Empowered by Innovation NEC

Thoughts on Automatic Latent Processing and Matching Algorithms

NIST Latent Testing Workshop
April 5-6, 2006

Masanori Hara

Preliminary Questions

- **Is Automatic Latent Encoding feasible?**
- **Is Automatic Latent Verification feasible?**
- **How much Accuracy Degradation is acceptable?**
- **How much Cost Increase is acceptable?**

Contents

1. Operational Needs

1.1 Examiner Driven & Manual Intensive

- Expert Examiner's skill - Far better than the present technology

- Operational Accuracy: Mostly dependent on Examiner's skill
  (*) Especially for low quality latent prints

- Expert Examiners tend to put energy to find serious criminals
  (*) Verify 30 through 100 candidates for serious cases

- LE (Law Enforcement) customers prefer "AFIS which maximizes accuracy with Examiner's full assistance" rather than "AFIS which is limited to perform average accuracy and which Examiner's skill is NOT fully utilized"

- Features needs to be compatible to Examiner's definition
  (*) Minutia, Ridge Count, Core&Delta, etc.

1.2 Fully Automatic (Light Out)

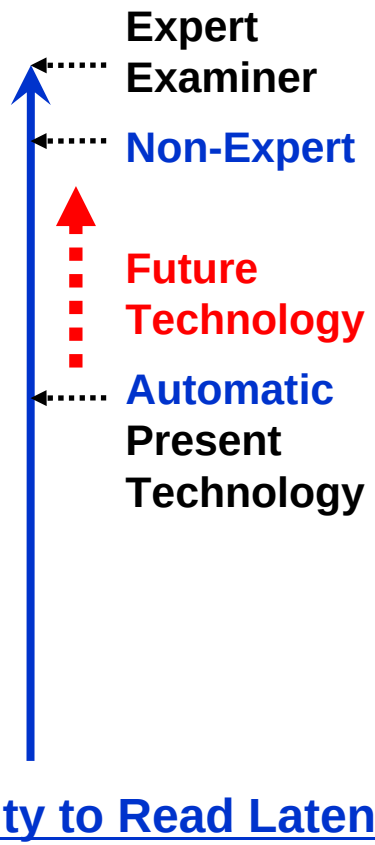
- Targeted to New Demands (such as Homeland Security) rather than traditional Criminal Investigation
- Quicker response expected
- Hiring sufficient Expert Examiners very Unlikely
- Significant Accuracy Degradation estimated
- Automatic Identification very Unlikely for most latent prints
- Features NOT have to be compatible to Examiner's definition

1.3 Semi-Automatic

- Targeted to New Demands that put up with Accuracy Degradation to some extent
- Expert Examiners workload be reduced without ”significant” Accuracy Degradation
- Worth to use Non-Experts on easier fields

(*) Non-Expert’s skill – Still far better than the present technology

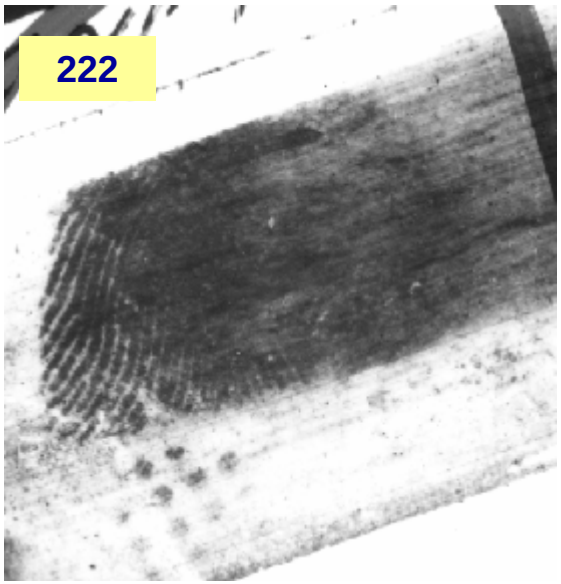
- Final Identification must be conducted by Expert Examiners
- Features NOT have to be compatible to Examiner’s definition

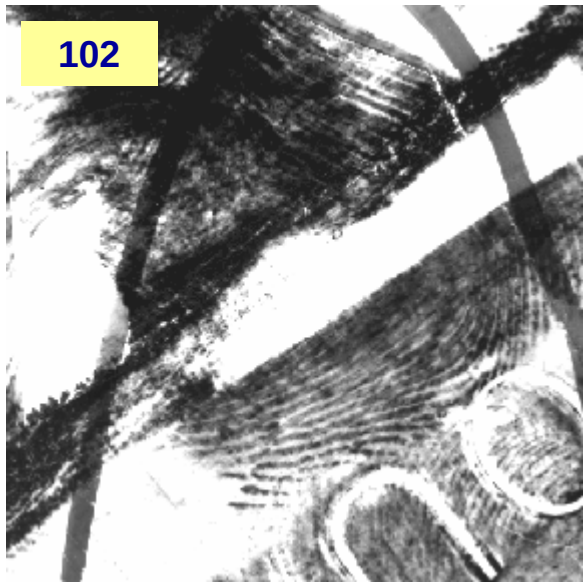


2. Latent Encoding

2.1 Manual Encoding (or Edit)

1) Time consuming and cumbersome
2) Requires Expert skill, and difficult for Non-Experts
3) Cost: Dependent on quality and the target area size
4) Features: Orientation, Target Area,
   Minutia, Ridge Count, Core/Delta, etc. or
   Zone, Skeleton, UCX; See Chap 4 for detail





2. Latent Encoding

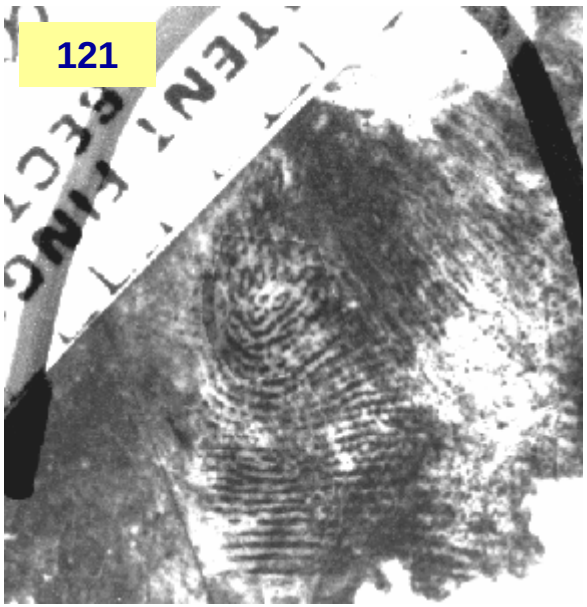
2.2 Fully Automatic Encoding

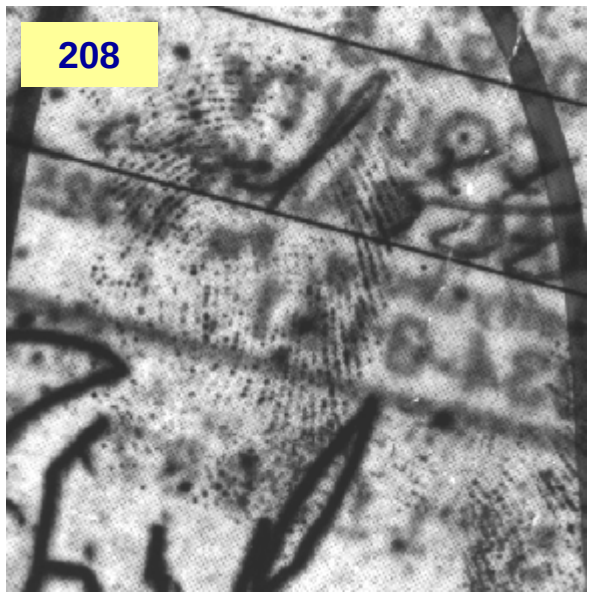
Dependent on the image quality

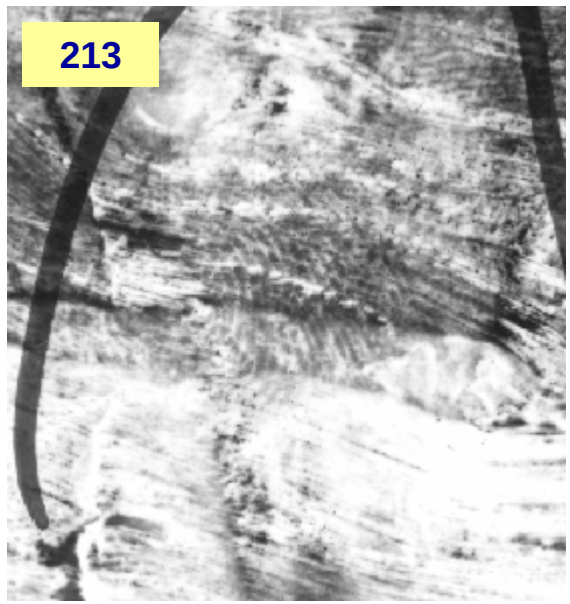
- **OK for very good quality latent prints**
- **Probably OK for good quality latent prints & Mate in Data Base**

However, difficult to predict at latent input,

i.e. without assessing matching results







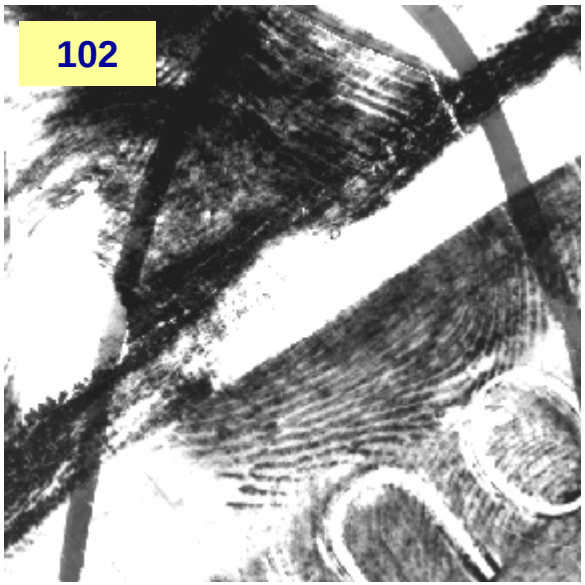
2. Latent Encoding

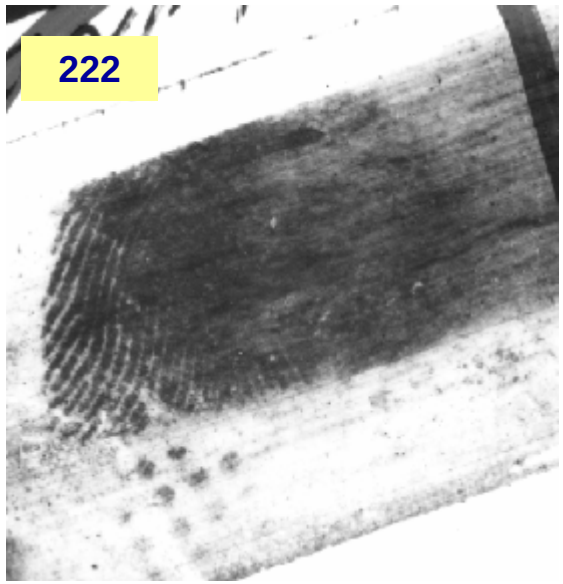
2.3 Semi-Automatic (Combined)

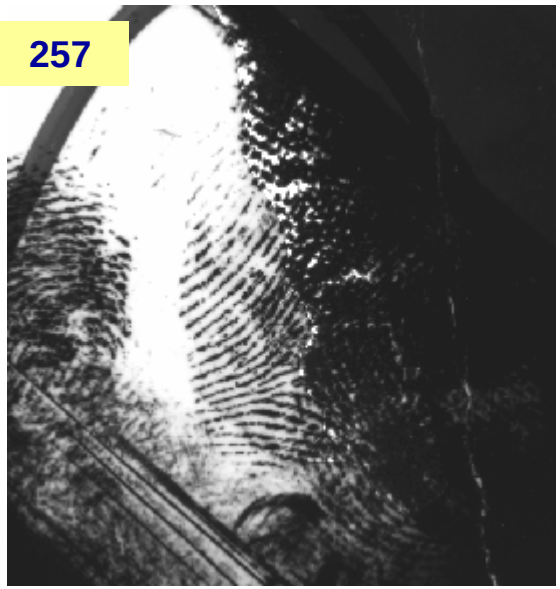
1) Orientation: Manual Recommended

- **Automatic Detection → Far less accurate than Manual**
- **Manual Decision → Assisted by supplemental information**
- **Easy Manual Input → Manual cost almost negligible**

(*) 360-degree search possible but NOT usually recommended
→ Unnecessary increase Matching Cost (Algorithm Cost)
→ Unnecessary degrade Accuracy (i.e. More False Candidates)







2. Latent Encoding

2) Target (Search) Area Finding: Manual Recommended

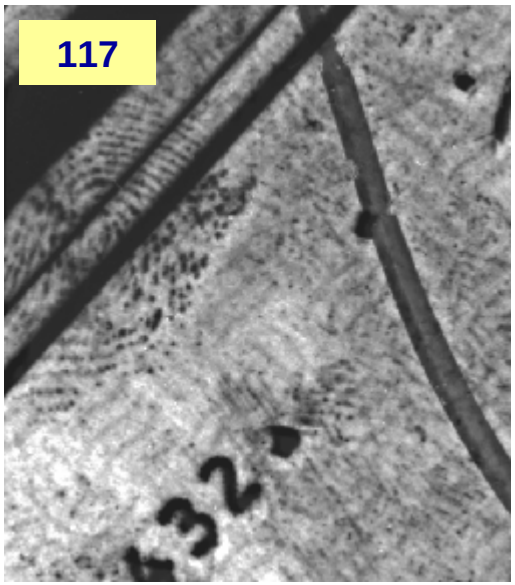
- Automatic Detection → Far less accurate than Manual

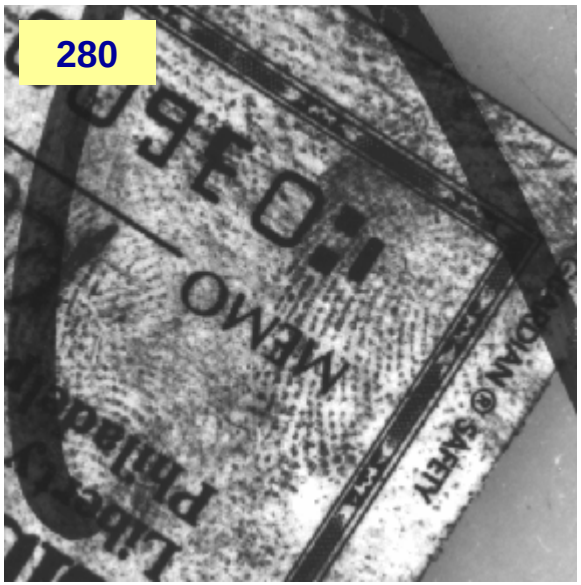
- Easy Manual Setting → Manual cost NOT significant

(*) Multiple-area search technically possible but NOT recommended

→ Significantly increase Matching Cost (Algorithm Cost)

→ Significantly degrade Accuracy (i.e. More False Candidates)





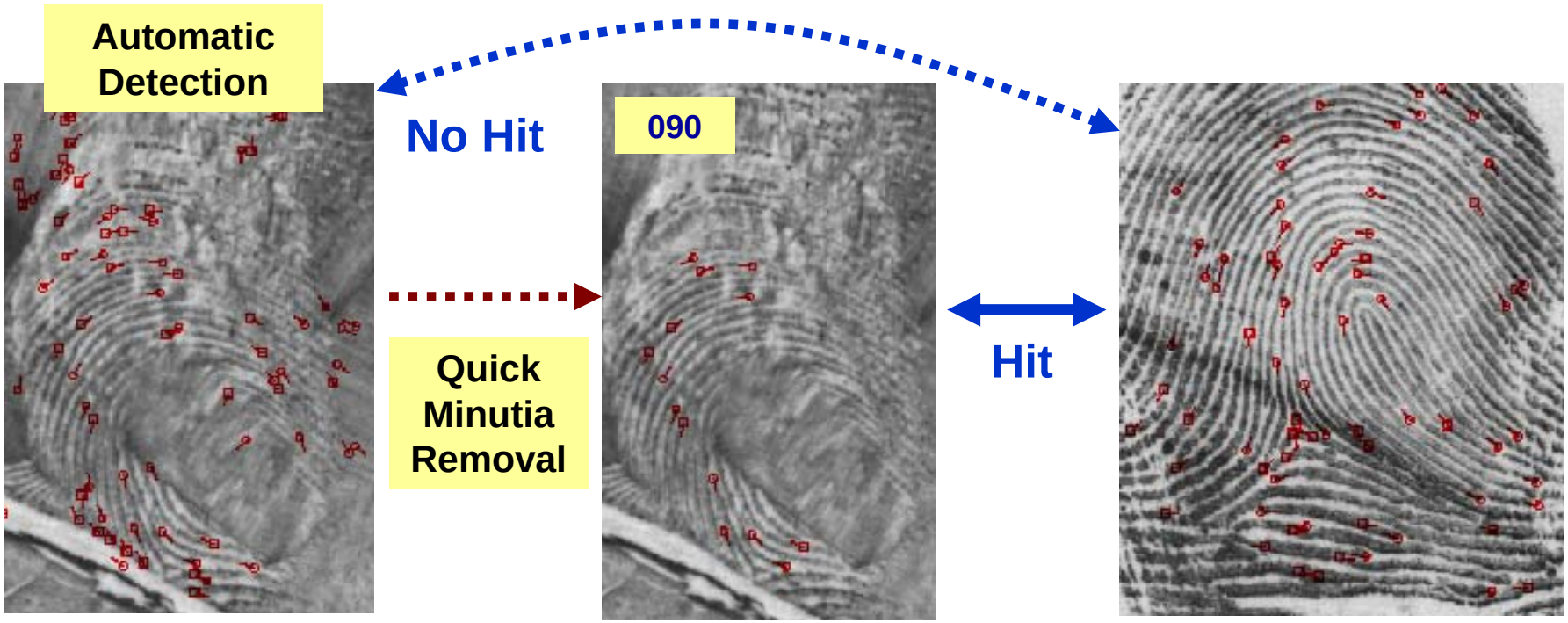
2. Latent Encoding

3) Minutia Plot (Edit): Manual NOT Recommended

- Too costly except for very serious cases
- Limited plot (addition) only recommended for serious cases

Automatic Detection & Manual Removal Recommended

- Aggressive Automatic Detection → Many Unreliable Minutiae
- Easy Manual Removal → Manual cost NOT significant



2. Latent Encoding

4) Ridge Count: Manual Edit NOT Recommended

- Too costly except for very serious cases

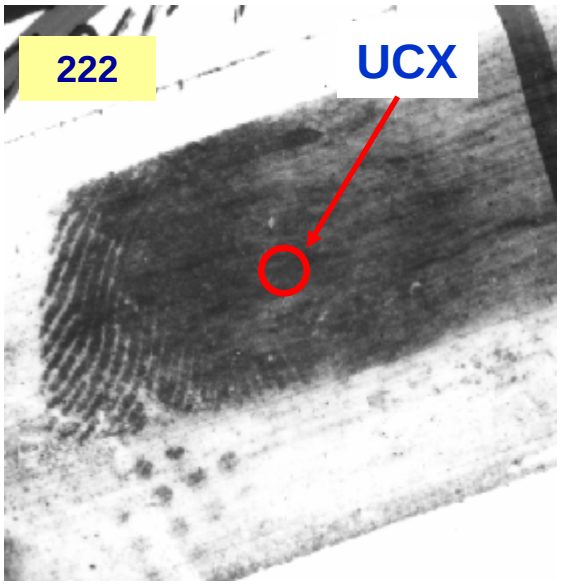
5) Core/Delta: Manual Edit Recommended

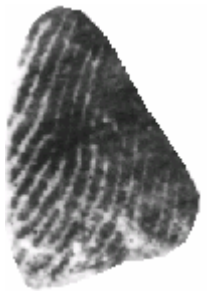
- Manual Edit NOT difficult → Maybe worth for manual cost

6) UCX: Manual Edit Recommended

Note: See Chap 4 for detail of UCX

- Effective to reduce matching cost and to reduce false candidates
- Easy Manual Edit; Easier than Core&Delta





Cropped Area
(Visible Ridge)

2. Latent Encoding

7) Zone: Manual Edit Recommended

- Easy method to remove Unreliable Features (Minutiae or Skeleton)
- Effective to reduce false candidates
- Manual cost NOT significant

8) Skeleton: Manual Edit NOT Recommended

- Too costly except for very serious cases
- Limited Edit only recommended for serious cases
- Skeleton Edit maybe useful at Visual Verification stage

3. Latent (Visual) Verification

3.1 Manual (Visual) Verification

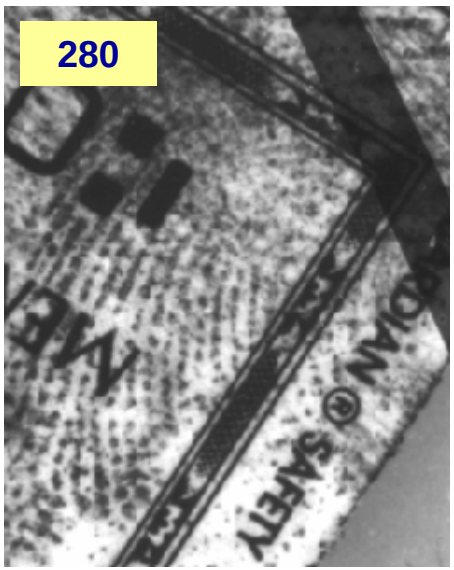
1) Time consuming

2) Requires Expert skill, and difficult for Non-Experts

(*) Difficult cases require long time even for Expert Examiners

3) Cost: Dependent on size of candidates

Cost for intensive Visual Verification (for difficult cases) could exceed cost for Manual Latent Encoding



3. Latent (Visual) Verification

3.2 Fully Automatic Verification

Dependent on Image Quality

- OK for very good quality latent prints & Mate in Data Base
  (*) 1st rank & Big score gap → Confident Hit
- Very difficult (almost impossible) for most latent prints
  (*) 1st rank; Small score gap → Inconclusive
  2nd rank or lower → Almost impossible

According to the statistics of Illinois State Police (ISP),
"Fully Automatic Verification" could lose identifications (hits)
by 50% compared to "Manual Verification of Top 10 candidates" .

Can we afford to lose so many hits?

If not, how much can we afford to lose?

3. Latent (Visual) Verification

Reference: An Evaluation of NEC's Enhanced Matching Algorithms
Mike Murphy, Illinois State Police (ISP)

(*) A part of page 6

15th Annual
AFIS Internet Conference
Biloxi, Mississippi
August 19 - 23, 2001

An Evaluation Of
NEC's Enhanced Matching Algorithms

Mike Murphy
Latent Print Group Supervisor
Illinois State Police

Only candidates falling in the top ten positions of the Candidate List will be recorded as hits. (Latent Candidate Merge hits will not be counted as hits for the purposes of this study.)

Search limiters (Sex, Region, Finger #, Pattern Type, etc.) will be applied to the search as dictated by the case and latent print.

Results

As of the end of July 2001, 156 latent print identifications from AFIS searches have been reported and are included in these results. Three laboratories have reported making AFIS identifications using the Enhanced Matching Algorithms that had not been made when the latent prints were searched using the Standard Matching Algorithm. The identifications reported by these laboratories are recorded in Attachment 2.

The number of identifications produced by each matching algorithm is contained in the following table.

Number of Identifications	Matching Algorithm Used
134	Standard
146	Enhanced A
146	Enhanced B

3. Latent (Visual) Verification

Out of 82 Hits
(Enhanced-A Algorithm)

1st Rank & Big gap: 43 (52%)
1st Rank: 66 (80%)
Up to 3: 76 (93%)
Up to 10: 82 (100%)

Note1: See USP4,956,870 for score gap calculation concept

Note2: Out of 146 hits, only 82 hits have score and rank data.

Data Base Size : 3.3M Subjects

(*) A part of page 7

Latent	Finger	Minutia	Score (Hit in Top 10 Candidates)			Position On Candidate List			Spread Score Gap		
			Standard	Enhanced - A	Enhanced - B	Standard	Enhanced - A	Enhanced - B	Standard	Enhanced - A	Enhanced - B
1	8	25	-	334	-	-	6	-	-	15	-
2	8	17	-	435	601	-	2	3	-	7	17
3	9	30	-	-	1617	-	-	2	-	-	34
4	6	41	-	1631	2109	-	3	3	-	39	56
5	3	14	-	1087	-	-	2	-	-	294	-
6	9	17	401	424	606	6	5	4	6	2	78
7	8	28	7259	7116	9633	1	1	1	4478	?	?
8	2	18	-	1225	1253	-	1	1	-	?	37
9	4	52	2954	3423	2976	1	1	1	2238	2655	1871
10	7	15	-	1692	1691	-	1	1	-	211	211
11	1	8	-	116	-	-	6	-	-	6	-
12	2	26	2563	2601	3087	1	1	1	1448	1669	2657
13	6	54	6790	9999	9377	1	1	1	5428	8810	7486
14	10	32	4021	6031	5647	1	1	1	2552	4591	4019
15	6	27	2866	2263	3087	1	1	1	1329	1023	1180

3. Latent (Visual) Verification

3.3 Semi-Automatic (Combined)

Visual Verification seems to be very effective even if Non-Experts do Primary Verification (screening)

1st Rank & Big Score Gap →

Identification by Expert Examiner (1st Rank Only)

Small Score Gap between 1st and 2nd →

1) Visual Verification (screening) by Non-Experts for up to 3 (?) candidates

2) Identification by Expert Examiner only for probable cases

ex. Zero or one (?) candidate per search

4. Matching Algorithms and Costs

4.1 Reliability of Features and Algorithms

- **Suitable algorithm depends on "reliability" of features**
- **NEC has three algorithm lines**
  - **a) LE: Latent Matching for Fingerprints**
  - **b) PID: Flat/Slap Matching for Positive ID**
  - **c) PALM: Latent Matching for Palm prints**
- **LE and PALM Algorithms: Assuming Reliable Features on latent prints (Reliable Minutiae, Reliable Ridge Count and Proper Zones)**
- **PID Algorithm: Tuned for fully Automatic Features (Less Reliable Minutiae per Examiner's definition)**

  (*) Features needs NOT to be compatible to Examiner's definition

4. Matching Algorithms and Costs

4.2 NEC's LE Algorithms

Algorithm Name	Release (First User)	Proc. Cost	Accuracy (%improve)	Template Size (bytes)
1) Standard	1981 (NPA)	1.0	Baseline	800
2) Enhanced-A	**Aug. 1999 (ISP)**	**1.3**	**+10%**	**800**
3) Zone Matching	**Jan. 2006 (NPA)**	**1.5**	**+12%**	**800**
4) Ridge Flow Mat.	**Jan. 2006 (NPA)**	**2.0**	**+15%**	**2500**
5) Flexible Match	Aug. 2005 (S/W)	5.0	+20%	2500/800
6) Skeleton Match	**2010?**	**200?**	**+35%?**	**5000?**

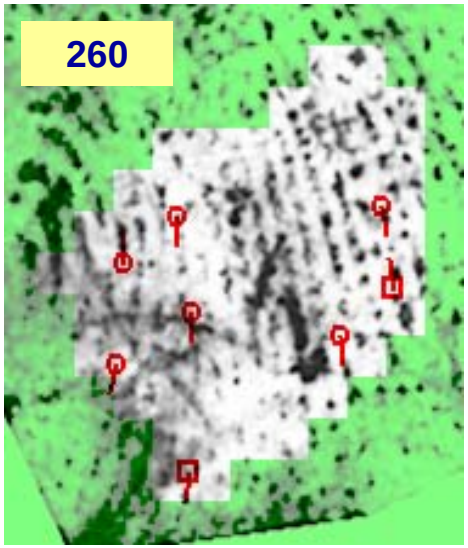
Notes: a) Figures with Suffix "?" : Estimated Figures
b) "%improve": Relative Improvement Percent
(*) Rate of Accuracy Improvement against the Room of Improvement
ex: Accuracy Gain from 60% to 70% → "%improve" is 25% (10% out of 40%)
Legend: NPA - National Police Agency of Japan
ISP - Illinois State Police; S/W – Software Library Available

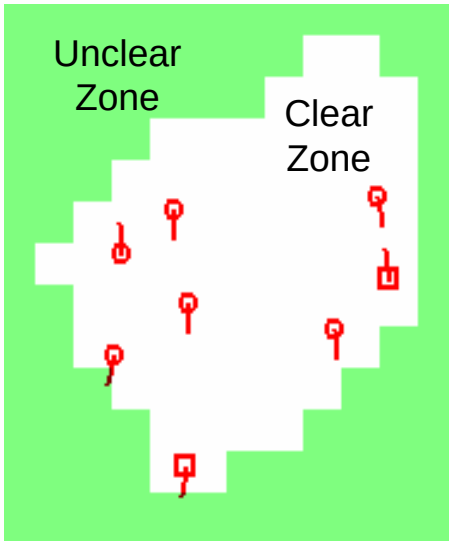
4. Matching Algorithms and Costs

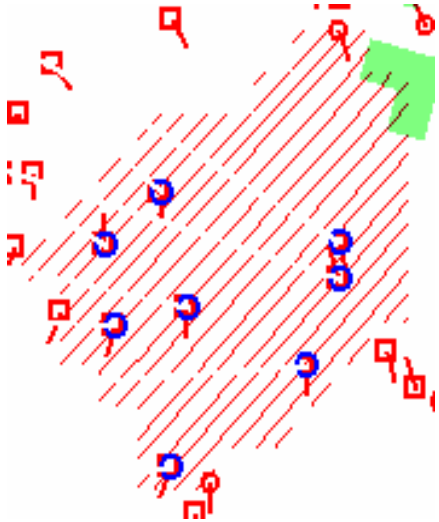
3) Zone Matching

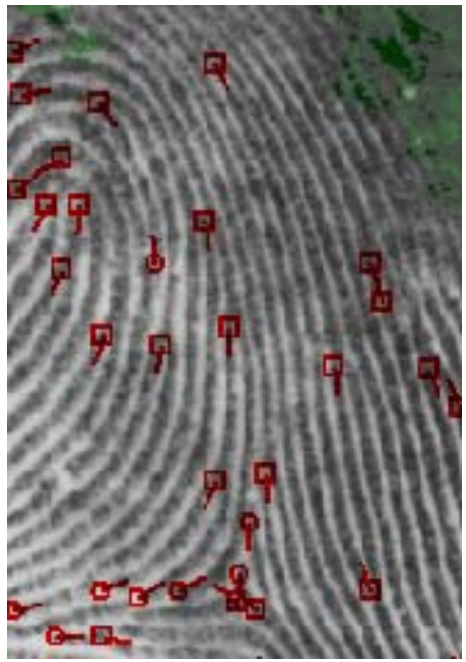
Zone: 16x16 pixel block which represents “Clearness” of ridges

- Score enhancement based on the area size of Common Zone
- Score penalized if Unpaired Minutiae exist on Common Zone area
- Zone can be compared even where minutia does not exist







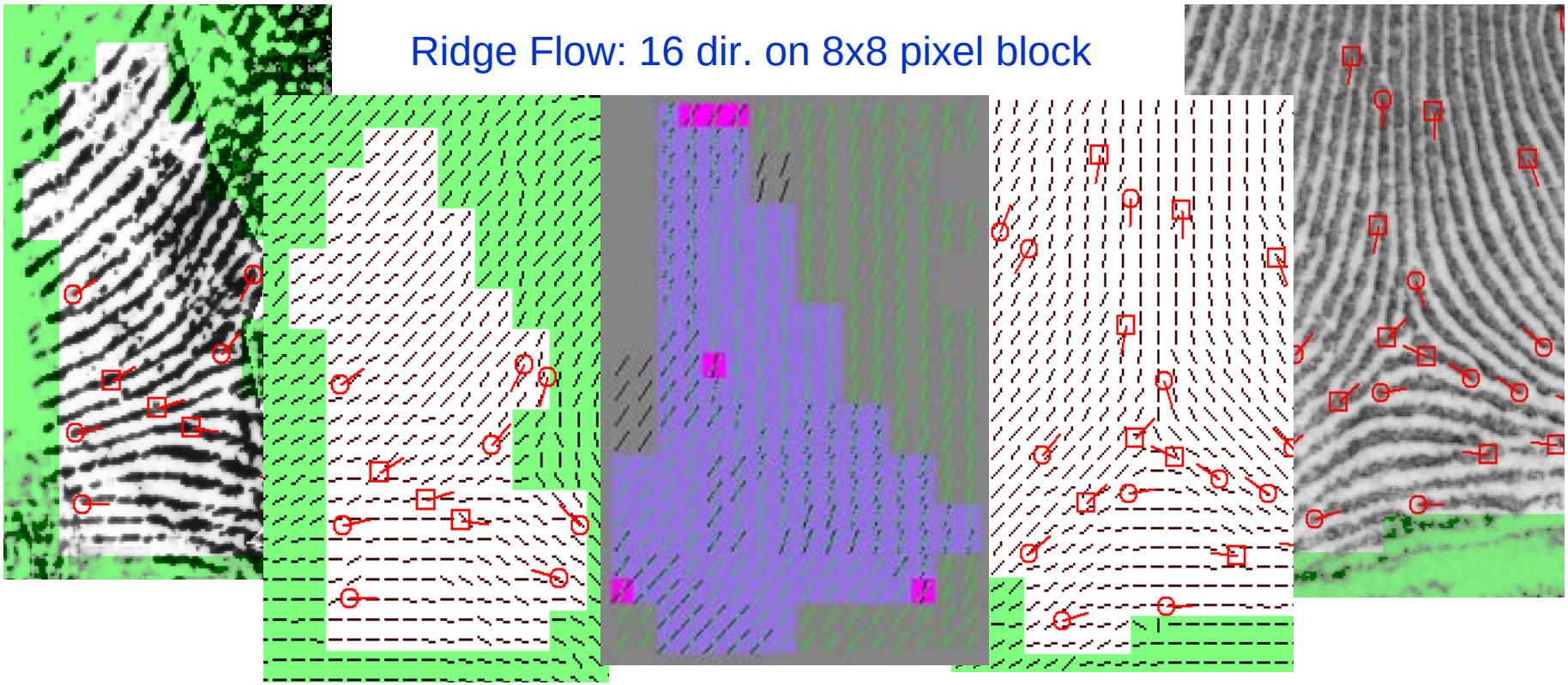


4. Matching Algorithms and Costs

4) Ridge Flow Matching

- Score enhancement based on the area size of matching Ridge Flow
- Score penalized if Ridge Flow does not match
- Ridge Flow is automatically extracted from skeleton for latent print
- Ridge Flow can be compared even where minutia does not exist

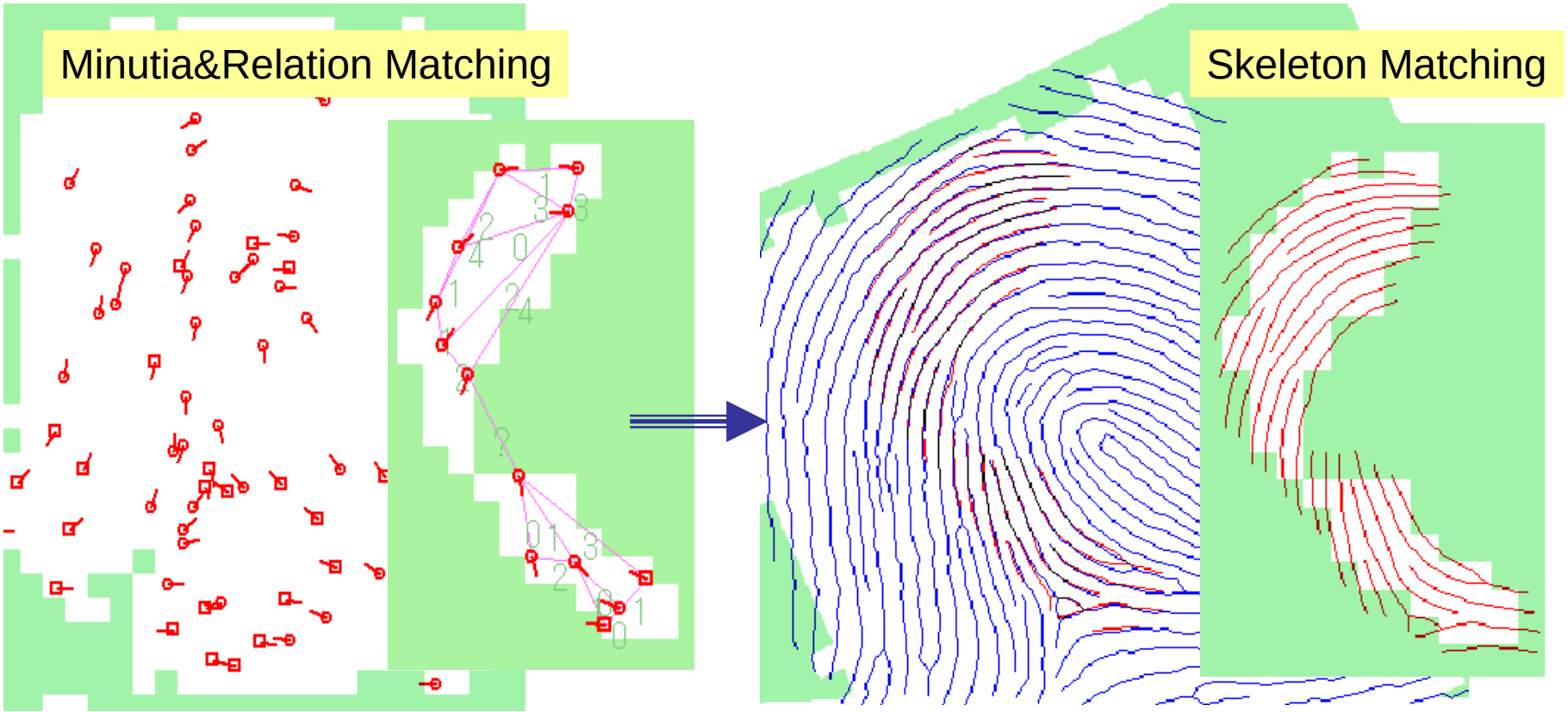
Ridge Flow: 16 dir. on 8x8 pixel block



4. Matching Algorithms and Costs

6) Skeleton Matching

- Score enhancement based on the length of matching Skeleton lines
- Score penalized for inconsistent structures of Skeleton
- Skeleton can be compared even where minutia does not exist

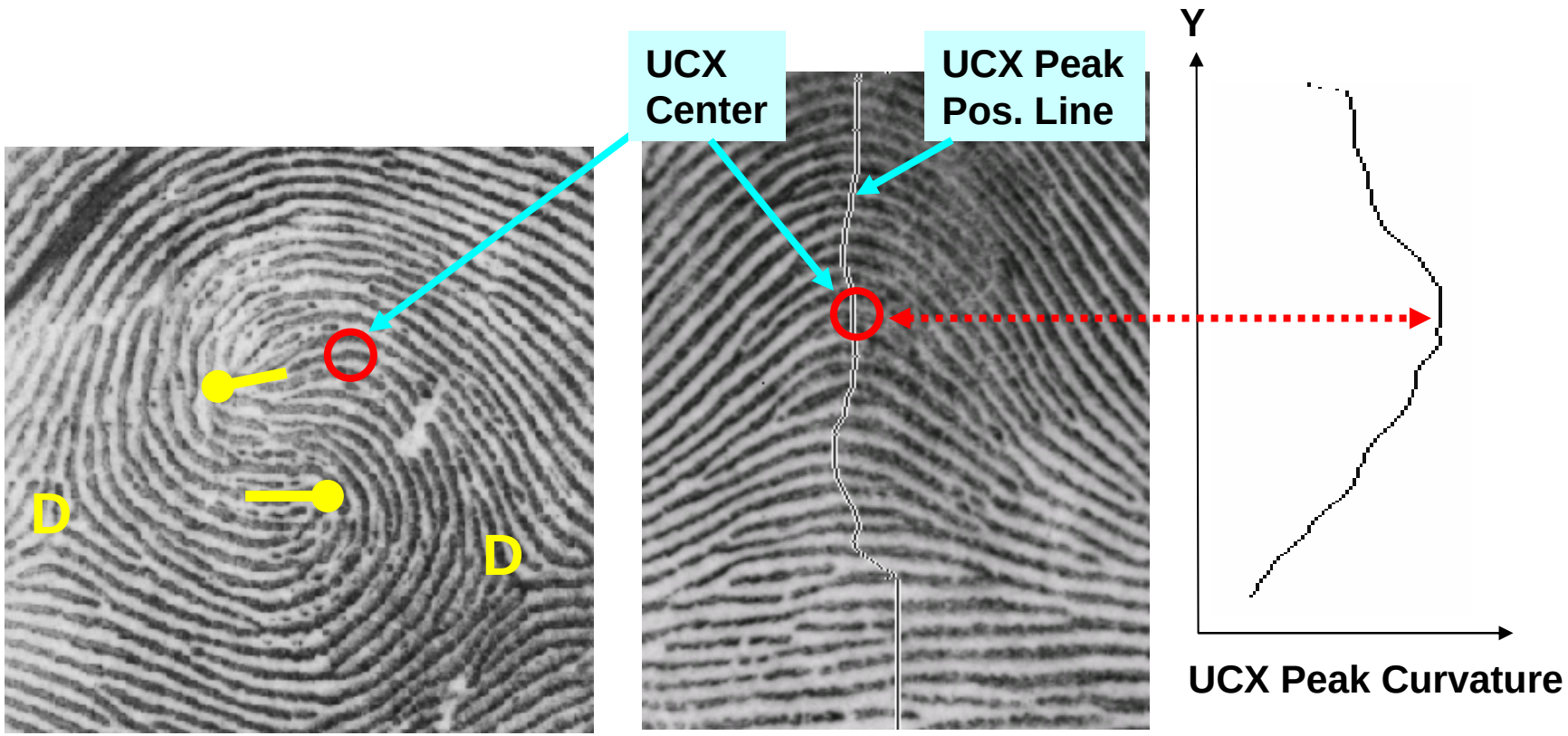


4. Matching Algorithms and Costs

UCX (Upward Convex) : New Definition for Center

Center of Peak Curvature Area with Upward Convex shape ridges

Unlike traditional core, UCX is defined on Arches and is more consistent



4.3 Algorithms for Auto Coded Latent Prints

- Matching algorithms for automatically coded latent prints have to deal with Unreliable Features

- Plan to develop new algorithm line (LE_A) for automatically coded features of latent prints if such demands really prospected

- LE_A will be much more sophisticated and complex (time consuming) than LE so that it can minimize Accuracy Degradation against "Unreliable Features"

5. Total Cost of Operation

5.1 Cost Factors for TCO

The following cost factors have to be considered:

1) Cost for Latent Encoding

2) Cost for Candidate Verification (Screening)

3) Cost for Identification (by Expert Examiner)

4) Cost for Matching Algorithm

5) Others

5.2 Feedback Operation (Reentry after No-Hit)

It is expected that the Feedback Operation could improve response time and save the cost for Manual Latent Encoding to some extent.

1) First, launch latent search with Fully Automatic

2) If no-match, Reenter the latent with Manual Encoding

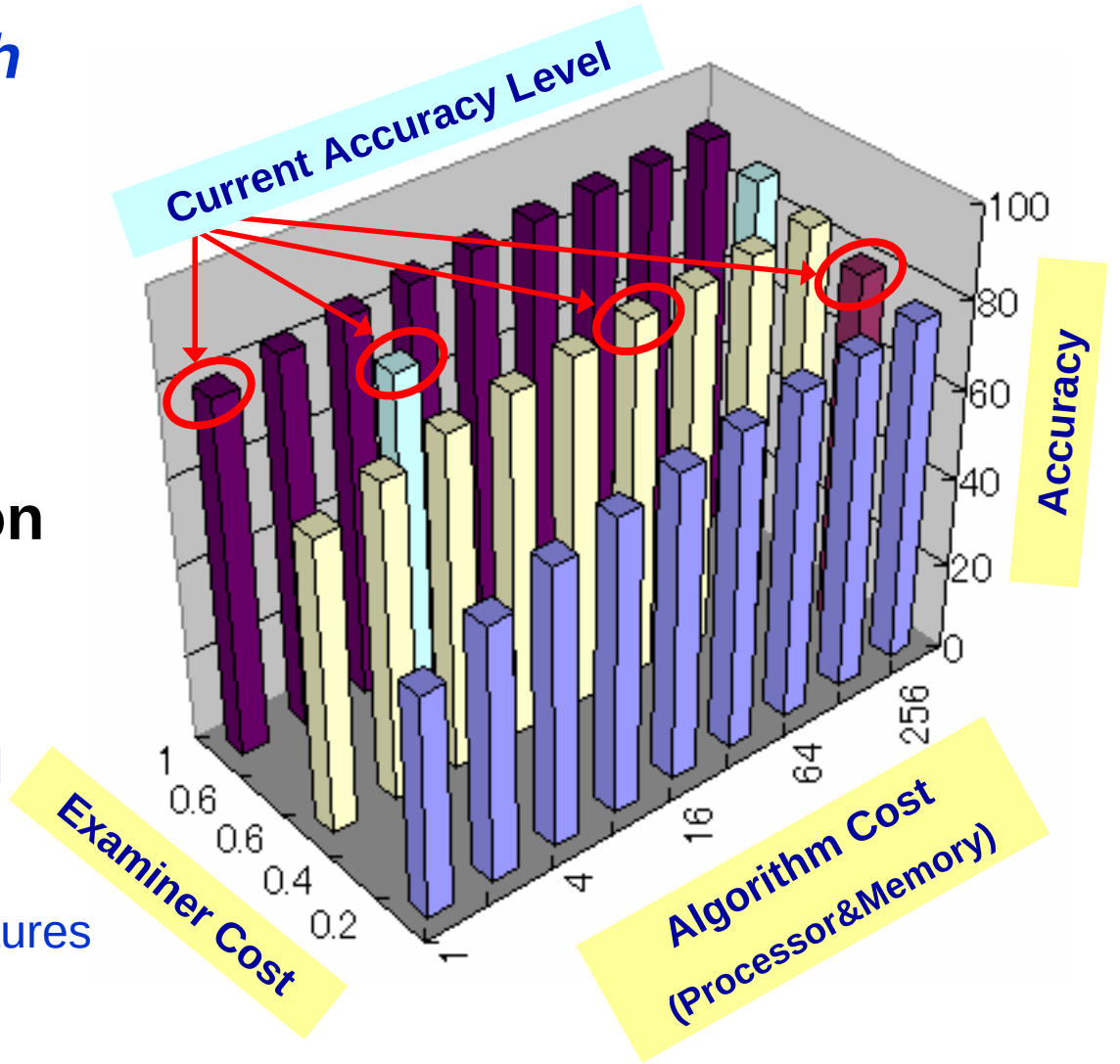
This expectation is valid only when mate prints exist in the repository (data base).

However, the latent hit rate is very low (10 %?) in the actual operation. Therefore, the Feedback Operation will almost double the matching workload and it does NOT seem to be worth for the additional cost.

5. Total Cost of Operation

5.3 Cost and Accuracy Tradeoff

Accuracy on Graph
→ Estimated
(Best Guess)



- Latent Encoding

Non-Experts Possible

- Primary Verification

Non-Experts Possible

- Identification

Expert Examiners Required

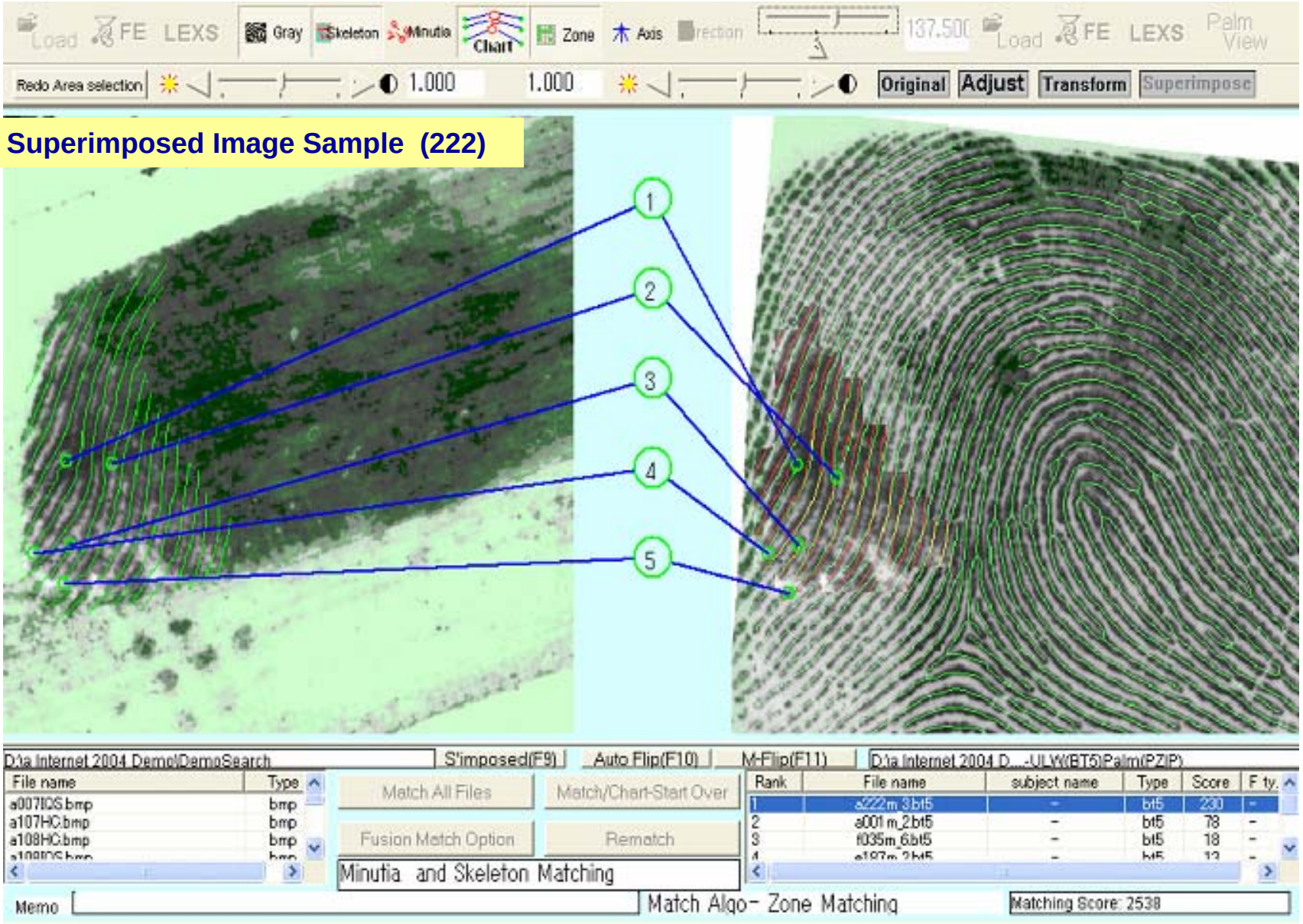
- Algorithm

Targeted to Unreliable Features

6. Visual Verification Support Tool

- **Friction Ridge Analyzer (FRA) is a GUI tool, that makes Visual Verification easy even for Non-Experts**
- **Skeleton Matching Function expands "matching area" to the area where minutia do not exist**
- **Image Transform Function, which transforms the shape of the candidate print image so that it looks like the shape of latent print image**
- **Superimpose Function overlays latent print image on the transformed candidate print image so that an operator can easily judge "conclusive non-hits" and find the "most probable candidate".**

6. Visual Verification Support Tool



7. Recommendations for New Demands

- **Latent Encoding → Semi-Auto**
  - **For Regular Cases: "Limited" Manual Edit by Non-Experts**
    1) Orientation and Crop (Search Area Setting)
    2) Minutia Removal or Zone Out (Unclear Area Setting)
    3) UCX and/or Core&Delta
  - **For Serious Cases: Full Manual Edit by Expert Examiners**
- **Primary Verification → Manual by Non-Experts**
  - **Up to 3 (?) candidates**
  - **Screening supported by the Superimposing tool**
- **Identification → Manual by Expert Examiners (Must)**
  - **Zero or one candidate per search for regular cases**
- **Algorithm → More Sophisticated Algorithm Needed**
  - **Allocate more cost to minimize Accuracy Degradation**
  - **Target to Automatically Coded Features**

Acknowledgements

All fingerprint images shown in this presentation are from NIST Special Data Base #27.

The contribution of this database to fingerprint matching research is notable.

I am grateful to Mr. Mike Murphy for his effort to produce an evaluation report for matching accuracy of ISP AFIS with detailed data.

I am also grateful to Mr. John Burt and Mr. Masanori Mizoguchi for their reviews and comments on the early drafts of this document.

References

"NIST Special Data Base 27" - http://www.nist.gov/srd/nistsd27.htm

"An Evaluation of NEC's Enhanced Matching Algorithms" Aug. 2001;
Mike Murphy, Illinois State Police (ISP)

Patents

Pending - JP Kai2004-078433 "Skeleton Matching"

Pending - JP Kai2003-173445 "Zone Matching"

Pending - JP Kai2004-078734 "Latent Ridge Analysis"

USP6,806,878 "Auto Skeleton Tracer"

USP5,703,958 "Distortion Correction"

USP5,040,224 "Fingerprint Core (UCX) Extraction"

USP4,956,870 "Dynamic Threshold"