New Chances and New Challenges in Camera-based Document Analysis and Recognition

In-Jung Kim

(*ijkim@inzisoft.com, ijkim@ai.kaist.ac.kr*)

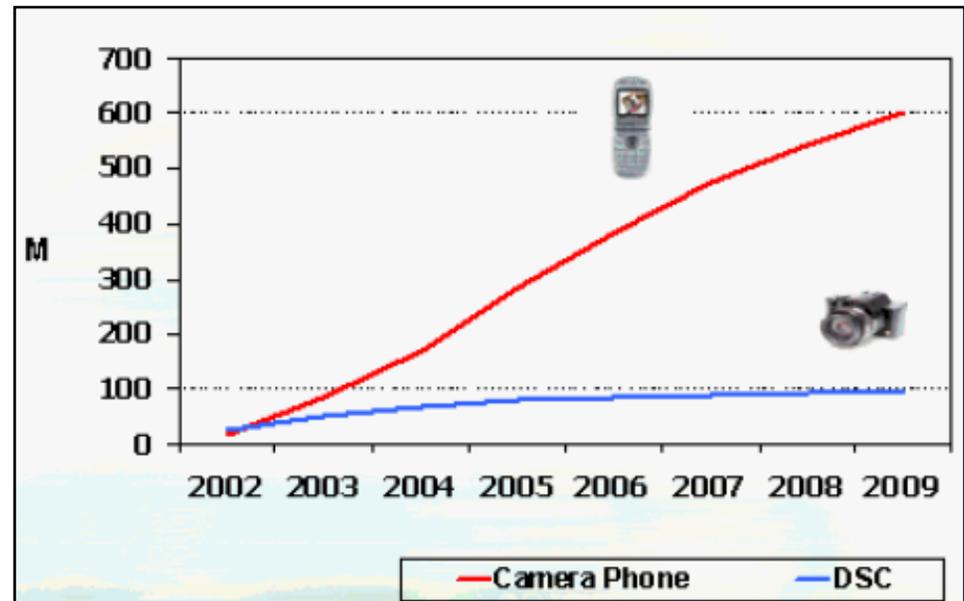CBDAR Aug. 29th, 2005

InziSoft CO.,LTD
www.inzisoft.com

◆ **Introduction**

◆ **Challenges in CBDAR**

◆ **A Perspective to CBDAR**

◆ *Mobile Reader^TM*: A Commercial CBDAR Product

◆ **Conclusion**

◆ Market of imaging devices

[IDC, JEITA, Data Quest, Samsung, …]

Year	Scanner	DSC	Camera Phone
2002	6.62M (highest)	28.3M/27.1M	–
2003	5.11M	46.8M/49.4M	80M
2004	4.37M	74M / 66.5M/62.4M	159M
2005	–	90M/77.3M/ 72M	–
2006	–	94M/85.2M/ 80M	–
2007	2.47M	90.2M	298M
2008	–	93.3M	366M
2009	–	94.9M/87M	–

[Strategy Analytics]

* DSC: Digital Still Camera

Market Trend

◆ **What's happening in the world of imaging device ?**

	Scanner	DSC	Camera Phone
Trend	**Shrinking**	**Expending**	**Exploding**
Extreme Point	**2002**	**2006~2009 (saturation)**	**Beyond 2009**
Max. #	**6.62M**	**94.9M**	**366~600M**

◆ **Performance improvement of camera is very rapid.**

< resolution of phone camera>

High-end phones

Mid-price phones

OCR-applicable

(pixel) — 10M, 7M, 5M, 4M, 3M, 2M, 1M

2002 2003 2004 2005 2006

(in Korean Market)
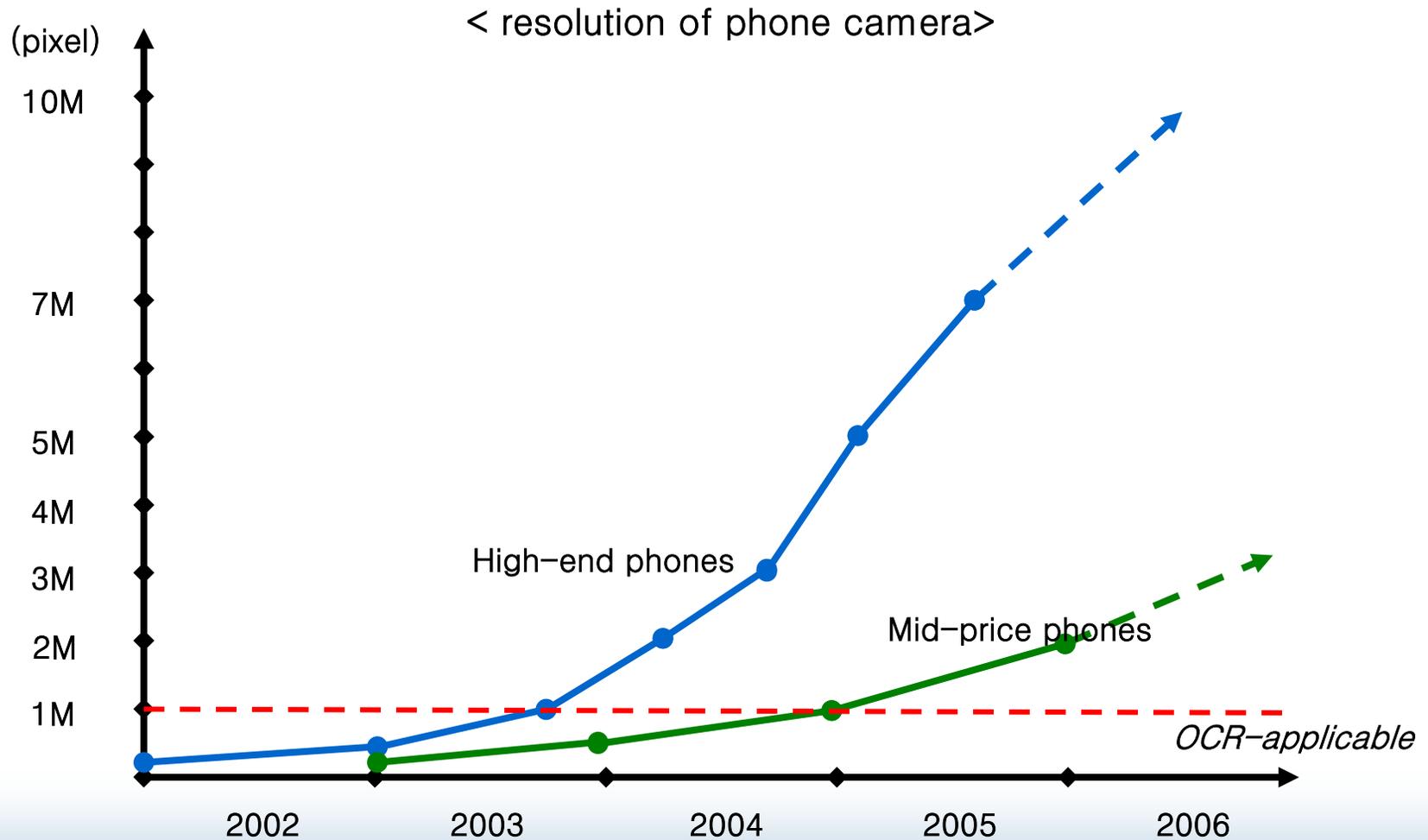
Future of Imaging Devices

Imaging Device for Document Recognition

High-speed Scanner

Large scale system
for Enterprise

Dominant in
scanning volume

Camera

General device for
Personal users

Dominant in
# of devices

ETC

SOHO/Special
purpose products

low-price scanner,
all-in-one device,
pen-type scanner,
...
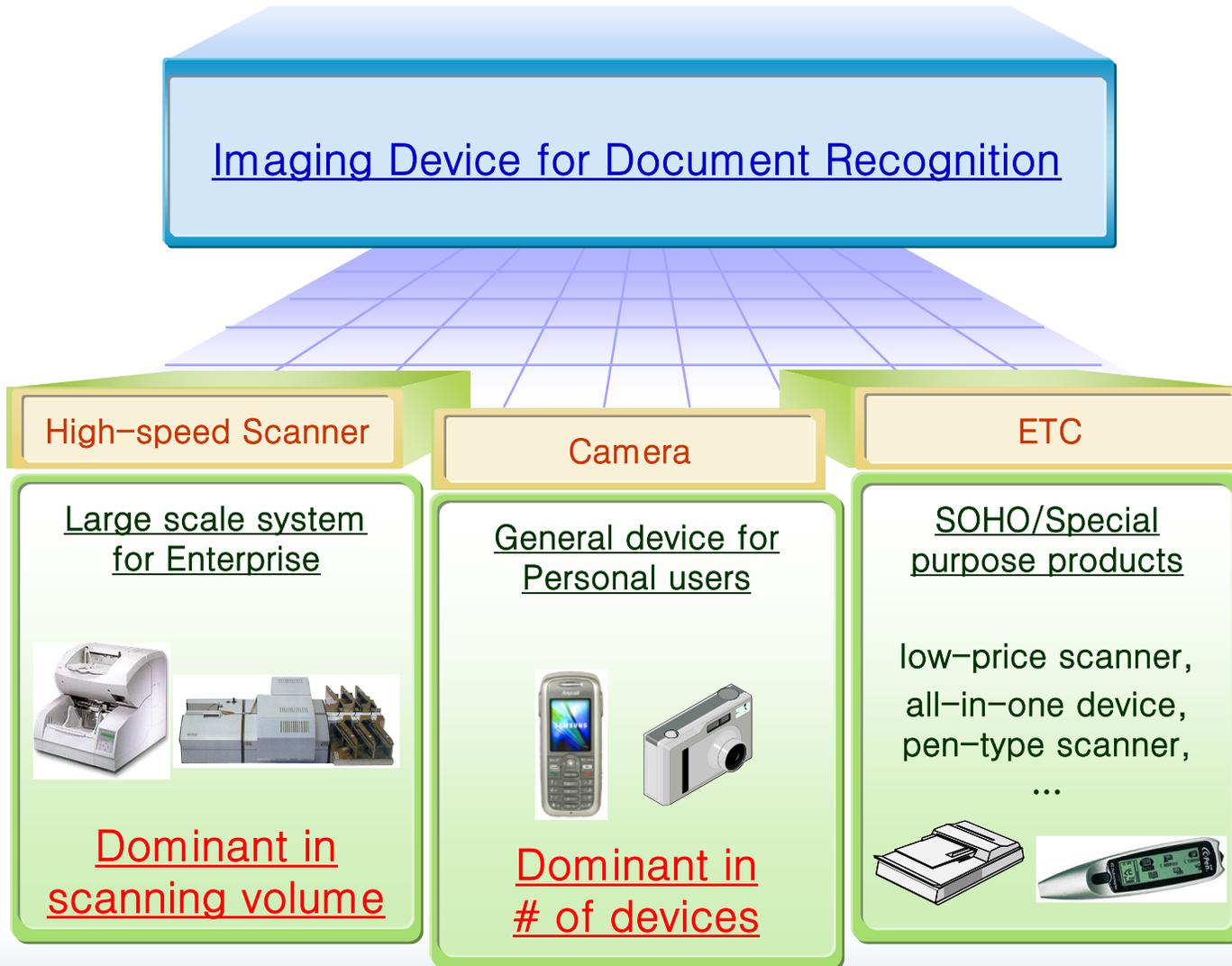
Input Device & Technology

Device

Pen-based Device

digitizer, e-Pen,
touch screen,
PDA, tablet PC,
Anoto Pen, …

Scanner

High-speed scanner,
Reader-sorter machine,
Flatbed scanner,
Pen-type scanner,
...

Camera Device

Camera Phone,
Digital Still Camera,
Camcorder,…

Technology

Motivation → Online Handwriting Recognition

Motivation → Offline Document Recognition

Motivation → Camera-based Document Analysis and Recognition

◆ **Introduction**

◆ **Challenges in CBDAR**

◆ **A Perspective to CBDAR**

◆ *Mobile Reader™*: **A Commercial CBDAR Product**

◆ **Conclusion**

Camera Device

Portable

Simple /fast

Contact−less

To user:
Freedom
to capture <u>anything</u>
under <u>any condition</u>

To researcher:
Opportunity & Challenge

<u>Variation in Imaging Condition</u>

• Lighting
• Angle
• Distance

<u>Variation in Target</u>

• Paper documents
• Non−paper objects
(sign, flag, 3D objects, monitor, screen, …)
• Objects in distance

Variation in
Imaging condition

Camera Document Images

Blurring,
Rotation,
Lighting/shadowing,
Scaling,
...

Difficult

Conventional
Document
Images

Out-door objects,
Non-paper objects,
Scene text,
...

Variation in
Target

Variation in Recognition Target

< Paper >

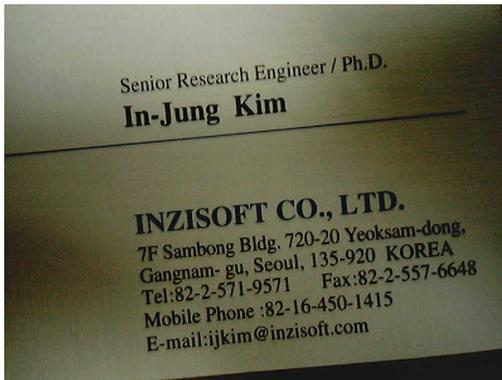### < Monitor >
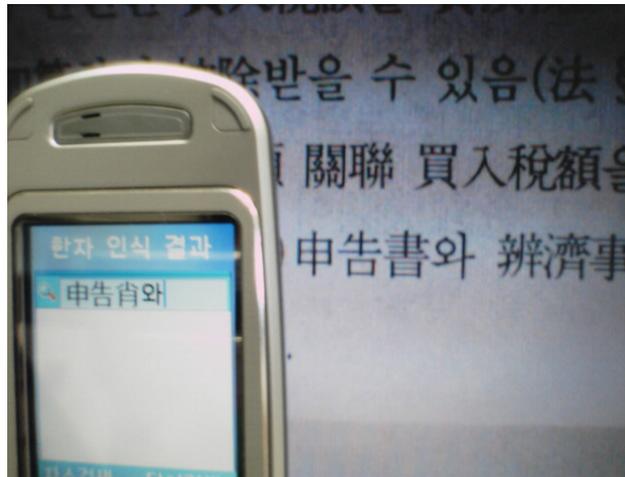
< Sign Boards >

< Screen >
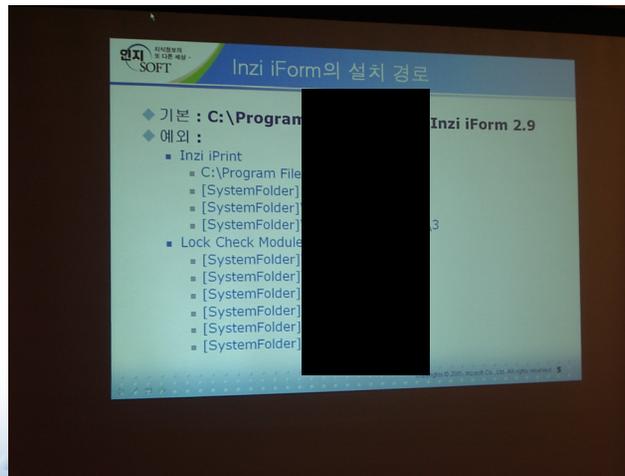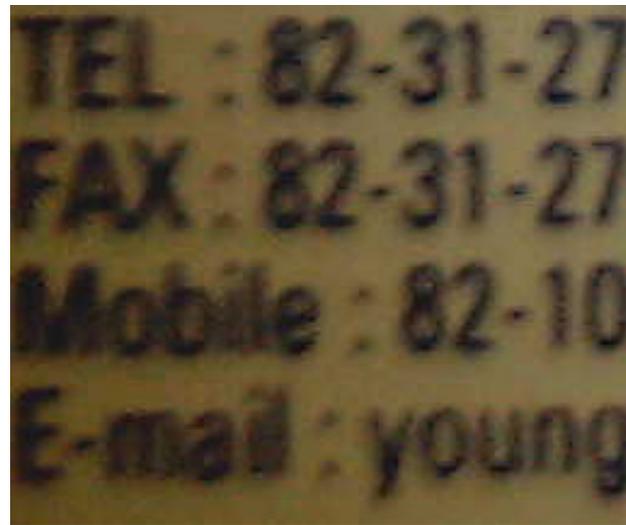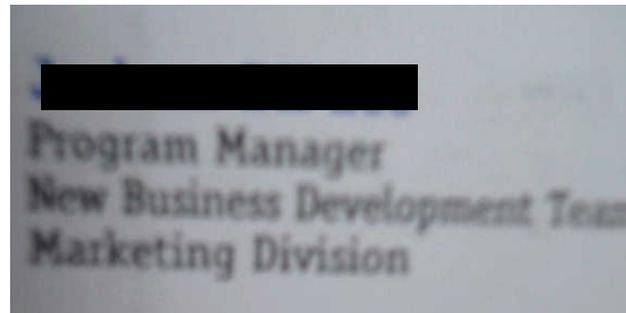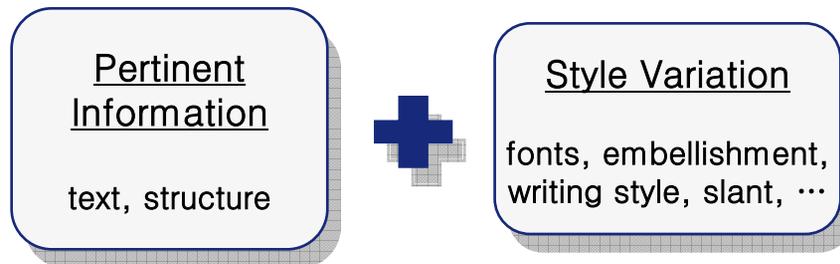
Variation in Imaging Condition

< Rotation >   < Blurring >   < Illumination/Shadowing >
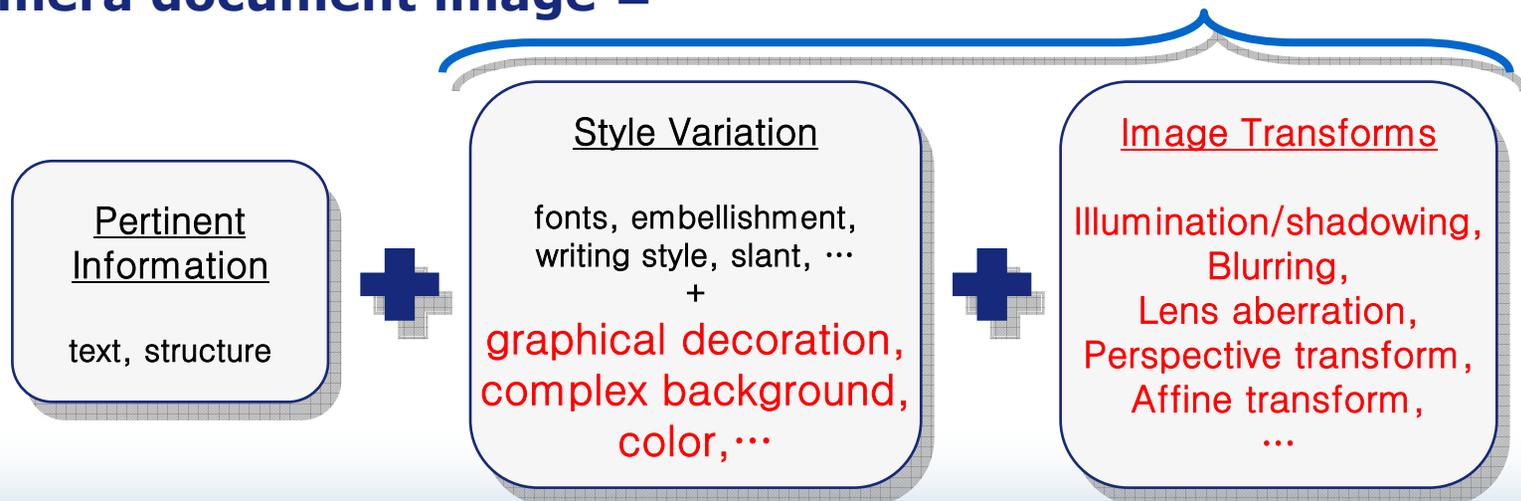
Conceptual Model of Doc. Image

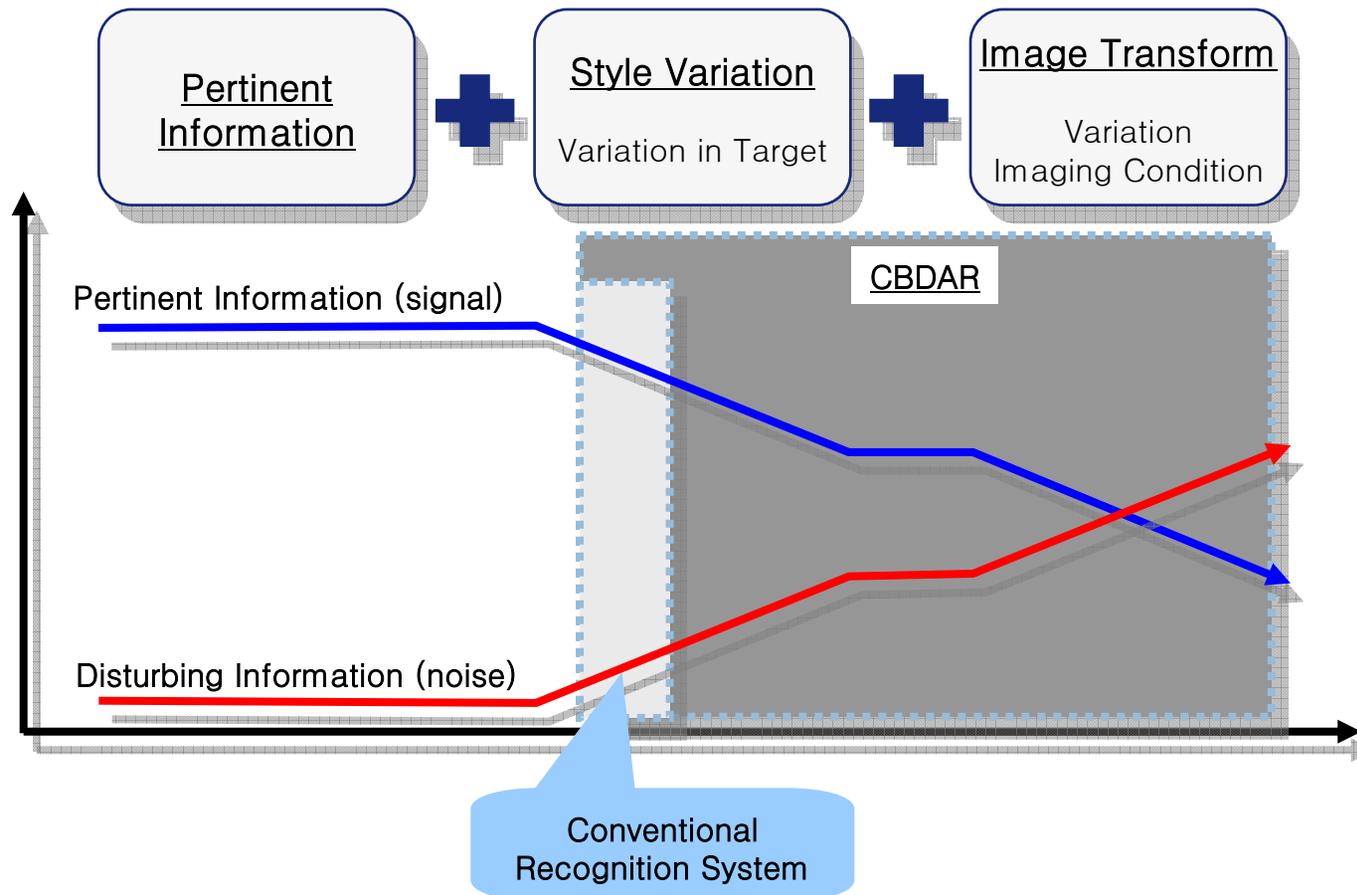◆ **Handwriting = Pertinent Info. + Style Variation [Lorette98]**

◆ **Document image =**

> **Pertinent Information**
>
> text, structure

+

> **Style Variation**
>
> fonts, embellishment, writing style, slant, ⋯

◆ **Camera document image =**

Irrelevant/disturbing information

> **Pertinent Information**
>
> text, structure

+

> **Style Variation**
>
> fonts, embellishment, writing style, slant, ⋯
> +
> graphical decoration, complex background, color,⋯

+

> **Image Transforms**
>
> Illumination/shadowing, Blurring, Lens aberration, Perspective transform, Affine transform, ...
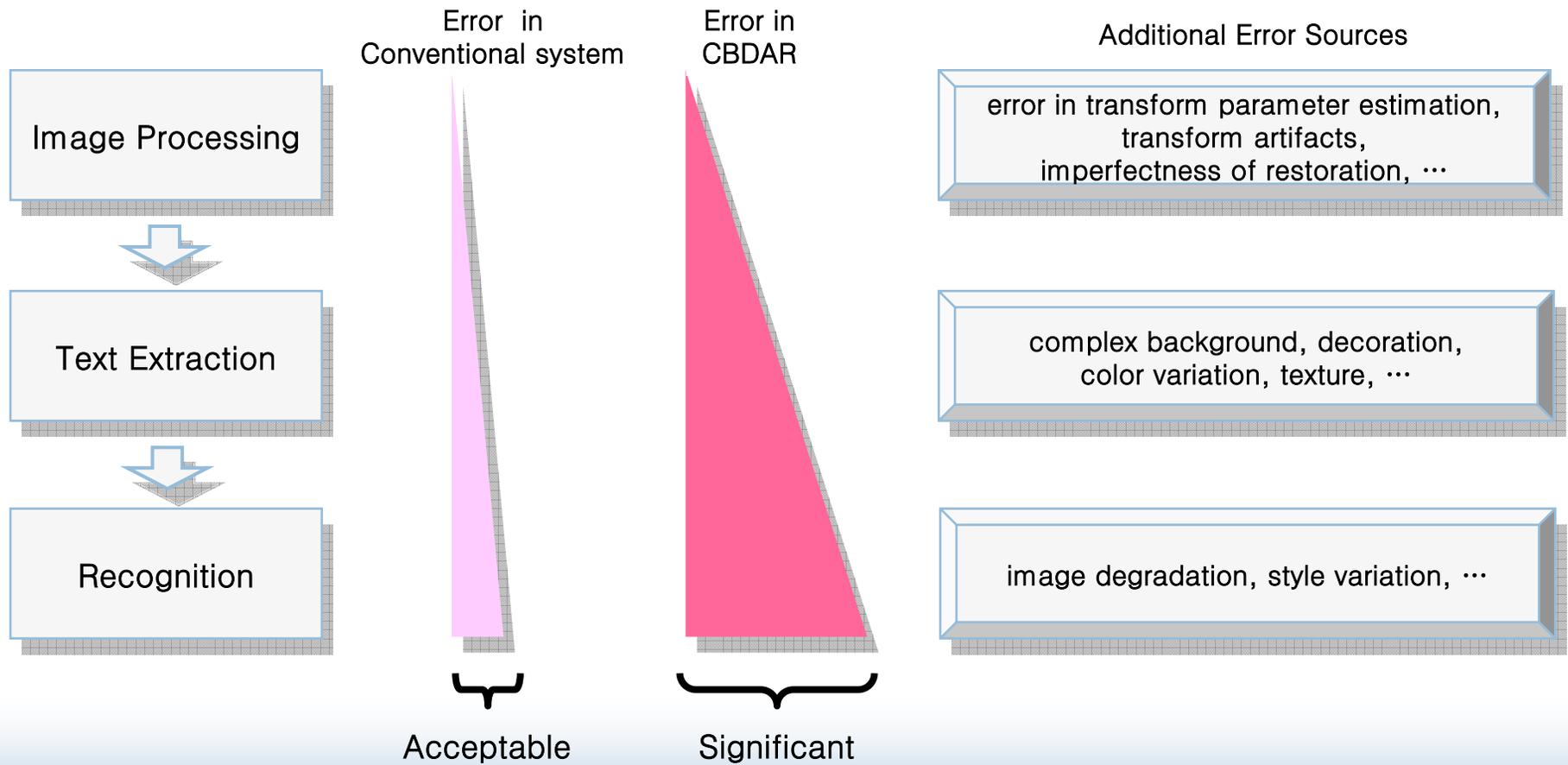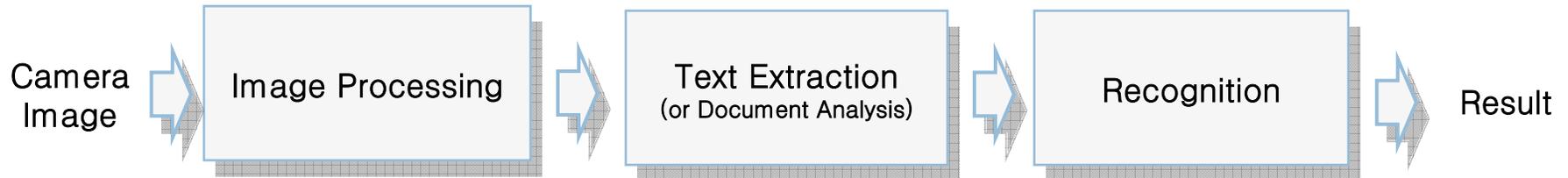
Loss of pertinent information

Growth of disturbing information

Difficulty of CBDAR

◆ Error growth can be very rapid !!

Image Processing

⬇

Text Extraction

⬇

Recognition

Error in Conventional system

Error in CBDAR

Additional Error Sources

error in transform parameter estimation, transform artifacts, imperfectness of restoration, ···

complex background, decoration, color variation, texture, ···

image degradation, style variation, ···

Acceptable   Significant

◆ **Introduction**

◆ **Challenges in CBDAR**

◆ **A Perspective to CBDAR**

◆ *Mobile Reader™*: **A Commercial CBDAR Product**

◆ **Conclusion**

◆ **General Framework of Document Recognition System**

Camera Image → **Image Processing** → **Text Extraction** (or Document Analysis) → **Recognition** → Result

◆ **For CBDAR, we need**

- More intelligent procedures
  - Image processing
  - Text extraction
  - Recognition

  • Compensate <u>image transform</u>
  • Adapt to <u>target variation</u>

- More intelligent framework
  - Consolidated system
  - Knowledge-driven system

  • Alleviate accumulation of error
  • Supplement Information

Image Processing

◆ **New mission: <u>compensation of image transforms</u>**

◆ **Transforms in camera image**

- Affine transform
- Perspective transform   } Reversible
- Lens aberration
- Illumination / shadowing
- Blurring (ill-focus)   } Irreversible
- Aliasing/jagging (low resolution)
- ETC.

Image Processing

For Reversible Transforms

Inverse Transform

Process

1. Estimate transform parameters
2. Apply Inverse transform

Problems

- Estimation of transform parameter (angle, scale, translation, …)
- Transform artifacts

For Irreversible Transforms

Non-trivial Techniques

- Illumination/shadowing
  - Intensity normalization
  - Adaptive binarization
  - Illumination removal

- Blurring, aliasing/jagging
  - Super-resolution
  - Image analogy
  - Mosaic image generation
- …

◆ Normalize lighting variation

< original image >	< illumination >	< result >

◆ **Synthesizing a high-resolution image from a set of low-resolution image**

Knowledge about Document [Park05]

< low resolution images >   < high resolution image >

Image Analogy [Herzman01]
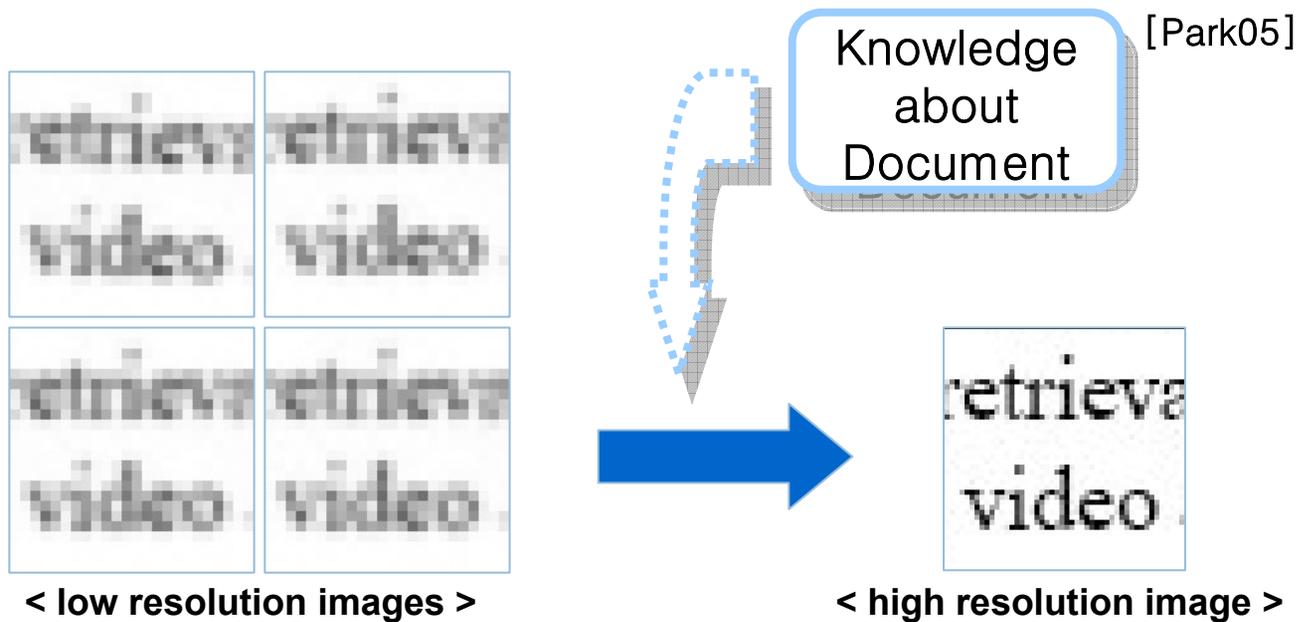
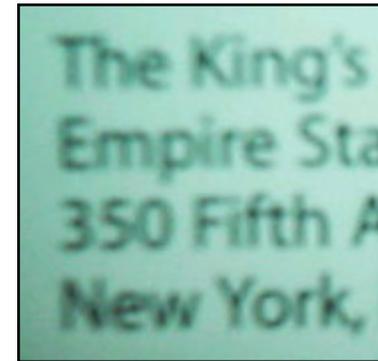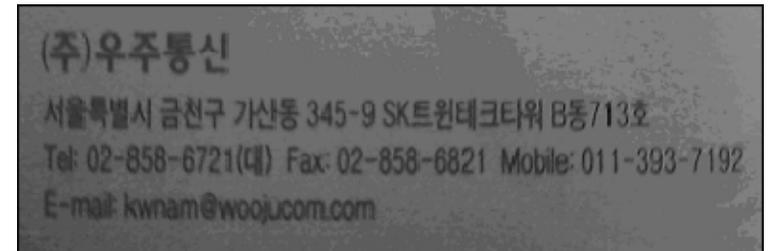< blurred images >   < sharp images >

Training →

< input image >

Image Processor

?

Text Extraction
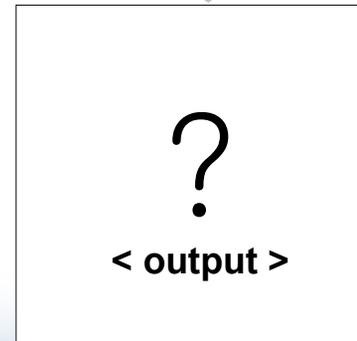
◆ **Mission: separating (decorated) text from (complicated) background (in row quality image)**
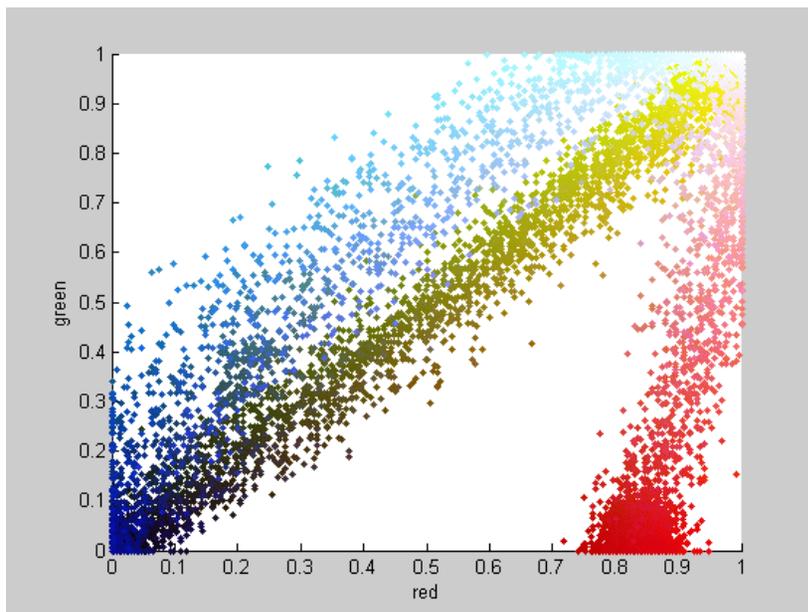
< outdoor text >

< paper document >
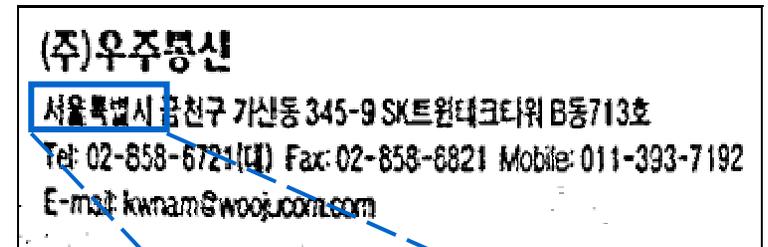
◆ **Available features**

- Homogeneity in color/intensity
- Special characteristics of text
    - Regularity, size, thickness, arrangement,…
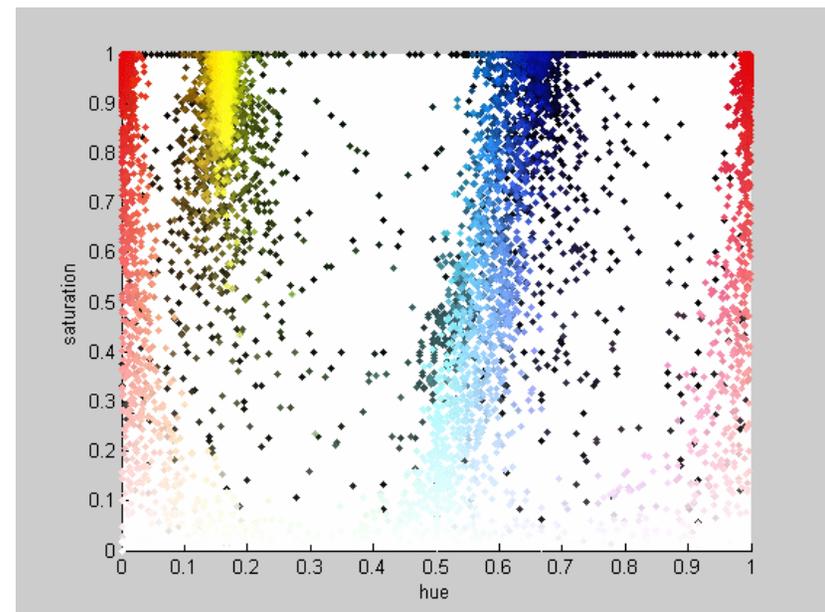
→ We need more ideas !

◆ **Color segmentation**
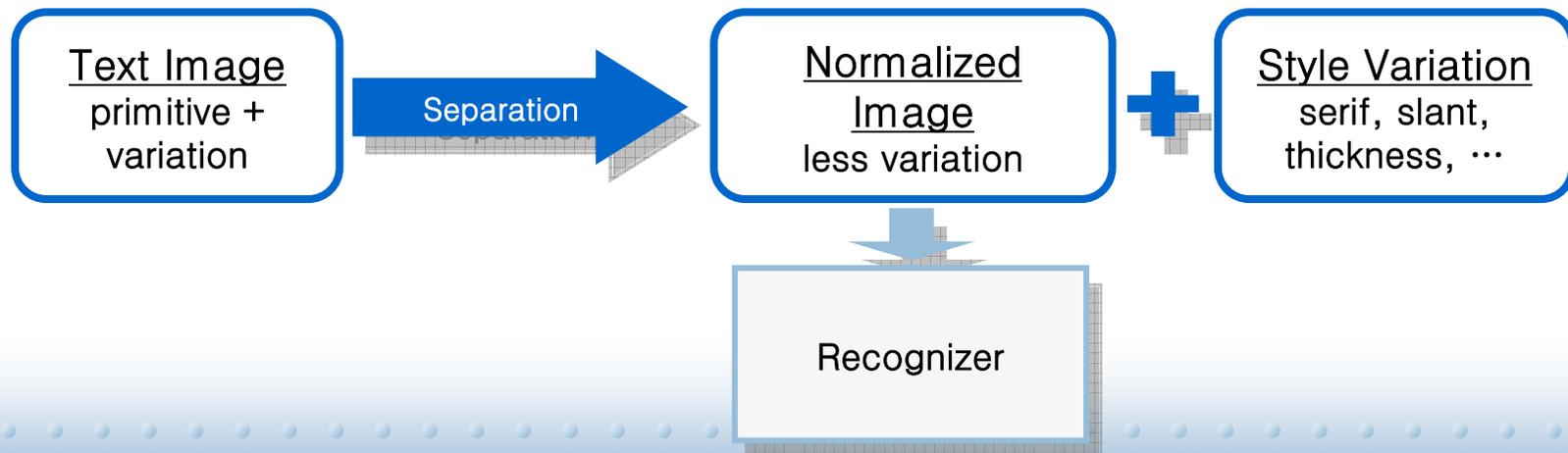

< Original Image >

< RGB >

< HSV >

◆ **New requirement: Robustness to <u>style variation</u>**

- ■ Difficult to collect training samples

◆ **(Proposal) <u>Explicit modeling</u> of style variation**

| <u>Text Image</u>
primitive +
variation | → Separation → | <u>Normalized
Image</u>
less variation | ✚ | <u>Style Variation</u>
serif, slant,
thickness, ⋯ |

↓

Recognizer

Recognition

◆ **New requirement: Robustness to <u>information loss</u>**

- ■ Intrinsic fuzziness

Loss of key feature →

Types	Examples
Indistinguishable Pattern	**C-c, P-p, O-0-o,1-l-I, g-9,...**
Confusing Pattern	ɑ-o, **O-Q,** l-J, **J-],** C-G, 5-S, f−t−l, ...
	cl-d, vv-w, ɑ-ci, IV-N, **l_-L,U-Ll...**

- ■ <u>Additional problem in CBDAR</u>: <u>Loss of key feature</u>

"L" or "I ," ?
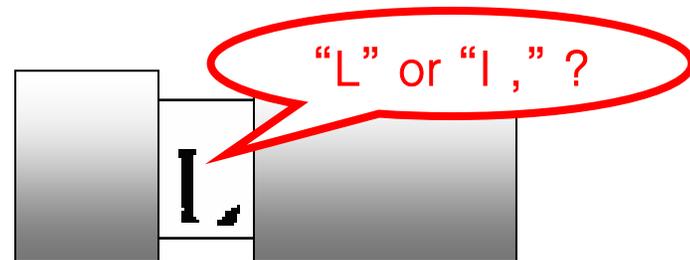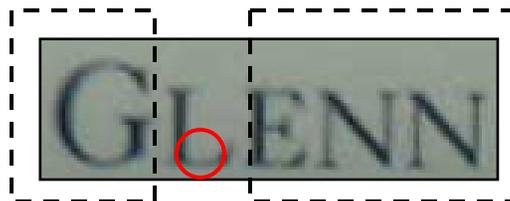
◆ **There are plenty of patterns like that.**

Internet: me◌d◌ ◼◼ it.edu

◌iel@ ◼◼ umbia.edu

e or c ?

Global Alliances Manager

a or o ?

cmr ◼◼ osstec,com

t or l ?

◆ General Framework of Document Recognition System

Camera Image → | Image Processing | → | Text Extraction (or Document Analysis) | → | Recognition | → Result
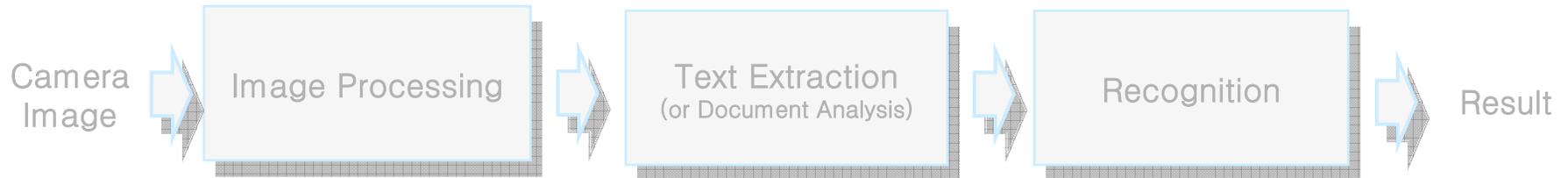
◆ New requirements for CBDAR
- More intelligent procedures
  - Image processing
  - Text extraction
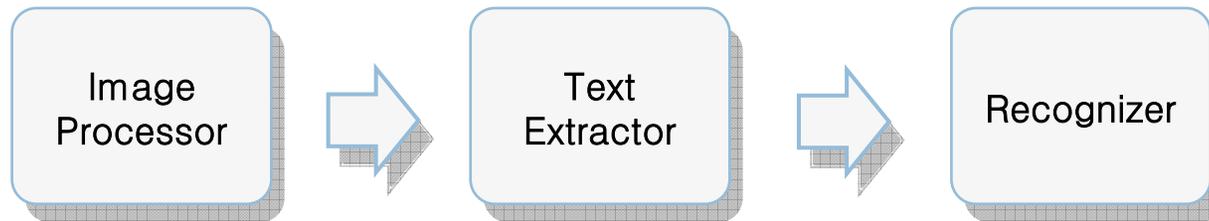 - Recognition

> • Compensate Variation in Imaging Condition
> • Adapt to Target Variation

- **More intelligent framework**
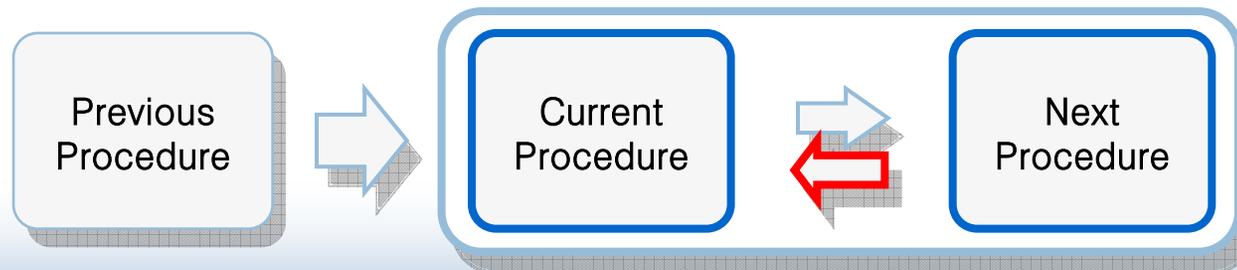 - **Consolidated system**
  - **Knowledge-driven system**

> • Alleviate error accumulation
> • Supplement Information loss

Consolidated System

◆ **Sequential system: No way to recover error from previous step**

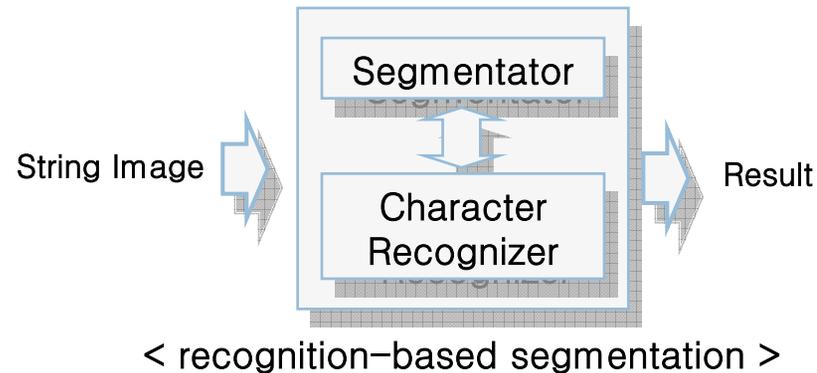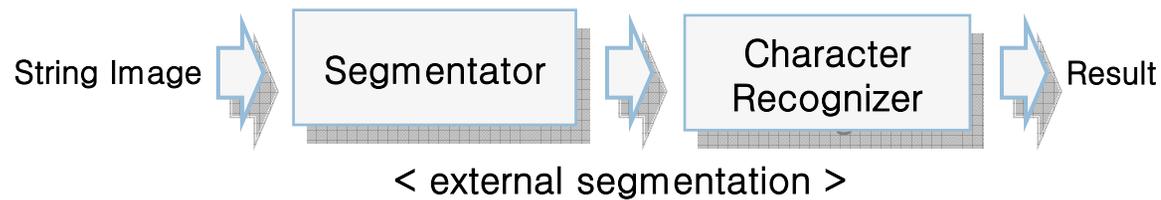| Image Processor	→	Text Extractor	→	Recognizer

→ In CBDAR, each procedure is not very reliable !

◆ **(Proposal) Consolidated system: Feed-back mechanism from next procedure**

Previous Procedure	→	Current Procedure	⇄	Next Procedure

◆ Recognition-based segmentation

```
String Image → [ Segmentator ] → [ Character Recognizer ] → Result
```

< external segmentation >

```
String Image → [ Segmentator ⇅ Character Recognizer ] → Result
```

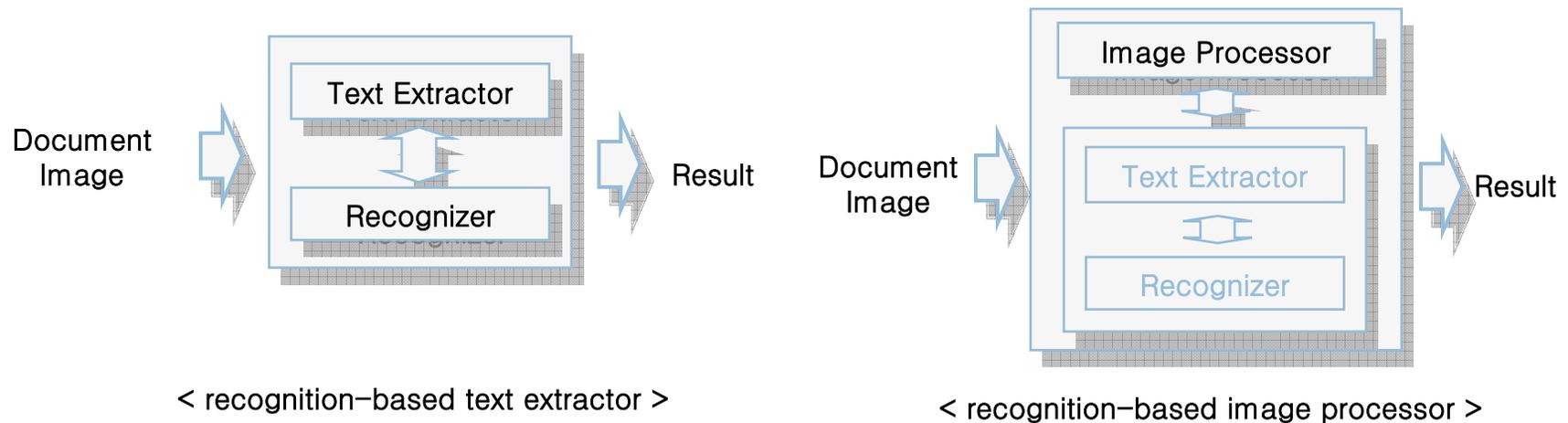< recognition−based segmentation >

Recognition-based segmentation is BETTER!!
Because it looks for global optimum.

◆ Possible consolidation of procedures

Document Image → [Text Extractor / Recognizer] → Result

< recognition−based text extractor >

Document Image → [Image Processor / Text Extractor / Recognizer] → Result

< recognition−based image processor >

◆ Totally consolidated system: Tree

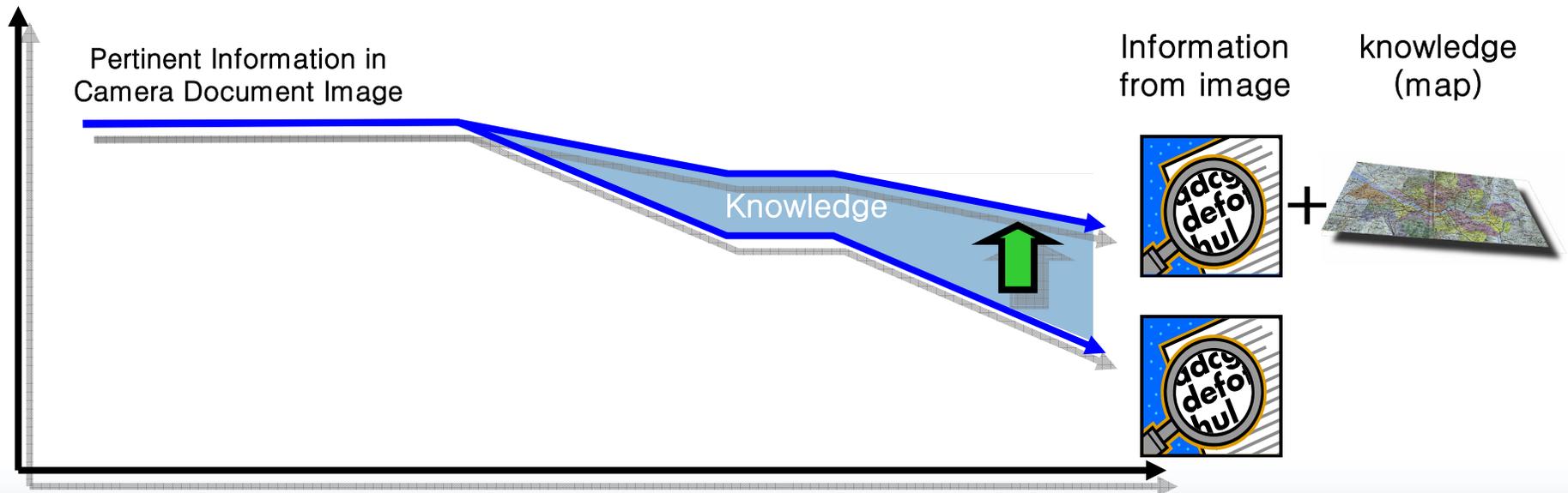Image Processing   Text Extraction   Recognition
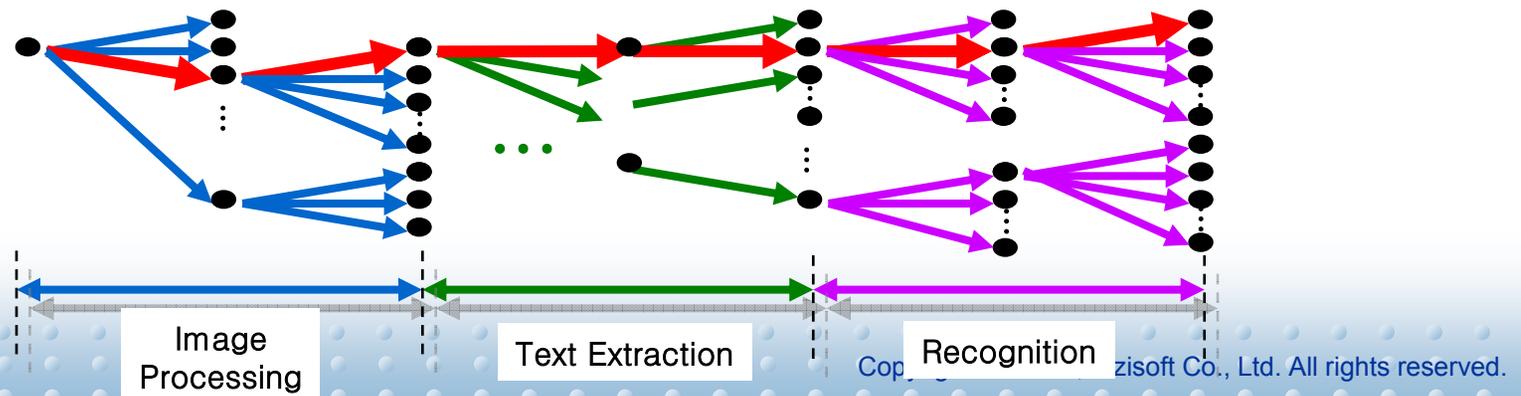
◆ Advantage

- Global optimization

◆ Issues

- Unified formulation for heterogeneous procedures
  - How to define optimization function

- Complexity reduction
  - Edge pruning
    - Definite decision at a particular procedure
  - Heuristic search
    - How to find a good heuristic function?

◆ Knowledge can mitigate information loss

Pertinent Information

Style Variation

Variation in Target

Image Transform

Variation in Imaging Condition

Pertinent Information in Camera Document Image

Information from image

knowledge (map)

Knowledge

◆ Available knowledge for each procedure

| Image Processing | What kind of transforms does the camera generate? | mathematical model of camera, Common properties of transforms, … |

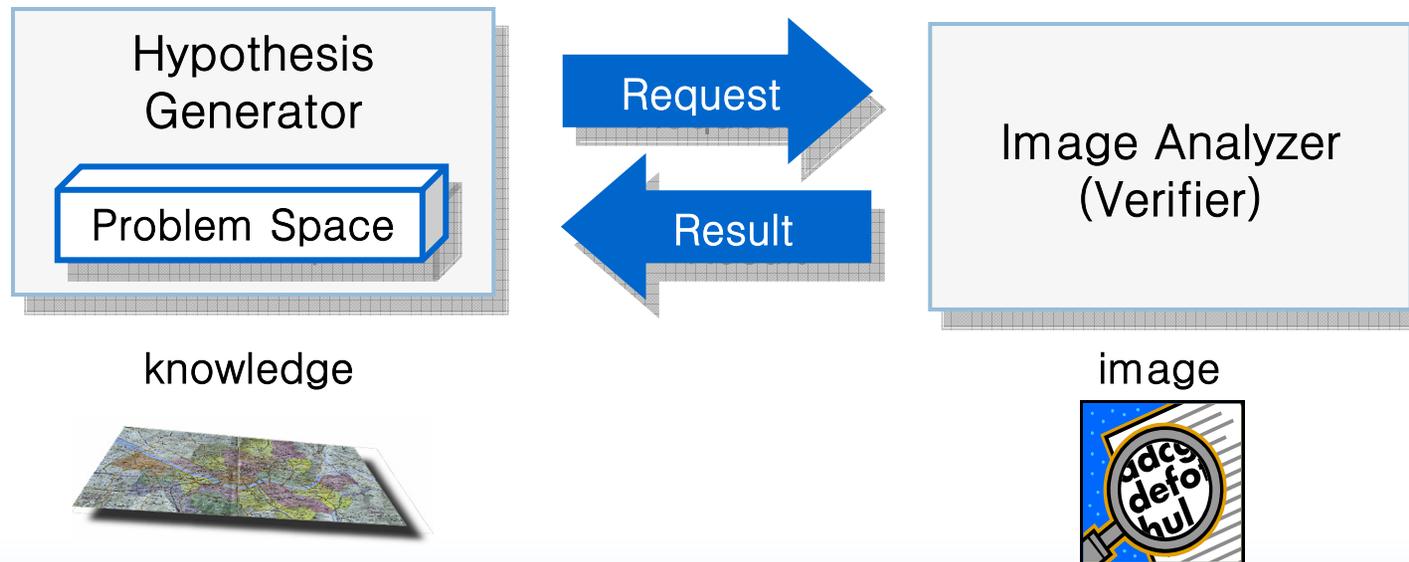| Text Extraction | What is intrinsic difference between text and background? | regularity, patterns of embellishment, printing mechanism, … |

| Recognition | What is the basic primitive of text? What is the popular property of text? | patterns of variation / decoration / degradation, lexicon, grammar, n−gram transition probability, … |

◆ Recognition system led by knowledge

- Feasible when strong domain knowledge is available
- Effective when information in image is not sufficient

Ex) address reader, URL reader, …

Hypothesis Generator	Request →	Image Analyzer (Verifier)
Problem Space	← Result	

knowledge

image

< Hypothesis−verification framework >

◆ **Introduction**

◆ **Challenges in CBDAR**

◆ **A Perspective to CBDAR**

◆ *Mobile Reader^{TM}*: **A Commercial CBDAR Product**

◆ **Conclusion**

◆ Processor speed and memory are limited

	Mobile Phone	PC
Speed	100~200 MHz	3 GHz
Memory	1~2 M	512 M ~ 1 G

- Not plentiful, but OCR is applicable
  - OCR should be well-optimized
- Memory or time consuming technologies are not applicable
  - Large scale linguistic processing, …
  - Super-resolution, image analogy, …
  - → Client-server architecture, specialized H/W

◆ **A Mobile OCR and Image Processing Tool Kit developed by Inzisoft**

< Capture >   < Text Extraction >   < Recognition >

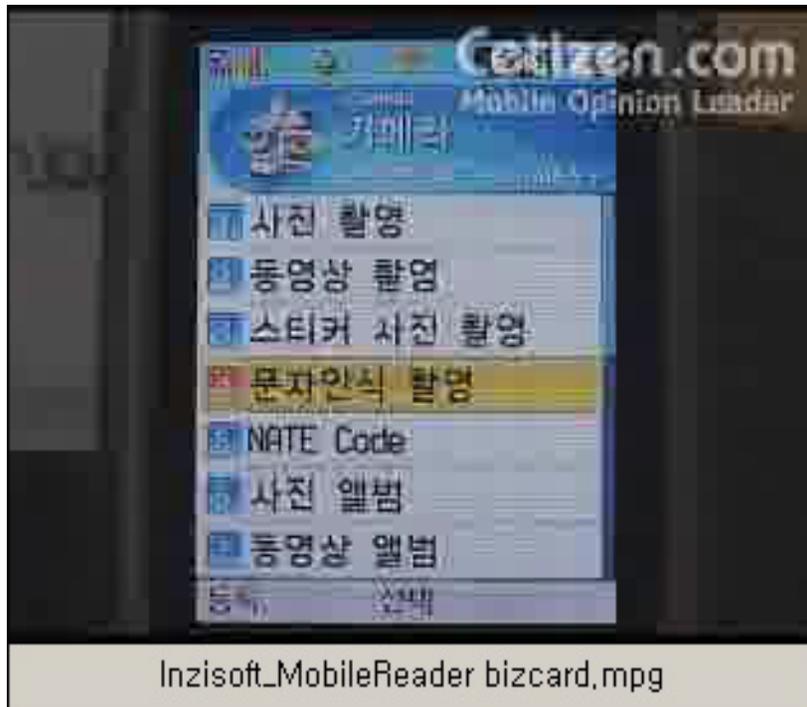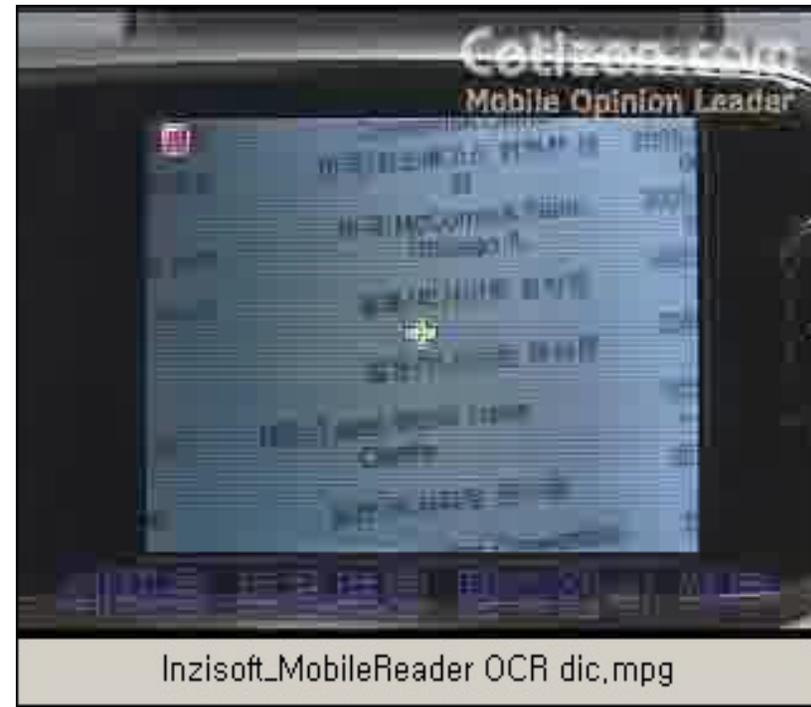◆ **H/W requirement**

- Processor: ARM-9 or comparable speed
- Camera: 1 M pixel, <u>AF or macro mode (5~8.5 cm)</u>

Mobile Reader Demo.
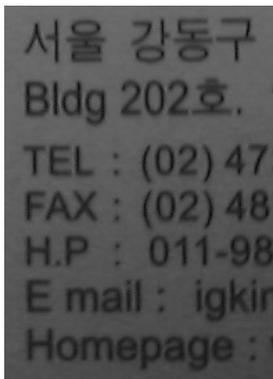

< Business Card Reader >

< OCR Dictionary >

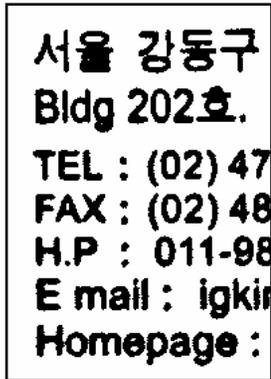* These movie clips were made by Cetizon, a mobile device review company

◆ Adaptive Binarization

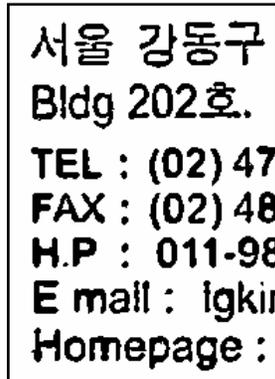- ■ Text / background separation
- ■ Adaptation to ill-focusing / noise

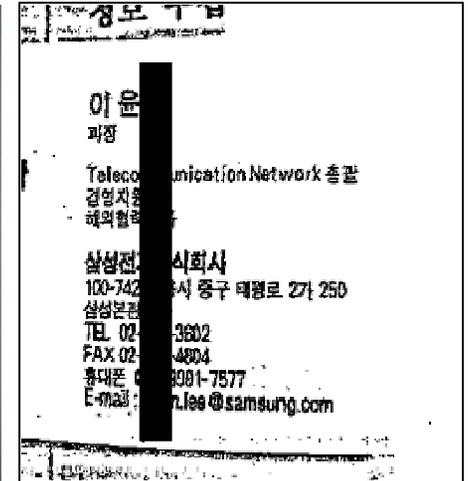Most important factors for high-performance

Dark Image

local binarization

Mobile Reader binarization

Rough Image

Mobile Reader Binarization

Mobile Reader Technology

◆ Tiny Recognizer

- ■ ROM (image processor + recog. engine + recog. model)

Language	# of characters	Rom
Alpha-numeral	< 100	600Kb
+ Hangul	2350	1Mb
+ Hanja	4888	1.9Mb
Chinese + alpha-numerals	6033	1.2~1.5Mb

(∗ Recognizers for other languages are under–development)

- ■ RAM (temporal memory for binarization & recognition)
  - ■ 500 Kb (cf. input image: 1Mb)
- ■ Performance for well-focused camera image
  - ■ Alpha-numeral: 98 ~ 99.5 %
  - ■ Hangul: 97 ~ 98.5 %
  - ■ Hanja, Chinese: 97 ~ 99 %

∗ Hangul: Korean char. ∗ Hanja: Chinese chars used in Korea

◆ *Mobile Reader* was embedded on more than 16 camera phones

(1) Pantech&Curitel

PH-K1000V(T)
[KTF]

PH-K1500
[KTF]

PH-K2500V
[KTF]

PH-K3000V
[KTF]

PT-S110
[SKT]

PT-K1100
[KTF]

PT-K1200
[KTF]

(2) LG

LG-KP3800
[KTF]

LG-SV360
[SKT]

LG VX-9800
[Verizon]

(3) Samsung

SPH-A800
[Sprint]

SPH-V7800
[KTF]

SCH-V770
[SKT]

◆ **Some love it, but some don't !**

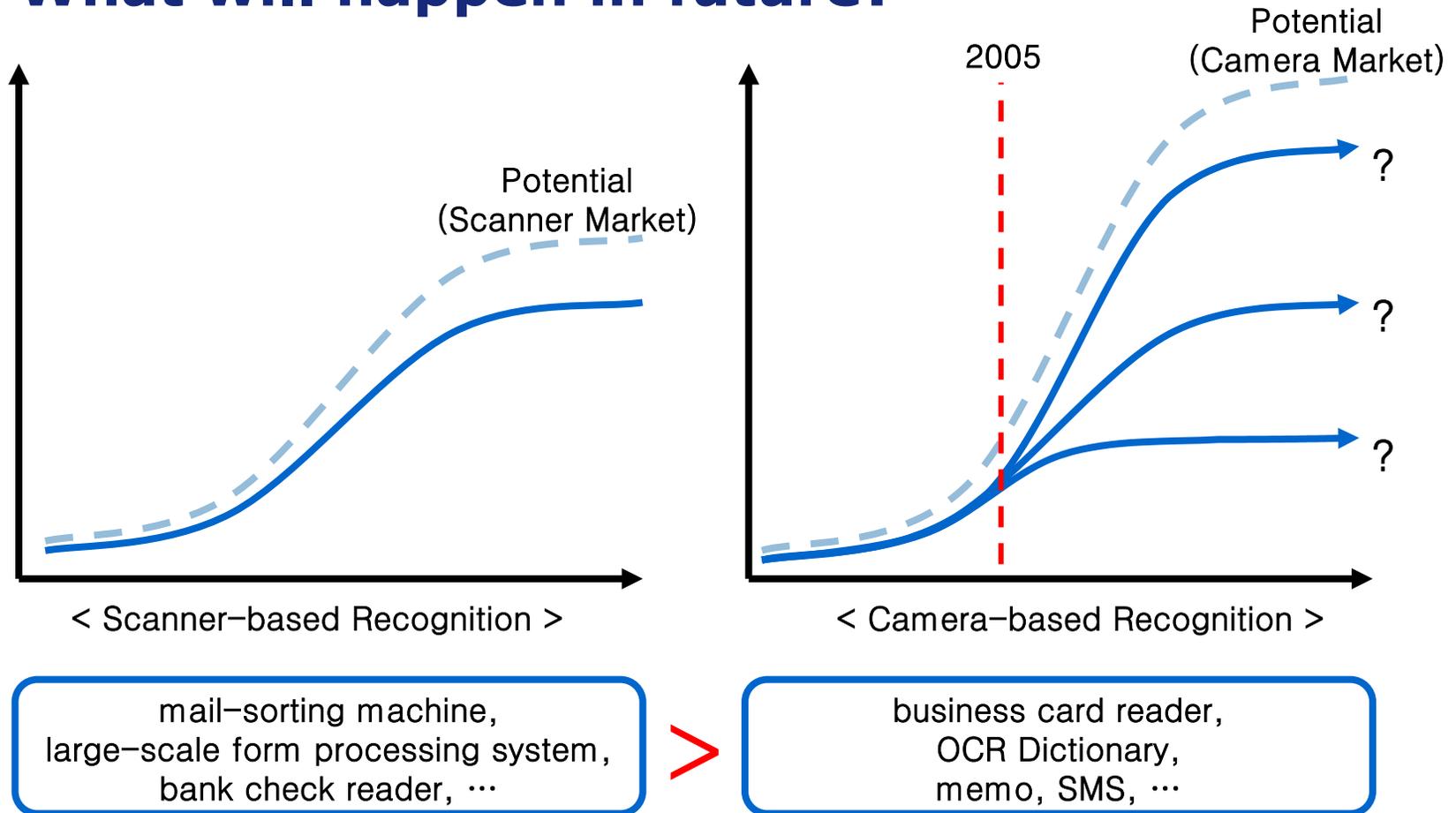|	Advocates	Critics	Impacting Factor
Function	**Very necessary**	**Not necessary**	**User's identity**
Performance	**Very nice**	**Unacceptable**	**Performance of Camera, Shooting skill**
Convenience	**Convenient**	**Worse than typing**	**Language, Typing skill**
Consequently...	**Wonderful / Interesting / Much better than expected**	**Useless/ Unconcerned**	

◆ **Mobile OCR market was successfully open**

◆ **But, we are still on left side of chasm.**
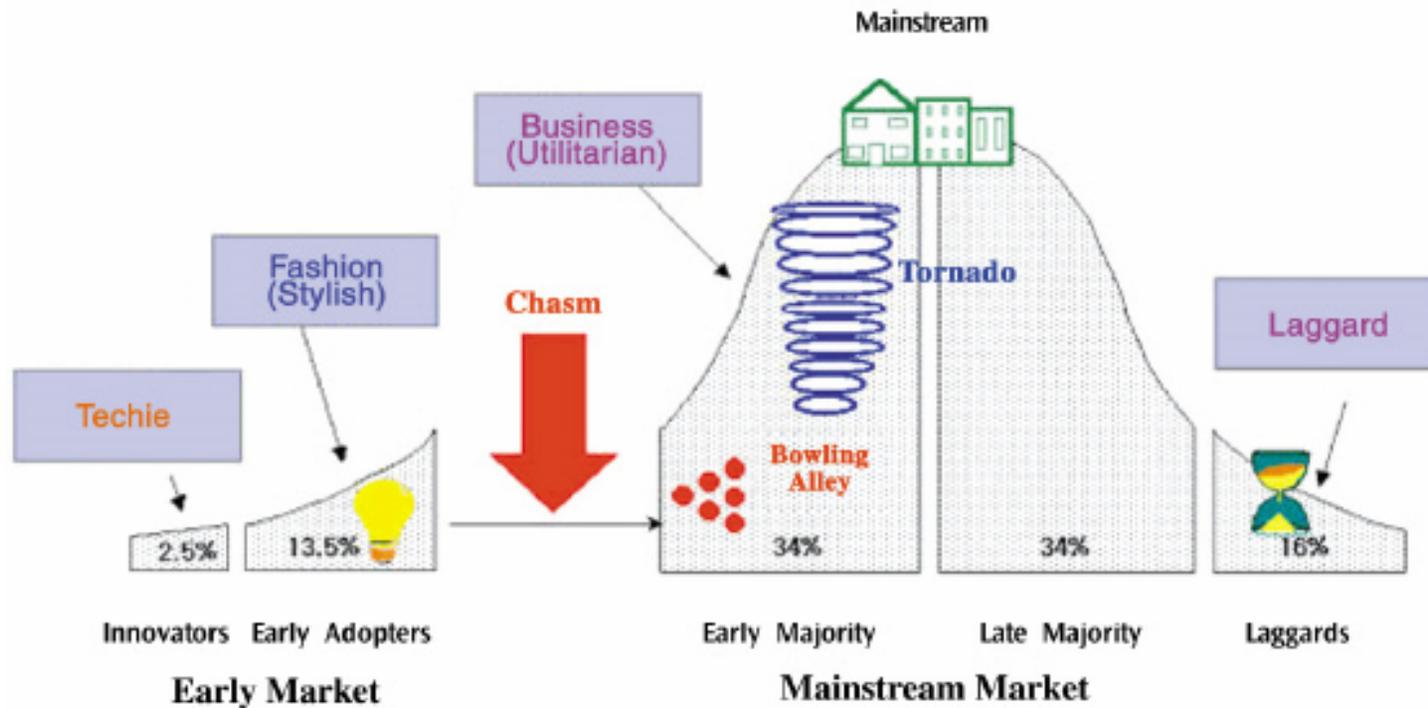
We need more idea and more improvement !!

◆ What will happen in future?

< Scanner-based Recognition >

< Camera-based Recognition >

| mail-sorting machine, large-scale form processing system, bank check reader, ⋯ | > | business card reader, OCR Dictionary, memo, SMS, ⋯ |

CBDAR need more Killer Applications !!

- **Mobile phone is more than a communication device**
  - Center of digital convergence.
    - Multimedia player, PIMS, payment method, …
  - Central device of Ubiquitous World

- **Main input method (numeric keypad) of mobile phone is inconvenient**
  - People want an alternative input method

- **Mobile phone embeds many technologies to make a synergy with CBDAR**
  - Text-to-Speech, voice recognition, network connection, …

Conclusion

◆ **Business opportunities are coming**
- **Several hundred millions of people** will carry **high-performance camera** and **processor** in their pocket

→ *A new continent for pattern recognition researchers*

◆ **But, there exist big challenges**
- Develop technologies to recognize camera image robustly
  - Seems more difficult than conventional document recognition system
- Needs killer applications of camera

Thank you for your attention !!

Acknowledgement to
Prof. Kim, Jin-Hyung
Mr. Kwon, Young-Hee
Dr. Kim, Kwang-In