



Beyond Camera-Based Document Analysis and Recognition: Towards Camera-Based Human Document Interactions

Koichi Kise

Osaka Prefecture University, Japan



Self-Introduction

- Koichi Kise
 - Professor, Osaka Prefecture Univ., Japan
 - Editor-in-Chief, IJDAR
 - Research
 - layout analysis
 - camera-based document retrieval
 - human document interaction
 - intelligence augmentation
 - Started CBDAR in 2005

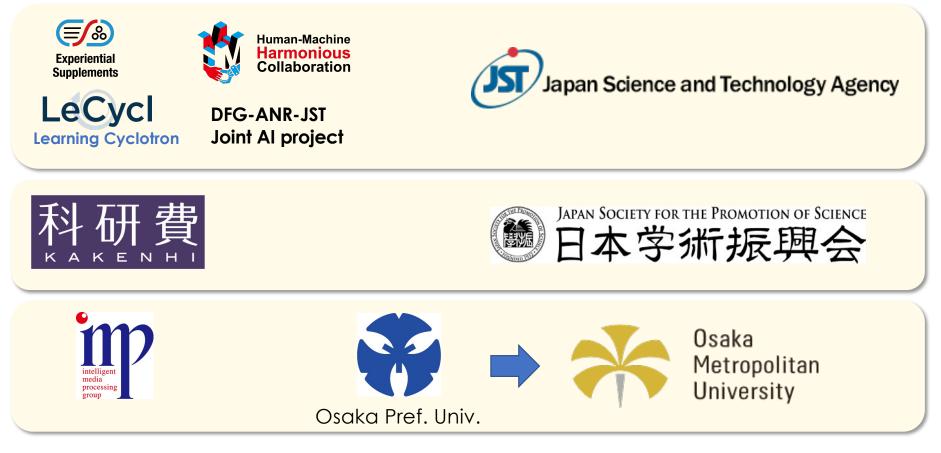








Acknowledgment



September 6, 2021

Goal of This Talk

- Not to give a comprehensive summary of SOTA
- Short overview of what we have done
- Think together future topics
- On-line anonymous questions: slido.com #777649





CBDAR



Document Analysis goes to the field.

You are a Collective of Every Experience

Jac Vanek

Abstract: You are the books you read, the films you watch, the music you listen to, the people you meet, the dreams you have, the conversations you engage in. You are what you take from these. You are the sound of the ocean, the breath of fresh air, the brightest light and the darkest corner. You are a collective of every experience you have had in your life. You are every single second of every single day. So drown

yourself in a sea of knowledge and existence. Let the words run through your veins and let the colors fill your mind until there is nothing left to do but explode. There are no wrong answers. Inspiration is everything. Sit back, relax, and take it all in. Now, go out and create something. reading.

People acquire a lot of information daily through reading. Thus by recording and analyzing reading activities, we are able to obtain various types of information about the reader including his / her interests and preferences. In addition, once the reading activities of many readers are recorded in relation to the contents of documents, we are able to consider the analysis of documents based on reading activities such as finding interesting and difficult parts for many readers. In other words we can open up a new research field of document analysis through the mutual analysis of reading behavior and document contents.

Our research on reading-life log has started with the motivation mentioned above. In our research the main source of information about the reading activities is the eye movement, because we think it contains rich information about reading activities. By analyzing eye movement, we can obtain various types of information about his/her reading. The simplest is the amount of reading: how many words are read by a reader in a specific time frame such as a day. In addition, the analysis of eye movement allows us to segment reading activities from others, to estimate a type of documents the reader is working on such as novels, magazines, newspapers and mangas. The contents of read documents can also be recorded as a log of

In this paper we are concerned with a higher level of readinglife log as compared to the previous approaches. As such a highlevel log we have selected the estimation of a level of language ability by analyzing reading activities. A simplest and traditional method for obtaining the language

abilities is based on paper tests. For the case of English for example, standardized tests such as TOEFL and TOEIC are well accepted for this purpose. However, such tests require time to take. Another problem would be that the test is available not so often, say once a month, and thus it is not possible to have a "real-time monitoring of the ability. Our research attempts to build a method for easy and real-time monitoring of the language ability with a special focus on English.

The proposed method analyzes eye movement obtained through an eve tracker to learn/classify



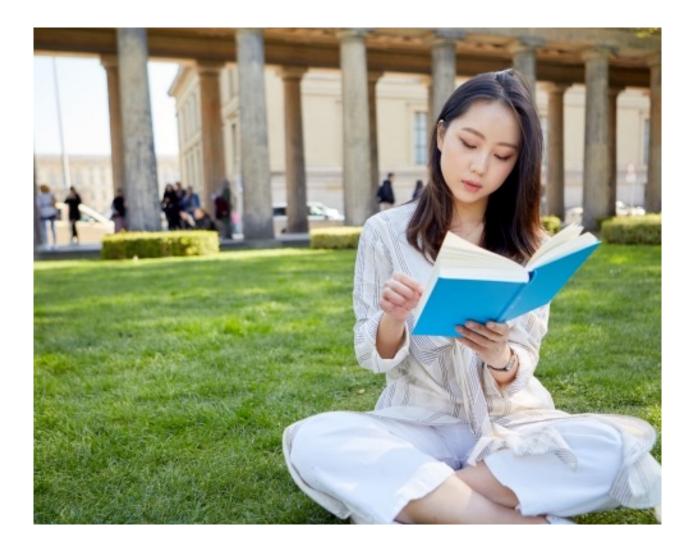
CBDAR2021 Keynote



^{1.} Introduction

We are doing CBDAR

by using our eyes as a camera

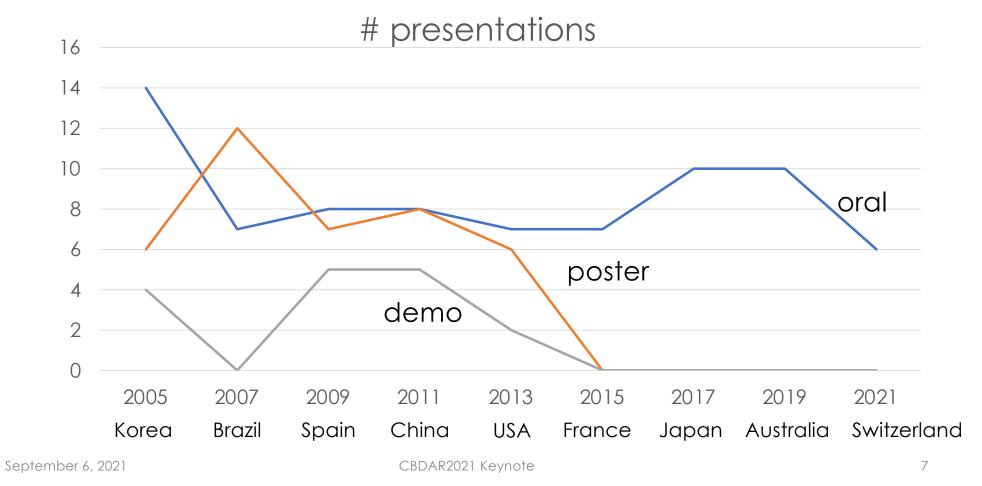


September 6, 2021



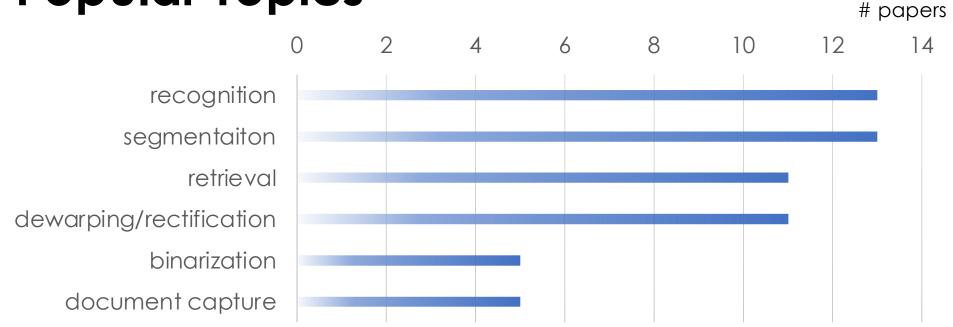


History & Stats





Popular Topics



DNN based methods started to appear from 2017

No more low-level processing and retrieval in recent years

September 6, 2021





Basic Questions

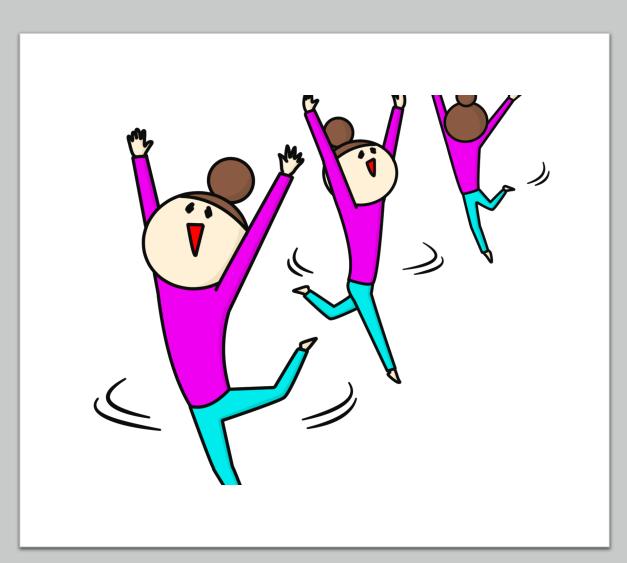
- Are you satisfied by current research topics and status?
 - end-to-end
 - accuracy continuously improved
- We have started the FDAR workshop since 2017 @Kyoto
 - Important question given by Prof. Uchida

What if you achieve 100% accuracy?

September 6, 2021

CBDAR2021 Keynote

10

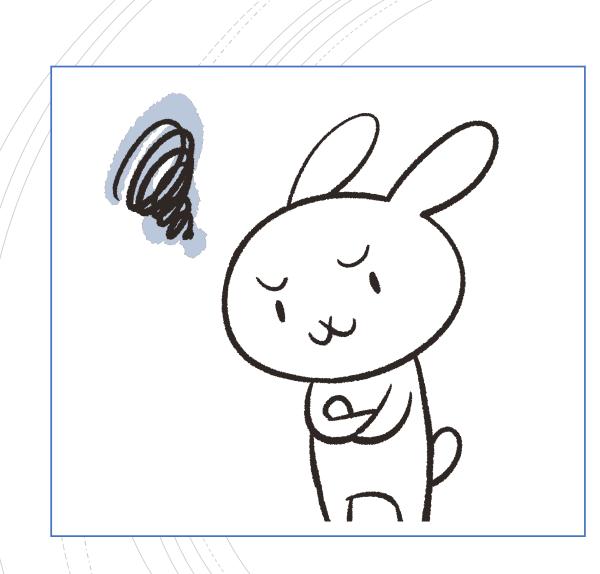




This has been the ultimate goal.

All problems can be solved.

September 6, 2021

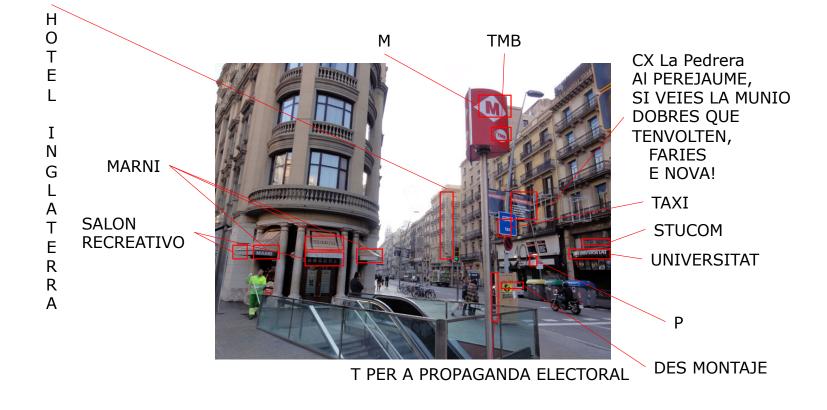


Nothing to do anymore.

September 6, 2021













Just telling the text around you...

Especially for the case of blind people

hotel ingratella, M, TMB, CX La Pedrera Al PEREJAUME, \cdots





Masa Iwamura and his students have been working on this topic.

September 6, 2021



What is necessary?



inteller provide and a second

What to recognize depends on:

Internal states of the user

Environment

Time



What we have done / What we should do

- What we have done:
 - General purpose "recognition"
 - Without thinking the use of results
- What we should do:
 - Focusing on use cases
 - Thinking more about the user

Experiential Supplements



Various Roles of Scene Texts

- Before discussing what to do, let's take a look at various roles of scene texts
- Their role is not just labeling

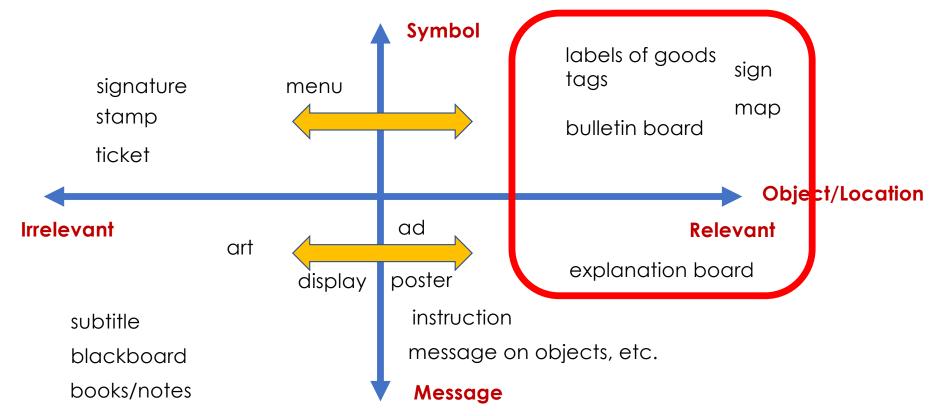


September 6, 2021











Labels of goods, tags





distinction

September 6, 2021

CBDAR2021 Keynote

19



Labels of goods, tags









September 6, 2021

Labels of goods, tags: exceptional cases

Labels detached from objects

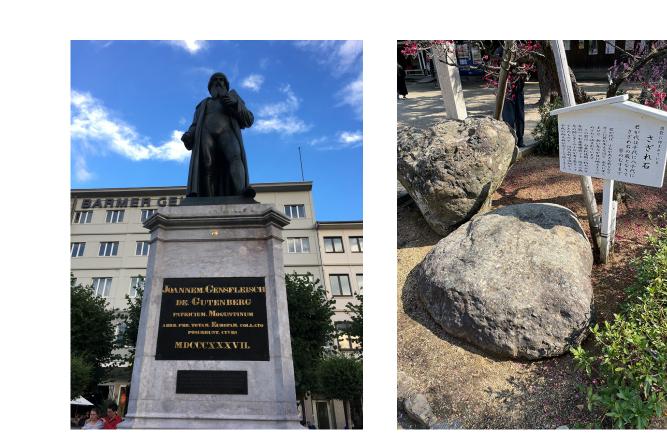
Incorrect labels



September 6, 2021









September 6, 2021





Signs and Maps



September 6, 2021

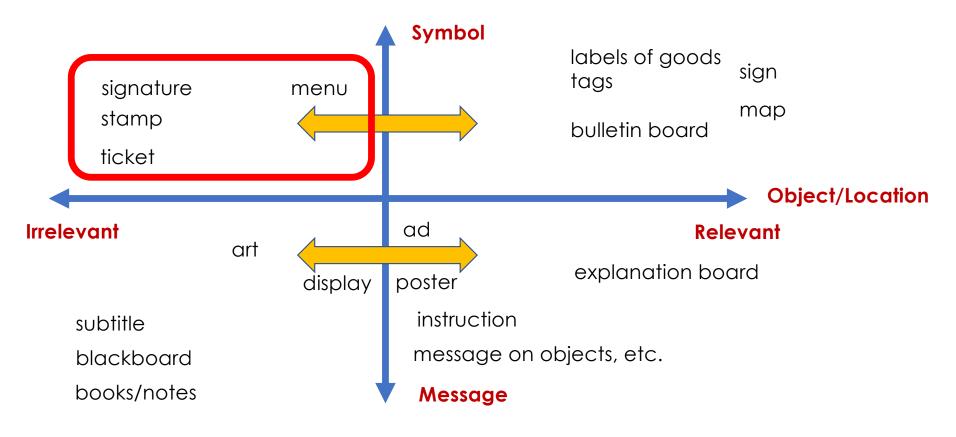








Spectrum





Menu

Location irrelevant

TEA SELECTION

Our selection of the finest whole leaf tea takes you on a voyage of discovery across the globe.

International Tea*



Founded by celebrated tea innovator Taha Bouqdib, TWG Tea is the finest luxury tea brand in the world, offering an unsurpassed tea list of over 800 different exclusive blends and fine barvest teas from every

1837 Black Tea

TWG Tea's renowned timeless classic. This is a unique blend of black tea with fruity and floral tones which leaves a lingering taste of ripe berries and caramel.

Silver Moon Tea

A green tea blend accented with grand berry and vanilla bouquet. It is a tea for that special moment, with richness in antioxidants and nutrients.

Vanilla Bourbon Tea

A theine-free red tea from South Africa, blended with sweet vanilla. With its high antioxidants, vitamin C, mineral salts and proteins, the tea is suitable for both adults and children.

Royal Darjeeling

An exquisite first flush black tea plucked during Spring harvests and is suitable for daytime consumption. This tea has a vibrant sparkling taste that develops remarkable overtones of

Location relevant



September 6, 2021

Signatures, Stamps, and Tickets



September 6, 2021

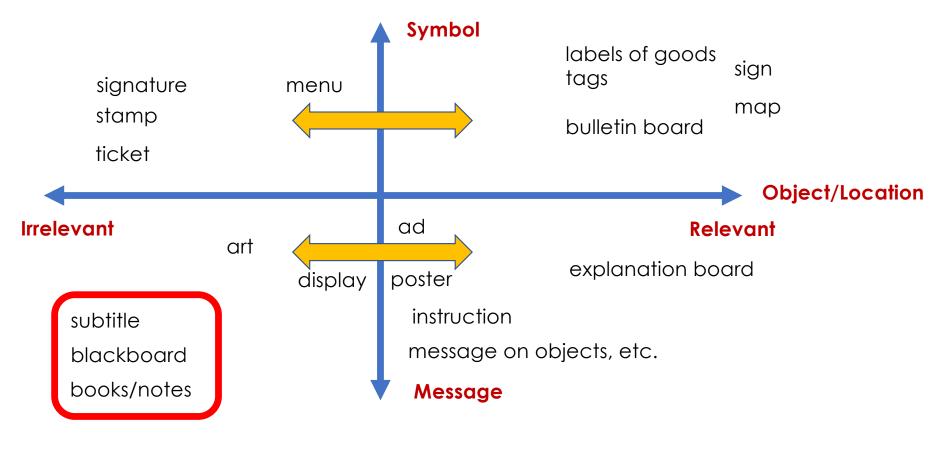
CBDAR2021 Keynote

28



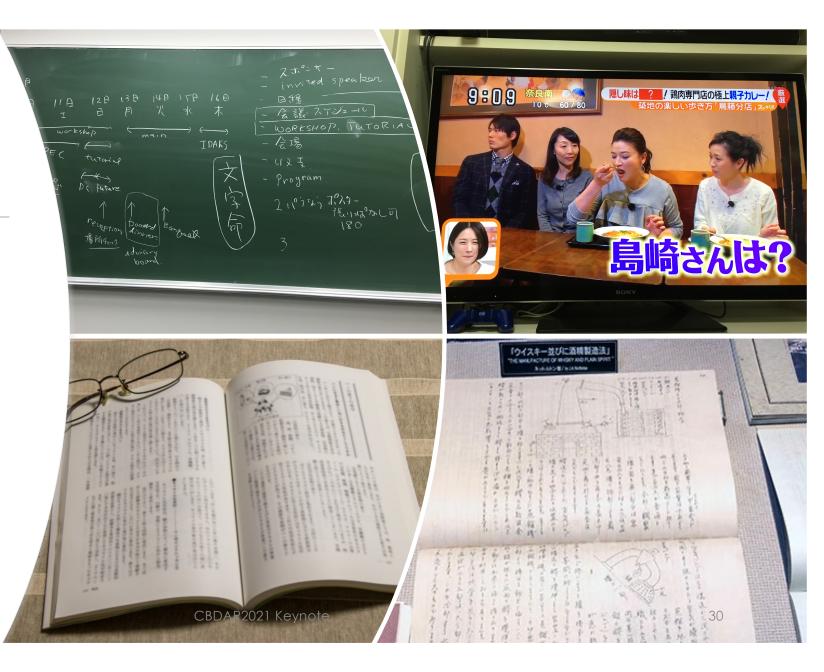






Subtitles, Blackboards, Books and Notes

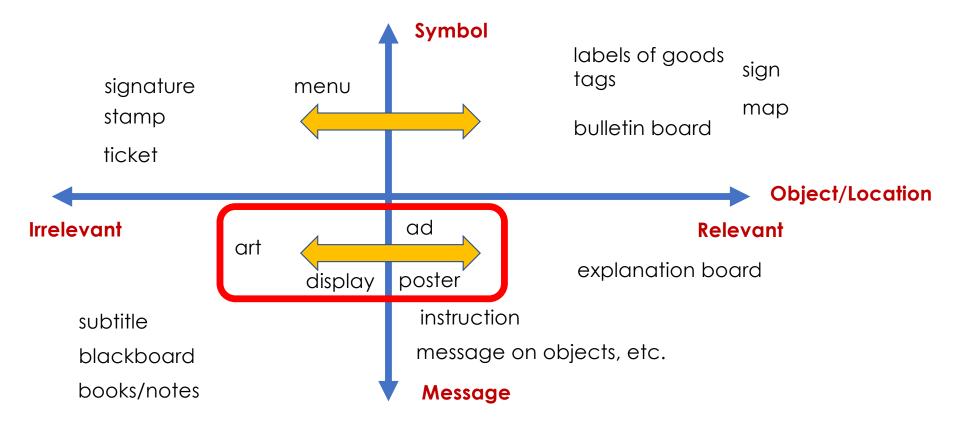
 These have been objects for recognition before CBDAR













Ads and Posters

September 6, 2021

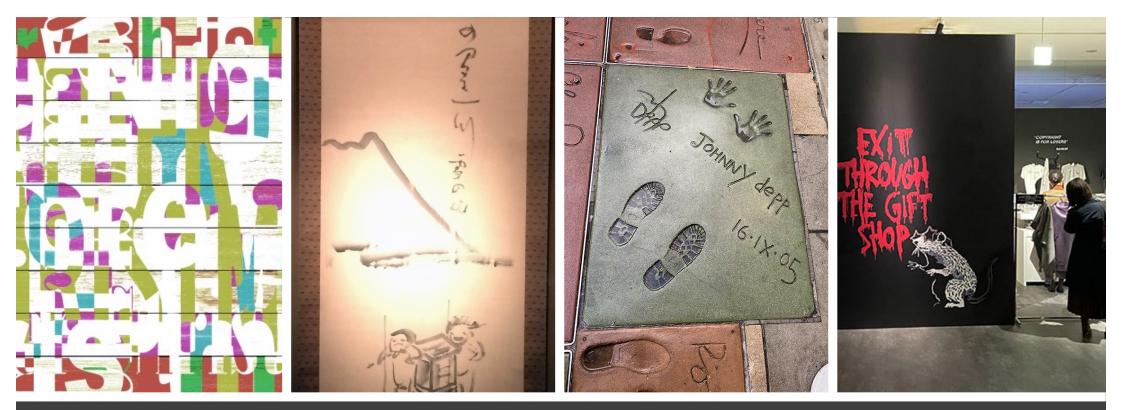


Display

• Location relevant and irrelevant



September 6, 2021



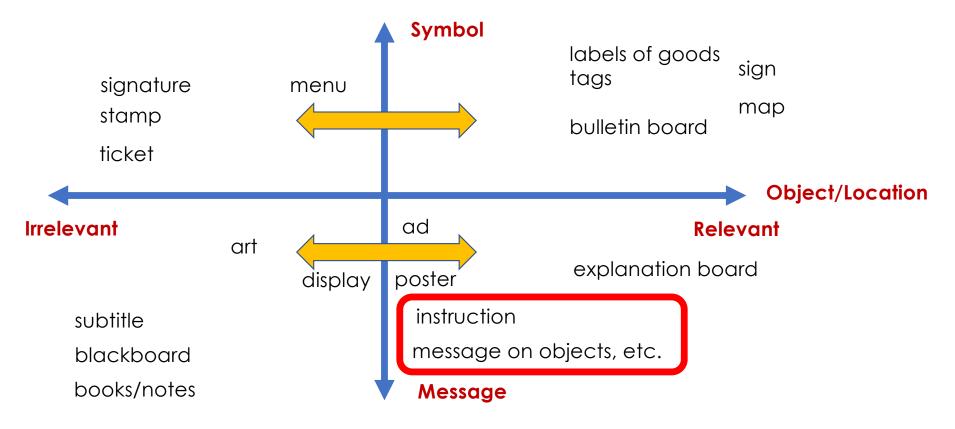
Arts

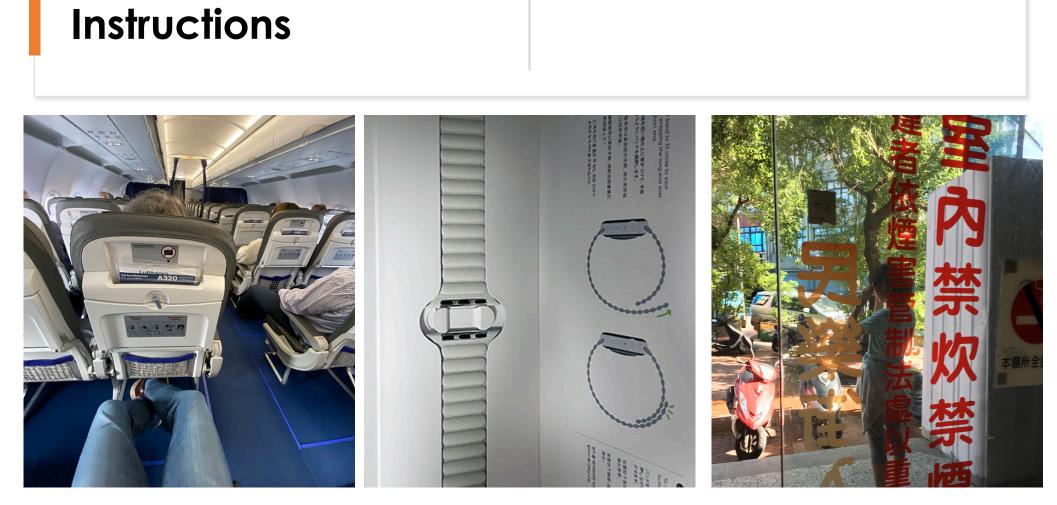
September 6, 2021











September 6, 2021







Messages on Objects



September 6, 2021

Intentions of Messages



September 6, 2021

Religious Messages





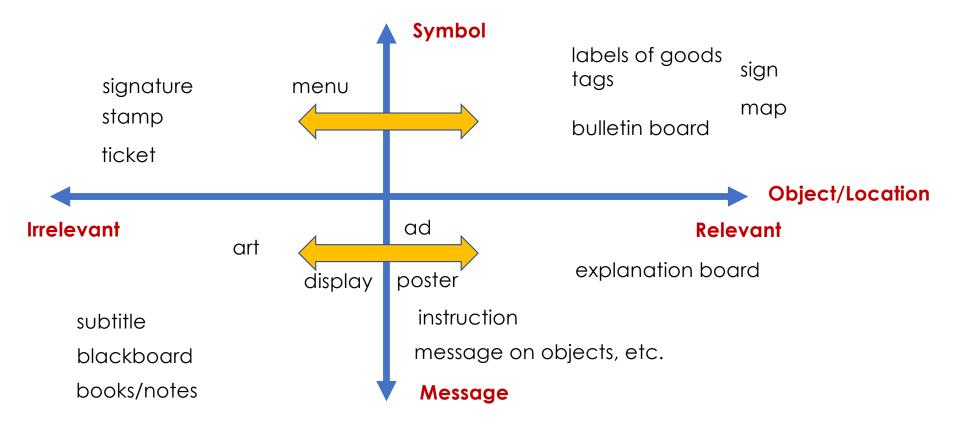


September 6, 2021













Points to Consider

- object/location relevant?
 - yes: label, tag, explanation of an object, or a location
 - no: books, notes, signatures, stamps, tickets, etc.
 - depends: bulletin board, menu, instruction, etc.
 - exceptional cases: detached label
- role of the scene text
 - typical roles: label, explanation, message
 - exceptional cases: design, art, ad, religion



What we have done / What we should do

- What we have done:
 - General purpose "recognition"
 - Without thinking the use of results
- What we should do:
 - Focusing on use cases
 - Thinking more about the user

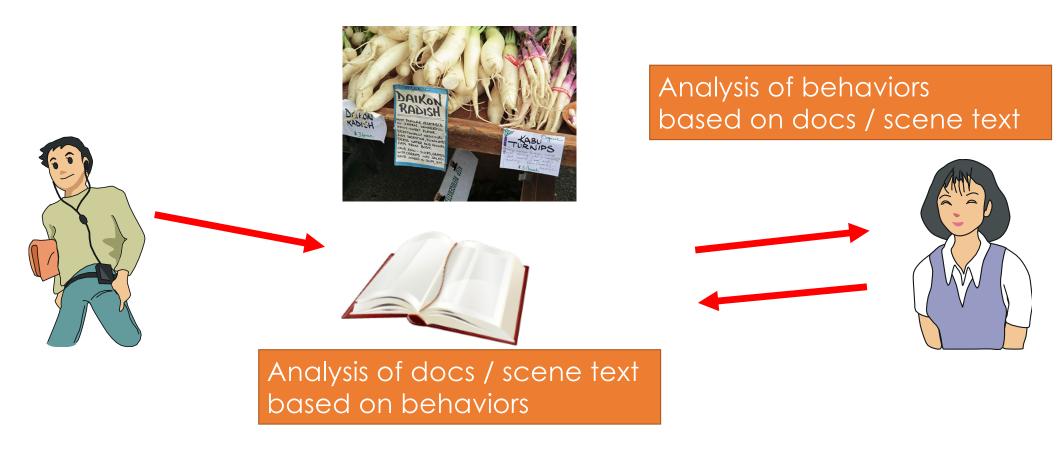
Human-Document Interaction

Experiential Supplements





Human-Document Interaction



September 6, 2021

Explanation of Objects

Analysis of **scene text** based on behaviors

- Audience rating
- Effectiveness

Analysis of **behaviors** based on scene text

• Buying with/without reading



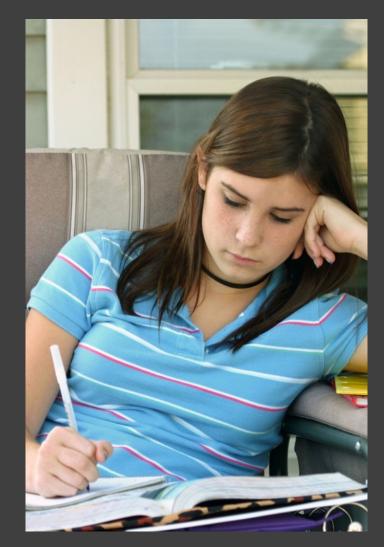
Acquiring knowledge

Analysis of **docs** based on behaviors

- Difficult parts to understand
- Interests

Analysis of **behaviors** based on docs

- what have been learnt
- level of understanding







What we know thru behaviors

- role of scene texts, intention
- property of scene texts (or objects)
 - interest
 - preference
 - difficulty
 - audience rating
- effectiveness to behavior change





What we know thru documents/scene texts

- property of humans
 - interest
 - preference
 - level of knowledge
- influence on his/her behavior





How to Implement

- Timing
 - real-time?
- Viewpoint
 - where to put the camera?





Where to put the camera?



3rd person

2nd person

1st person

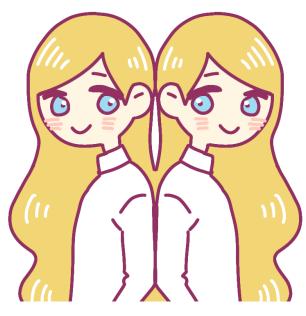
Applications of CBDAR

Personal Life Record

Digital Twin

Digital Transformation







September 6, 2021





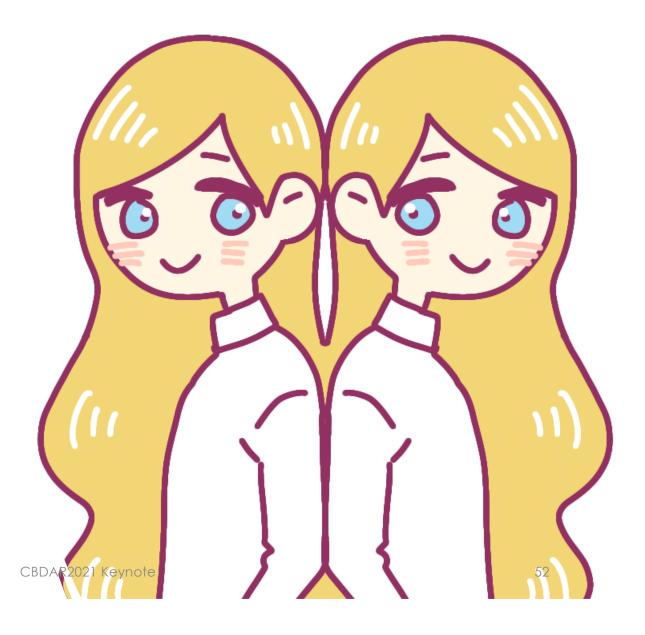
Personal Life Record

- Can be partially done on the digital world
- Full coverage requires interactions with the physical world
- CBDAR to capture such interactions

September 6, 2021

Digital Twin

- Human digital twin can be built upon the recognition of human properties
- CBDAR to capture such properties



September 6, 2021

Digital Transformation

- DX requires deeper analysis of business processes
- Human-document interactions gives us insights







Summary

- Current Status of CBDAR
 - Recognition of scene text
 - With a 3rd person camera
 - Without purposes
- Future CBDAR
 - Based on deeper understanding of the role of scene texts
 - With a camera on the appropriate viewpoint $(1^{st} 3^{rd})$
- Important Applications are waiting for your research