happiness (c) Blink (d) Surprise (e) Anger

Expression variation in the KFDB database.

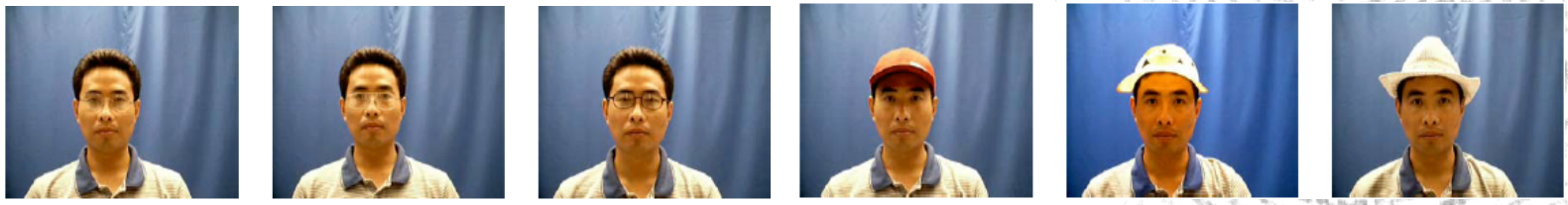
(a) Neutral (b) Smile (c) Anger (d) Scream

Expression variation in the AR database.

CVLab

Components of LoD for Face

- Other components
  - Accessory Variation



Accessory variation in the CAS-PEAL database.

  - Resolution

96 x 128 69 x 93 48 x 64 24 x 32

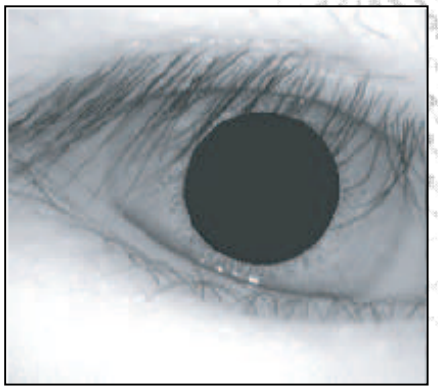
Resolution variation image.

  - Compression
  - etc.

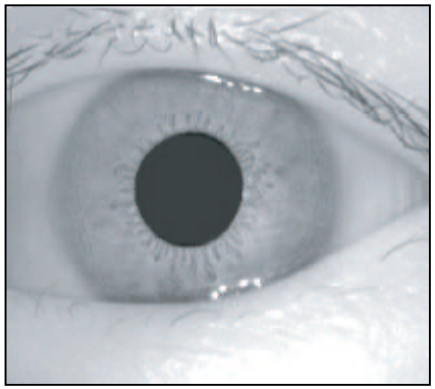
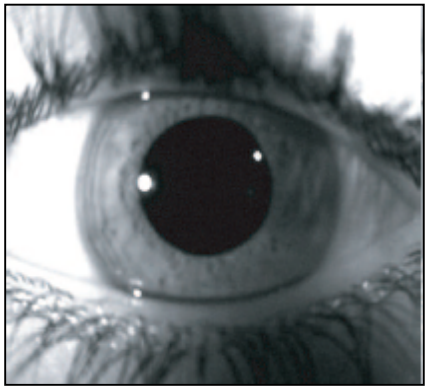
CVLab

Components of LoD for Iris

- Occlusion
  - Pupil loss by eyelash interference
  - Blinking

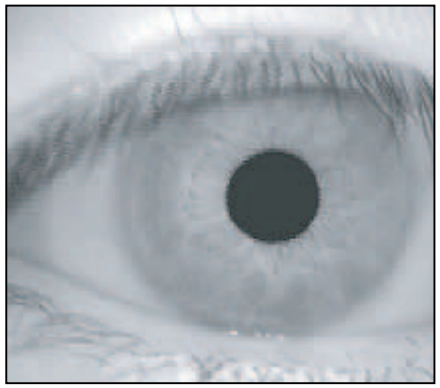



- Non-uniform illumination

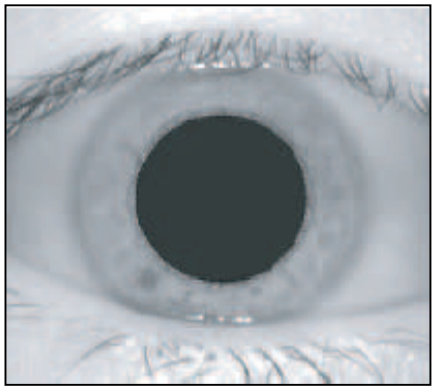
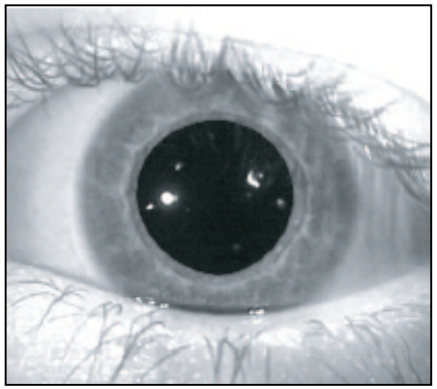



Components of LoD for Iris

- Poor focusing
  - Eye motion and motion blur




- Change in pupil area

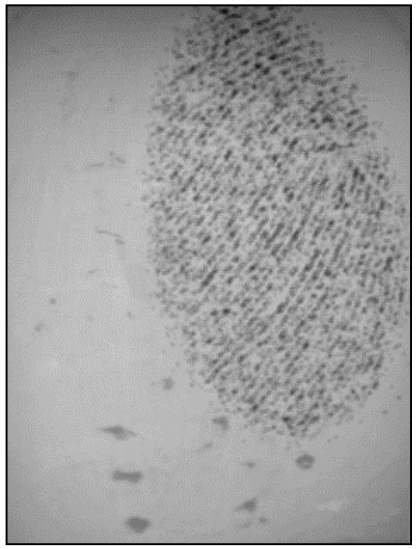



CVLab

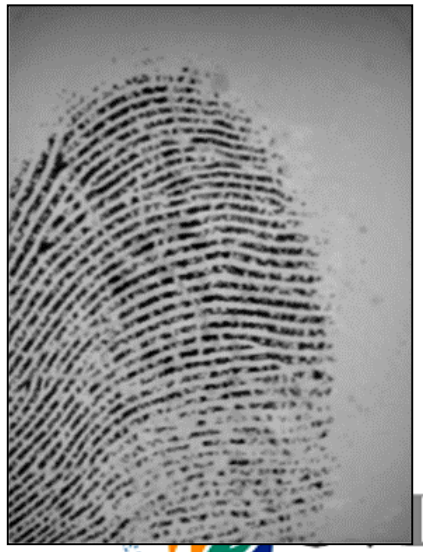
Experiments for LoD of Fingerprint Databases

- Information of Experiments
  - Target databases
    - Three DB4s for FVC2000, 2002, 2004
  - Number of subjects
    - 100 per DB4
  - Number of impressions
    - 8 images per subject
  - Resolution
    - About 500 dpi

FVC2000 DB4



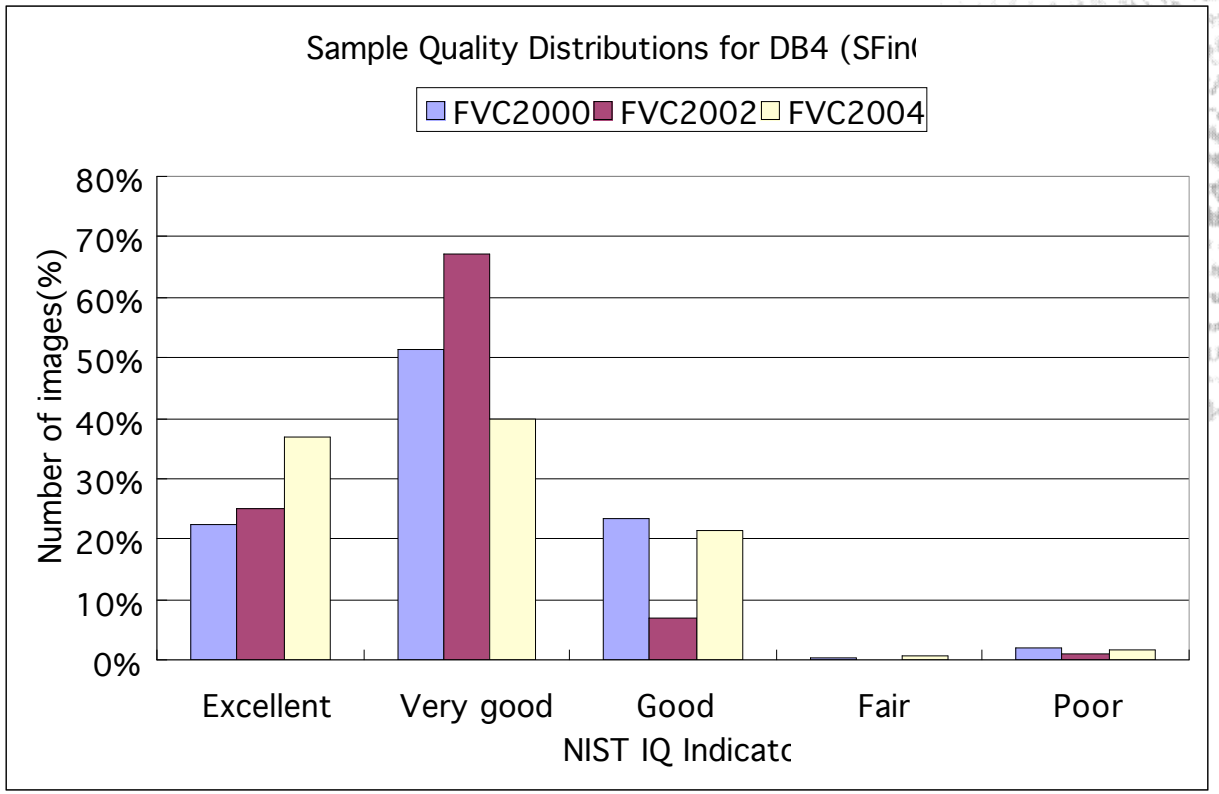
FVC2002 DB4



FVC2004 DB4

Experiments for LoD of Fingerprint Databases

❖ Distribution of Sample Quality

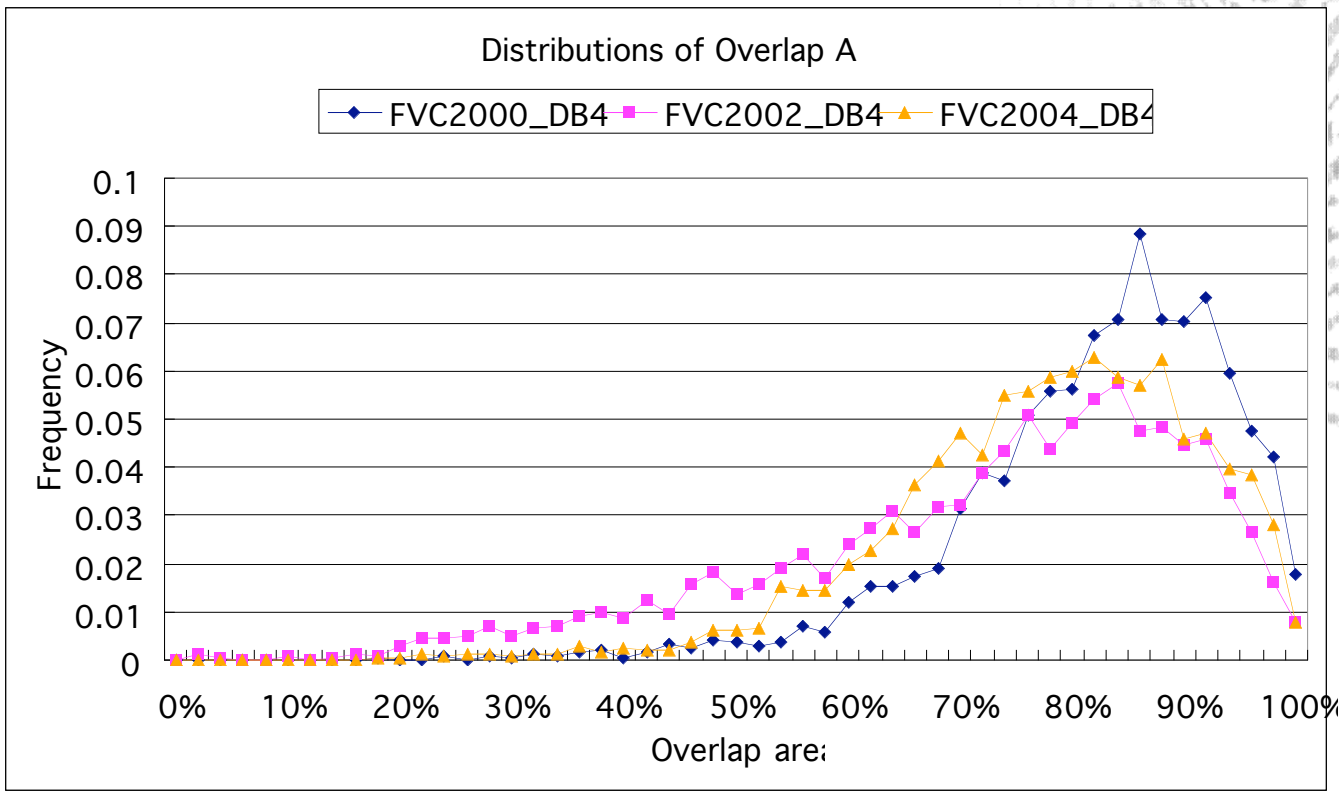


NIST_IQ_Indicator	NIST_IQ_Level	FVC2000	FVC2002	FVC2004
Excellent	1	180	200	294
Very good	2	412	537	320
Good	3	188	55	170
Fair	4	3	1	4
Poor	5	17	7	12
	Number of Images	800	800	800
		2.5%	1.0%	2.0%

CVLab

Experiments for LoD of Fingerprint Databases

- Distributions of Ratio of Overlapped Area

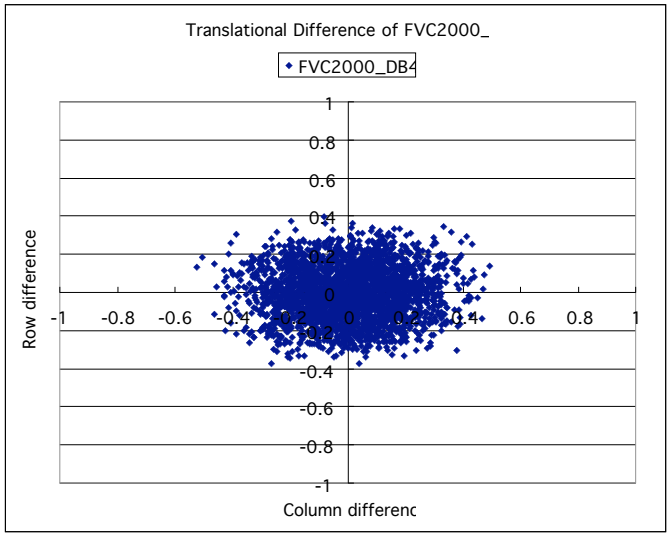


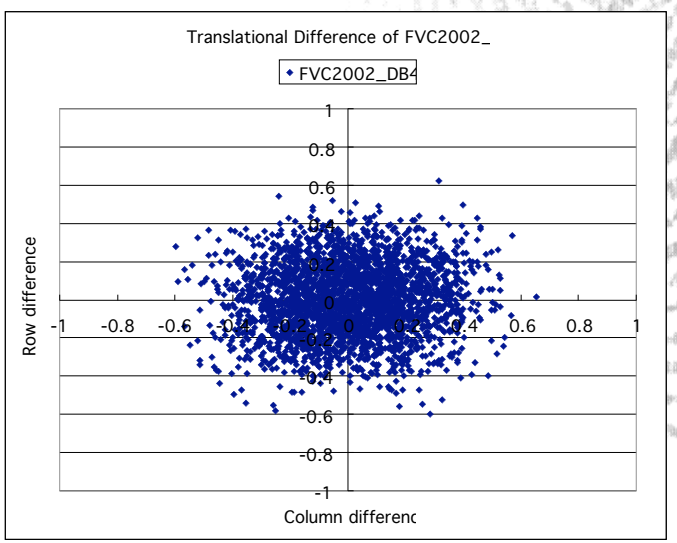
2002 > 2004 > 2000

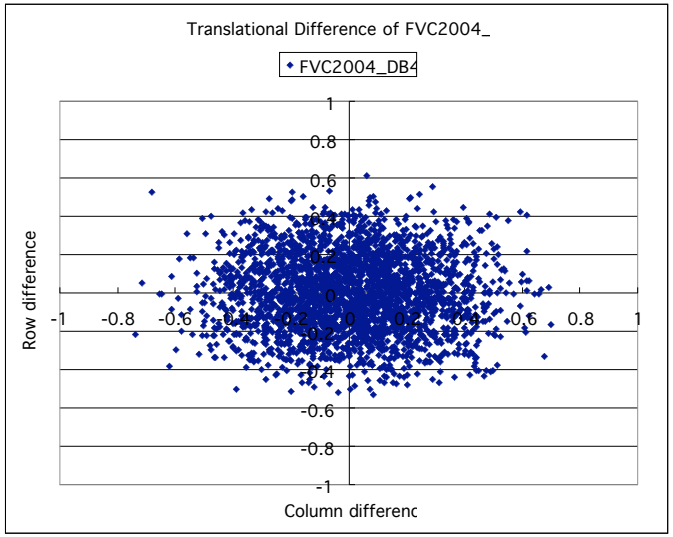
CVLab

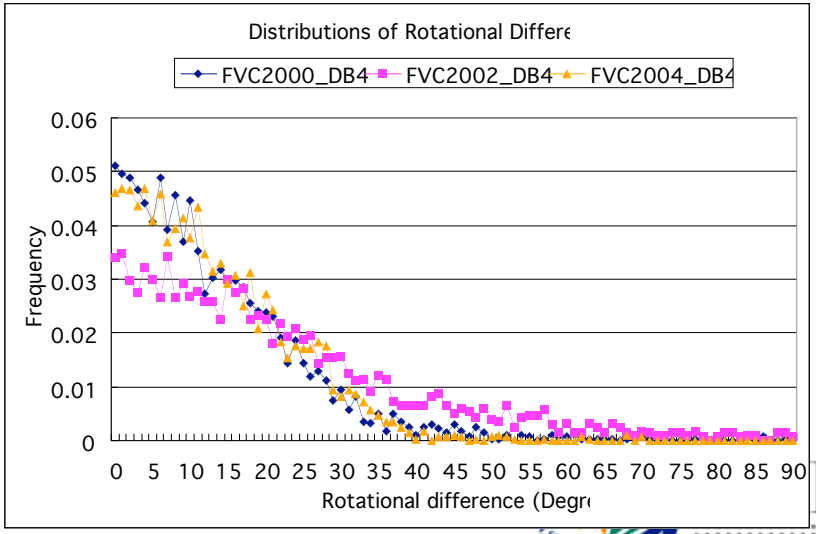
Experiments for LoD of Fingerprint Databases

❖ Distributions of Translational and Rotational Difference









Experiments for LoD of Fingerprint Databases

- Level-2 RPP

Enrolled \ Test	Excellent	Very Good	Good	Fair	Poor
Excellent	322	215	61	0	6
Very Good	266	947	250	0	3
Good	63	250	304	5	40
Fair	0	0	11	1	3
Poor	5	6	28	0	14

FVC2000 DB4

4.4%

Enrolled \ Test	Excellent	Very Good	Good	Fair	Poor
Excellent	380	275	12	0	8
Very Good	313	1482	98	0	5
Good	25	99	65	4	5
Fair	0	0	3	0	0
Poor	7	5	9	0	5

FVC2002 DB4

1.8%

Enrolled \ Test	Excellent	Very Good	Good	Fair	Poor
Excellent	690	246	91	4	3
Very Good	216	605	228	7	13
Good	112	294	209	8	18
Fair	0	6	1	0	0
Poor	6	20	20	2	1

FVC2004 DB4

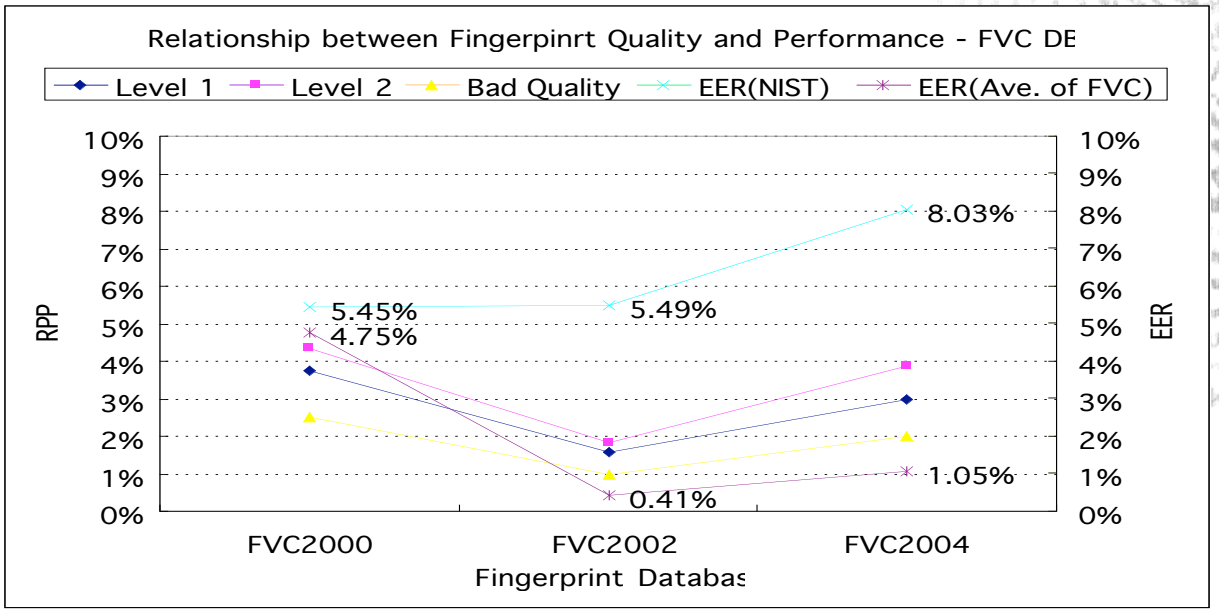
3.9%

2000 > 2004 > 2002

CVLab

Experiments for LoD of Fingerprint Databases

❖ Ratio of Poor Pairing vs. EER



DB4	Level 1	Level 2	EER	Ave. EER	Bad Quality
FVC2000	3.75%	4.36%	5.45%	4.75%	2.50%
FVC2002	1.57%	1.82%	5.49%	0.41%	1.00%
FVC2004	2.96%	3.89%	8.03%	1.05%	2.00%

CVLab

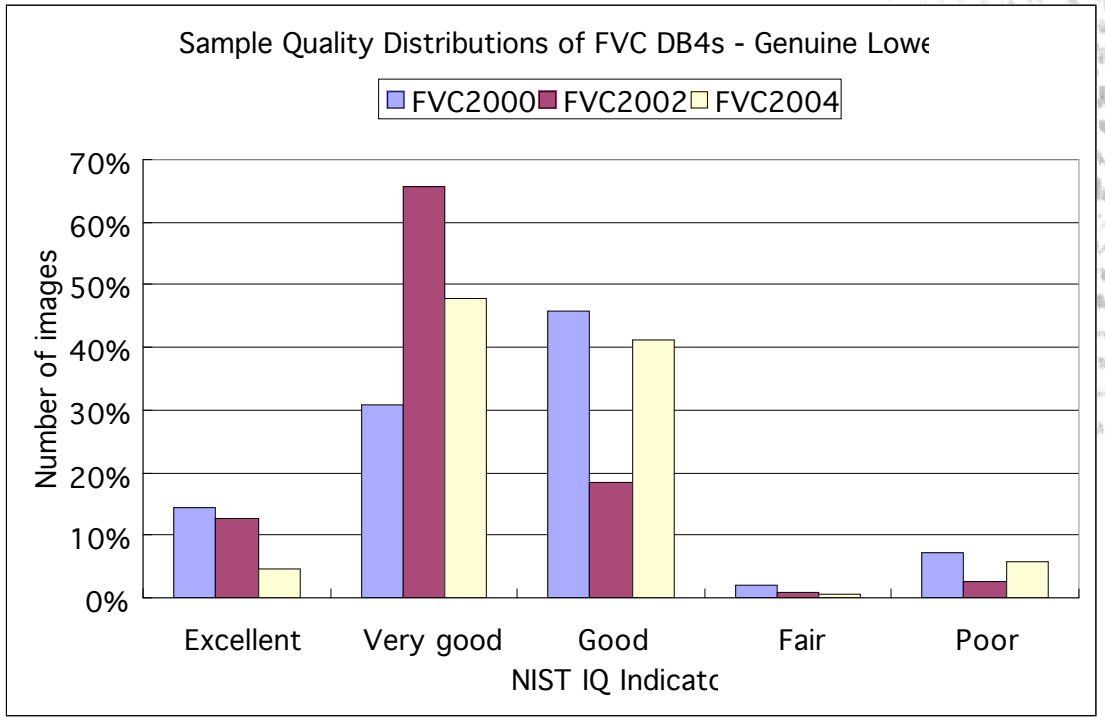
Analysis of Sample Pairs of Low Scores

- Intention
  - To realize how much sample quality affects to the performance
  - To find any other factors causing low genuine matching scores
- Experiment
  - Target DB: DB4(SFinGe) of FVC2000, 2002, 2004
  - Sample Quality Measure: NFIQ
  - Fingerprint Matcher: NIST's BOZORTH3
- Analysis
  - For the samples of bottom 5% scores in genuine matching, colle ct and analyze
    - Sample quality by NFIQ
    - Quality co-occurrence matrix of genuine pairs
    - Ratio of overlapped area of genuine pairs

CVLab

Analysis of Sample Pairs of Low Scores

❖ Distribution of Sample Quality by NFIQ



NIST_IQ_Indicator	NIST_IQ_Level	FVC2000	FVC2002	FVC2004
Excellent	1	20	15	7
Very good	2	43	78	73
Good	3	64	22	63
Fair	4	3	1	1
Poor	5	10	3	9
	Number of images	140	119	153
		9.3%	3.3%	6.5%

CVLab

Analysis of Sample Pairs of Low Scores

❖ Quality Co-occurrence matrix of Genuine sample pairs

Enrolled \ Test	Excellent	Very Good	Good	Fair	Poor
Excellent	5	5	4	0	2
Very Good	2	18	18	0	0
Good	2	8	45	2	13
Fair	0	0	4	0	2
Poor	0	2	7	0	1

FVC2000 DB4

76.4%

Enrolled \ Test	Excellent	Very Good	Good	Fair	Poor
Excellent	5	6	0	0	1
Very Good	4	66	13	0	0
Good	0	12	22	3	1
Fair	0	0	2	0	0
Poor	2	1	2	0	0

FVC2002 DB4

91.4%

Enrolled \ Test	Excellent	Very Good	Good	Fair	Poor
Excellent	0	1	1	0	0
Very Good	1	32	9	0	1
Good	5	28	47	1	7
Fair	0	0	0	0	0
Poor	0	3	3	0	0

FVC2004 DB4

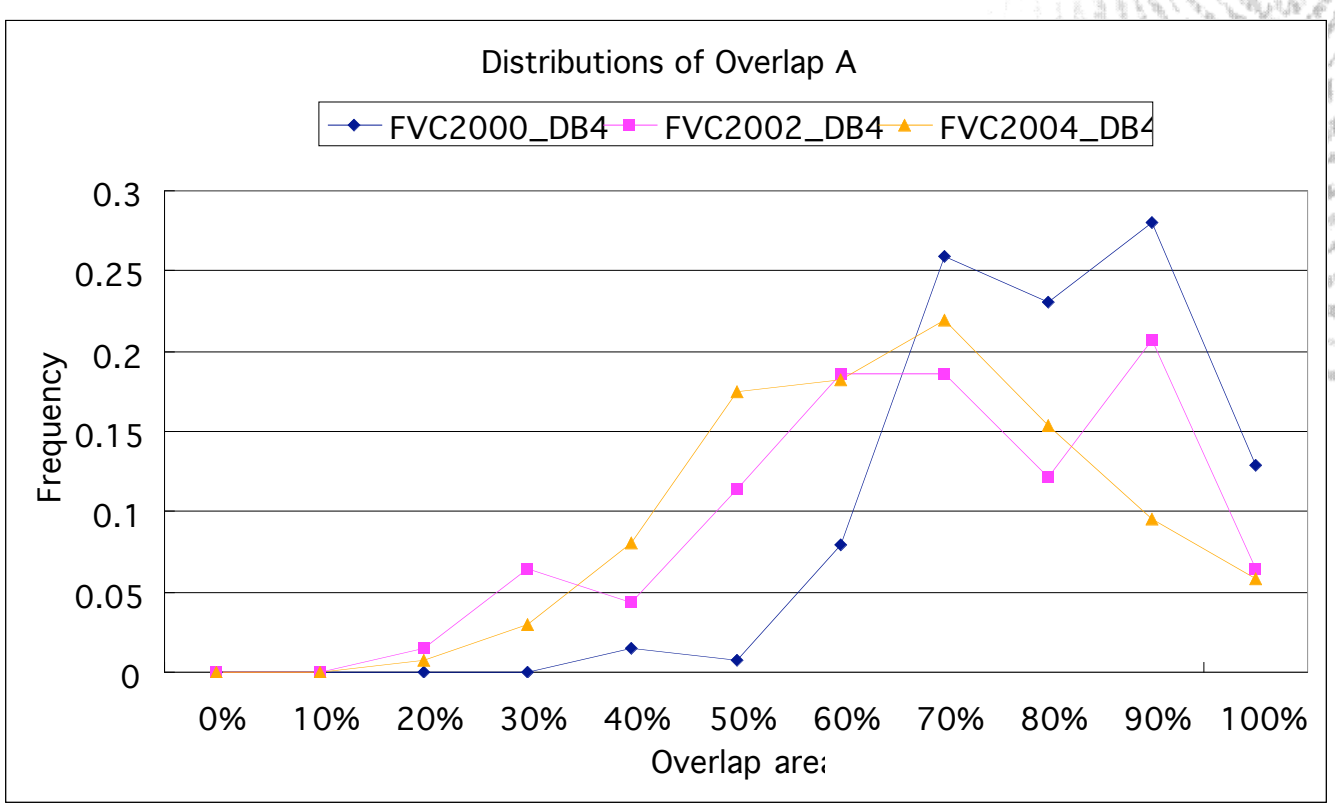
88.6%

✓In thses specific databases, there are a good portion of genuine pairs with "**good quality pairing**" but "**low matching scores**."

CVLab

Analysis of Sample Pairs of Low Scores

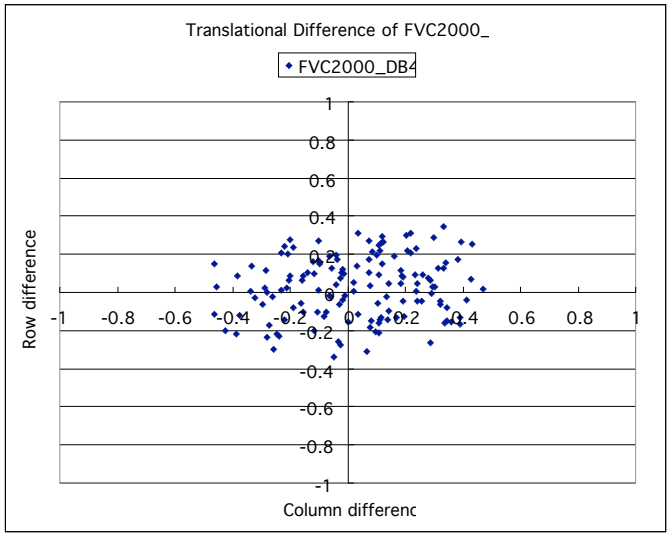
- Distribution of Ratio of overlapped area between Genuine pairs

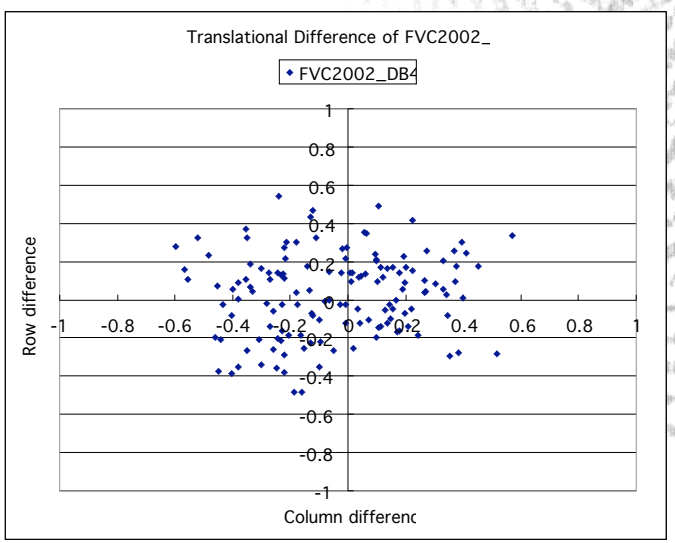


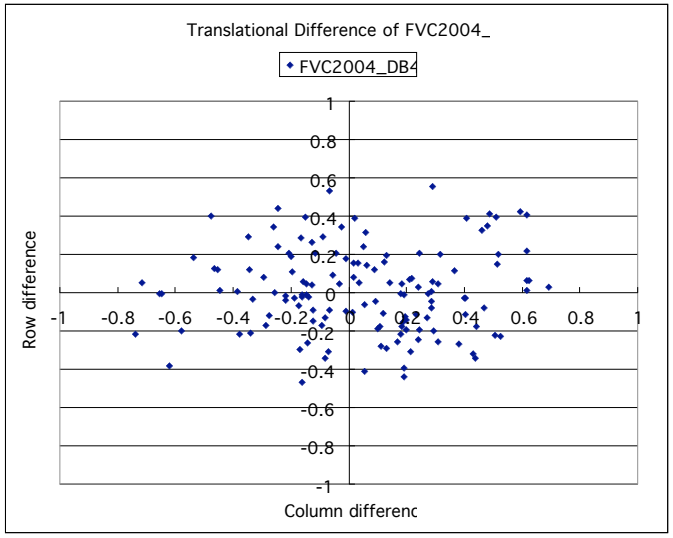
CVLab

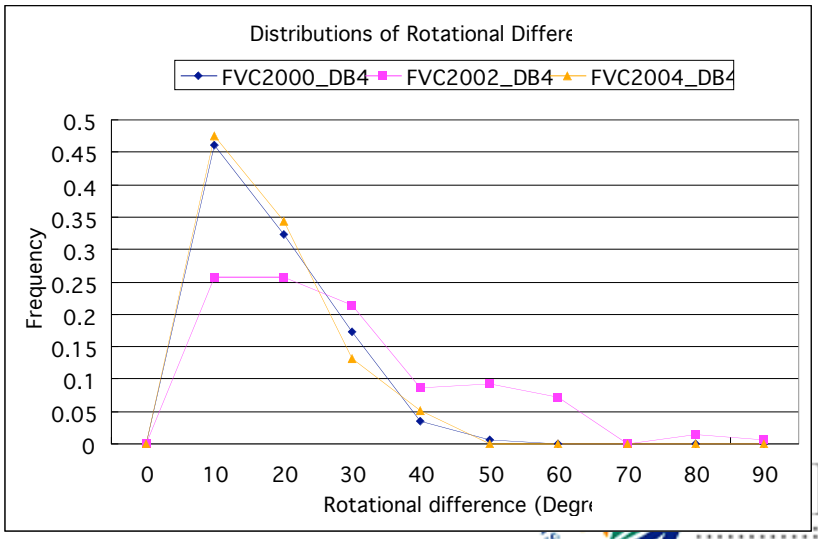
Analysis of Sample Pairs of Low Scores

- Distributions of Translational and Rotational Differences









Analysis of Sample Pairs of Low Scores

- Quality-pairing of Low overlapped genuine pairs
  - For FVC2000 DB4

Overlap < 40% 2

Enrolled \ Test	Excellent	Very Good	Good	Fair	Poor
Excellent	0	0	0	0	0
Very Good	0	0	1	0	0
Good	0	1	0	0	0
Fair	0	0	0	0	0
Poor	0	0	0	0	0

Overlap < 50% 3

Enrolled \ Test	Excellent	Very Good	Good	Fair	Poor
Excellent	0	1	0	0	0
Very Good	0	0	1	0	0
Good	0	1	0	0	0
Fair	0	0	0	0	0
Poor	0	0	0	0	0

CVLab

Analysis of Sample Pairs of Low Scores

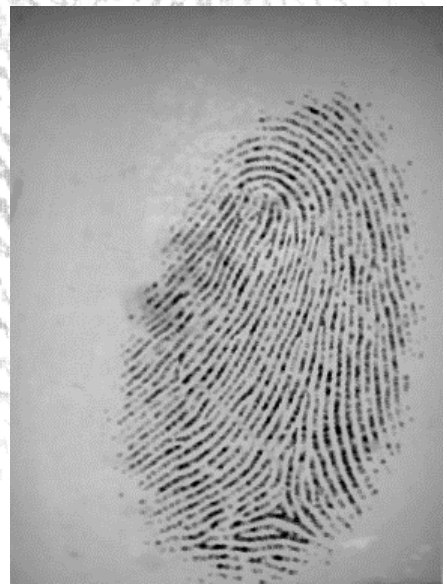
- Quality-pairing of Low overlapped genuine pairs
  - For FVC2002 DB4

Overlap < 40% 17

Enrolled \ Test	Excellent	Very Good	Good	Fair	Poor
Excellent	4	0	0	0	0
Very Good	2	9	0	0	0
Good	0	0	1	0	0
Fair	0	0	0	0	0
Poor	1	0	0	0	0

Overlap < 50% 33

Enrolled \ Test	Excellent	Very Good	Good	Fair	Poor
Excellent	4	0	0	0	0
Very Good	2	22	1	0	0
Good	0	0	2	0	0
Fair	0	0	0	0	0
Poor	2	0	0	0	0



Analysis of Sample Pairs of Low Scores

- Quality-pairing of Low overlapped genuine pairs
  - For FVC2004 DB4

Overlap < 40% 16

Enrolled \ Test	Excellent	Very Good	Good	Fair	Poor
Excellent	0	1	0	0	0
Very Good	0	5	0	0	0
Good	0	2	0	0	0
Fair	0	0	8	0	0
Poor	0	0	0	0	0

Overlap < 50% 40

Enrolled \ Test	Excellent	Very Good	Good	Fair	Poor
Excellent	0	1	1	0	0
Very Good	0	12	3	0	0
Good	1	6	14	0	1
Fair	0	0	0	0	0
Poor	0	1	0	0	0

✓ **Good Q-pair but low overlap → Low matching score**

CVLab

Analysis of Sample Pairs of Low Scores

- Observations
  - Sample quality seems the most influencing factor to the perform ance with the underlying assumptions that enrolled samples are of good quality and have enough overlapped area with test sam ples.
  - However, the above assumptions do not hold in technology (off-l ine) evaluation where sample quality control is not in use.
  - Even a genuine pair of excellent quality do not match.
  - Ratio of overlapped area can be a factor to be considered for pre dicting the performance, especially in technology evaluation.

Conclusions

- Defines the quality of biometric databases, called *Level of Difficulty*.
- Proposes possible components of LoD for fingerprint.
- Demonstrates the automatic processes of measuring the components.

- How to combine the multiple components into a single LoD?
- How to predict the relative performance of a recognition algorithm based on LoD's?
- How to develop automatic processes of measuring LoD of face and iris databases?

CVLab

Thank you for attention !

Hakil Kim
hikim@inha.ac.kr

CVLab