FUJITSU

shaping tomorrow with you

Video Retrieval

Yoshinobu Hotta
FUJITSU LABORATORIES LTD.

Copyright 2012 FUJITSU LABORATORIES LTD.

Who am I ?

FUJITSU

- My research areas : Image processing
  - Document image processing
    - Japanese OCR (Optical character recognition)
    - Text extraction from natural scene images
  - Video analysis ← today's topics
  - Gesture user interface

Text extraction

Gesture user interface

Copyright 2012 FUJITSU LABORATORIES LTD.

Outline

FUJITSU

0. Introduction
1. Caption extraction and recognition
2. Digest video generation
3. Visual user interface
4. Summary

Copyright 2012 FUJITSU LABORATORIES LTD.

0. Introduction

FUJITSU

- Importance of video retrieval

  “By 2013, more than 25 percent of the content that workers see in a day will be dominated by pictures, video or audio” (Gartner 2008)

Current status :

- Many video portals : YouTube, Google videos, Veoh..
- Many legacy videos : TV programs, Surveillance videos, E-learning videos, ..
- Wide spread of smart phones

⇨ Video retrieval is needed in various scenes

Copyright 2012 FUJITSU LABORATORIES LTD.

Applications in a wide variety of businesses

FUJITSU

Broadcast

Editing

Broadcast

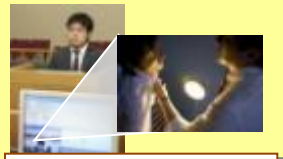
Manufacturing

Maintenance manual

Weld verification

Security

Accident inspection



Interrogation video

Healthcare

Surgery video

Informed consent

Video analysis/ retrieval

Delivery

Share

Production

Surveillance

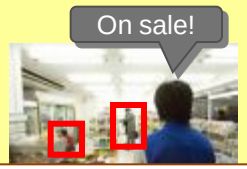
Agriculture

Growth observation

Damage control

Retail distribution

Shoplifting



Announcement effect

Education

Educational videos

Museum signage

Public service

Proceedings video

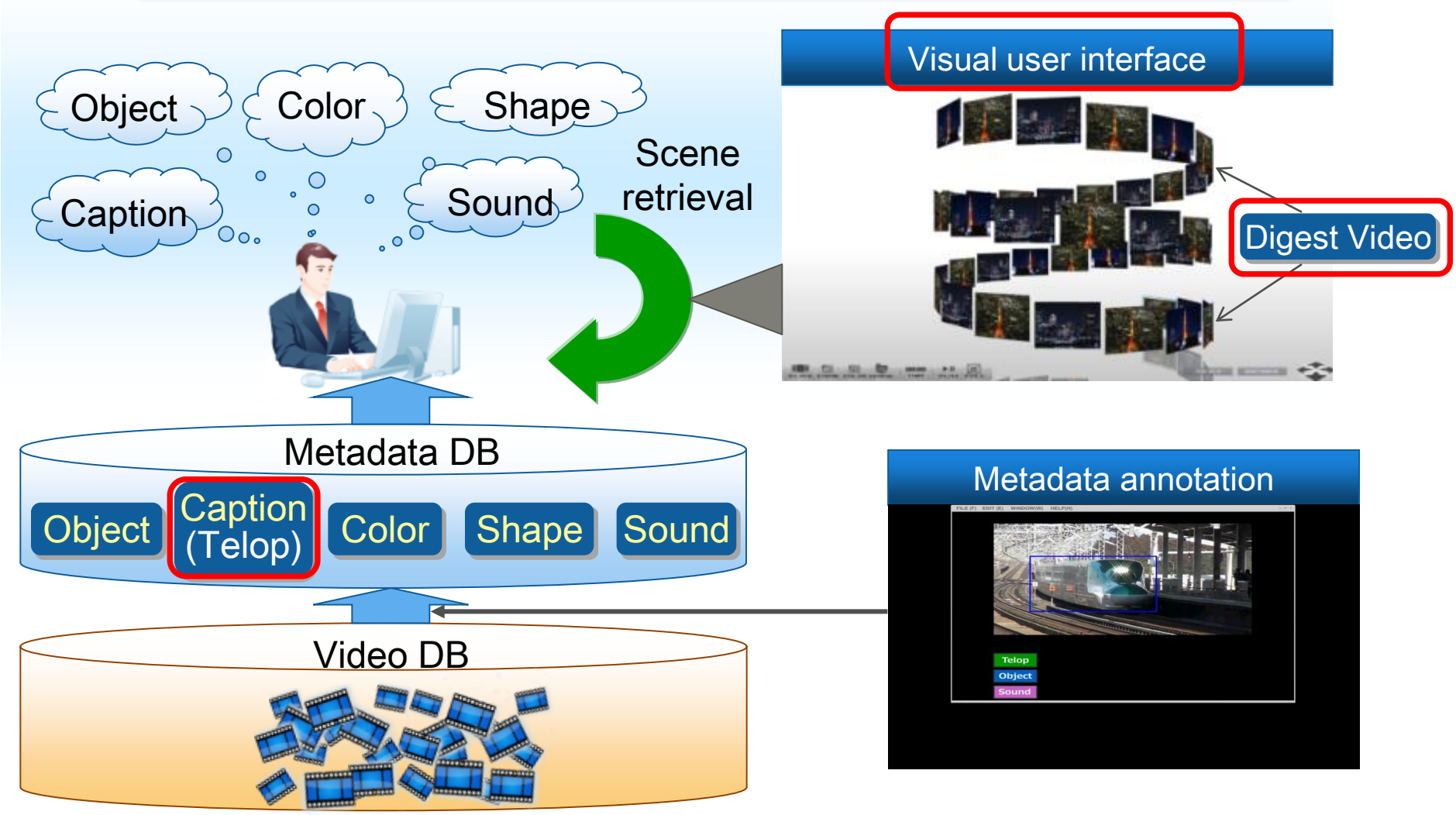
Illegal dumping

Copyright 2012 FUJITSU LABORATORIES LTD.

Video Retrieval System

FUJITSU

Scene retrieval from various aspects (metadata) such as object, caption, color, shape, and sound



Copyright 2012 FUJITSU LABORATORIES LTD.

1. Caption extraction and recognition

FUJITSU

- What is a caption ?
  - ➔ Character strings which appears in a frame

Video Database

:

ビストロ | Search

Retrieved result

Copyright 2012 FUJITSU LABORATORIES LTD.

Our target

FUJITSU

- Various types of captions used in Japanese TV program

	Uniform character color	Multiple character color
Uniform Background		
Complicated background		

Method A, B

Method C

Target of our method

Exception1 : movable caption
(ex. End credits of movies)

Exception2 : Gradation

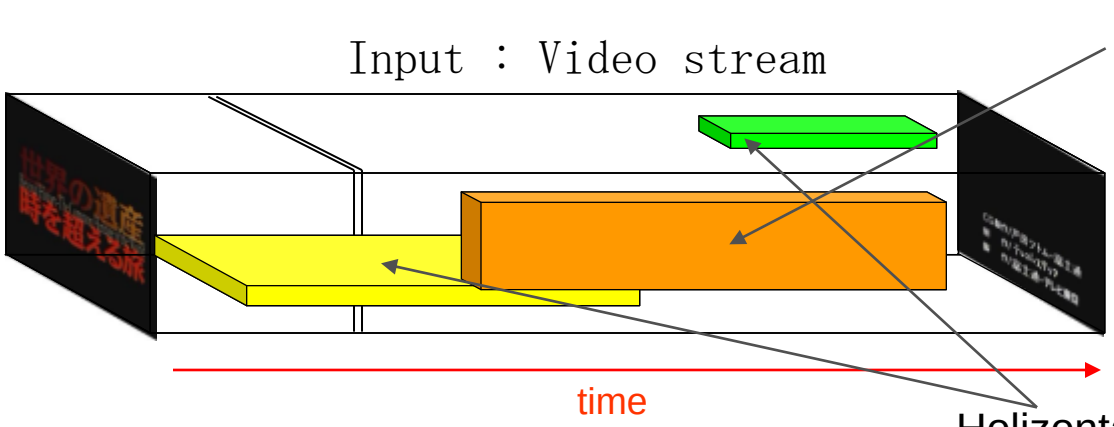
Copyright 2012 FUJITSU LABORATORIES LTD.

Outline of caption extraction / recognition

FUJITSU

Caption extraction

Extract caption region candidates from static part of the video



Caption recognition

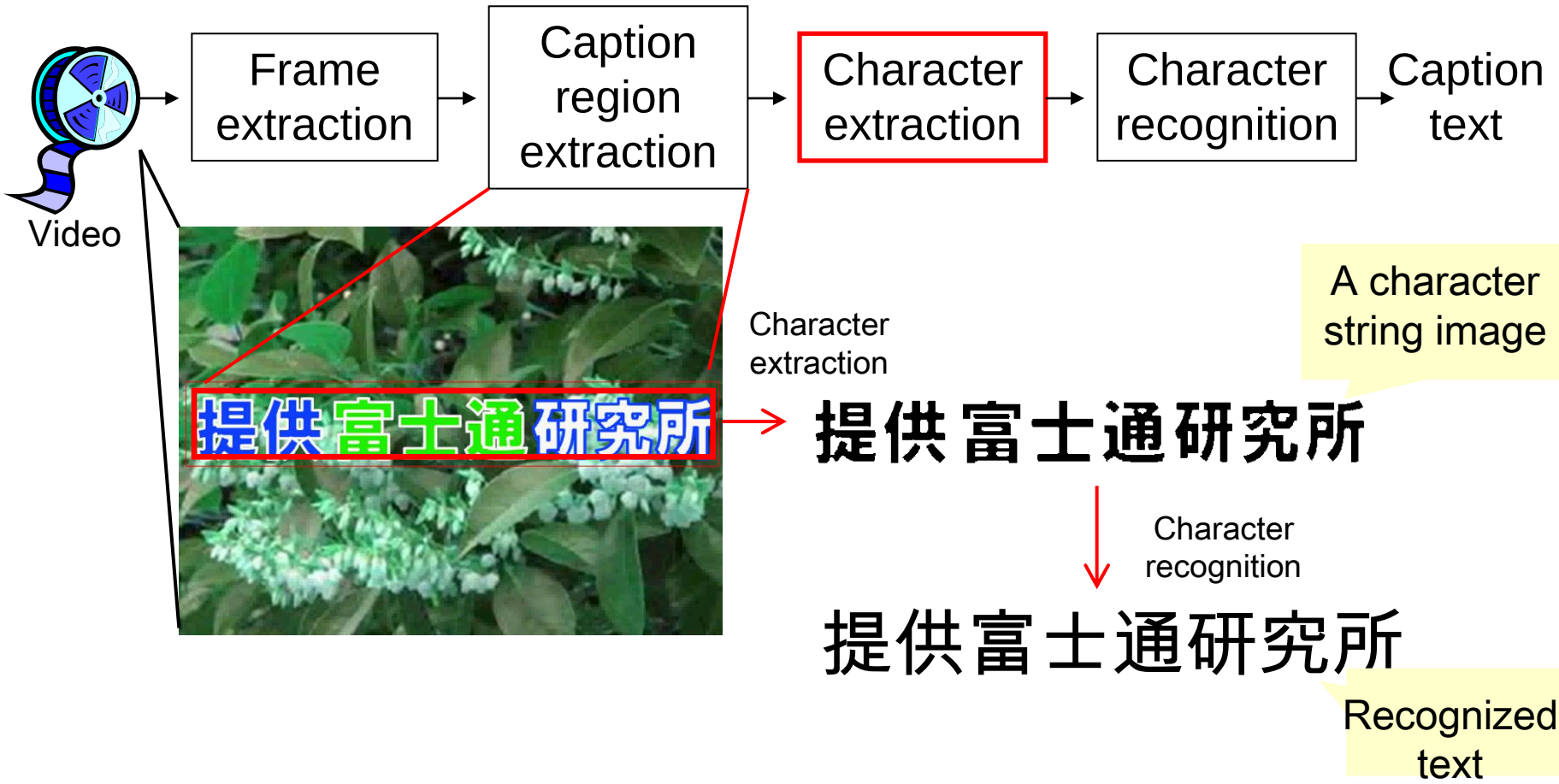
Recognize characters inside caption region

→「レトロ車両」

(utilized as video index)

Copyright 2012 FUJITSU LIMITED

Whole flow



It is important to extract character string from complicated background

Copyright 2012 FUJITSU LABORATORIES LTD.

Problems of conventional method

FUJITSU

(1)A method which utilizes uniform character color

- Effective for single colored caption
- Not effective for multi-colored caption

(2)A method which utilizes uniform background color

- Not effective in complicated background

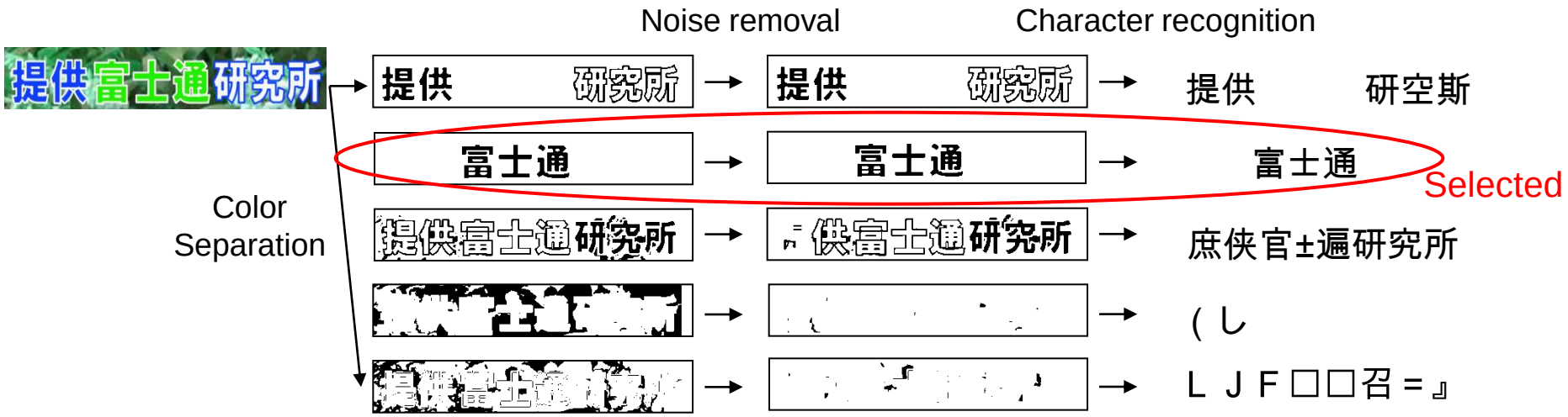
Character extraction with multiple-color from complicated background is not easy

Copyright 2012 FUJITSU LABORATORIES LTD.

Problems of conventional method (2)

FUJITSU

(3) A method which utilizes color separation and select single color plane

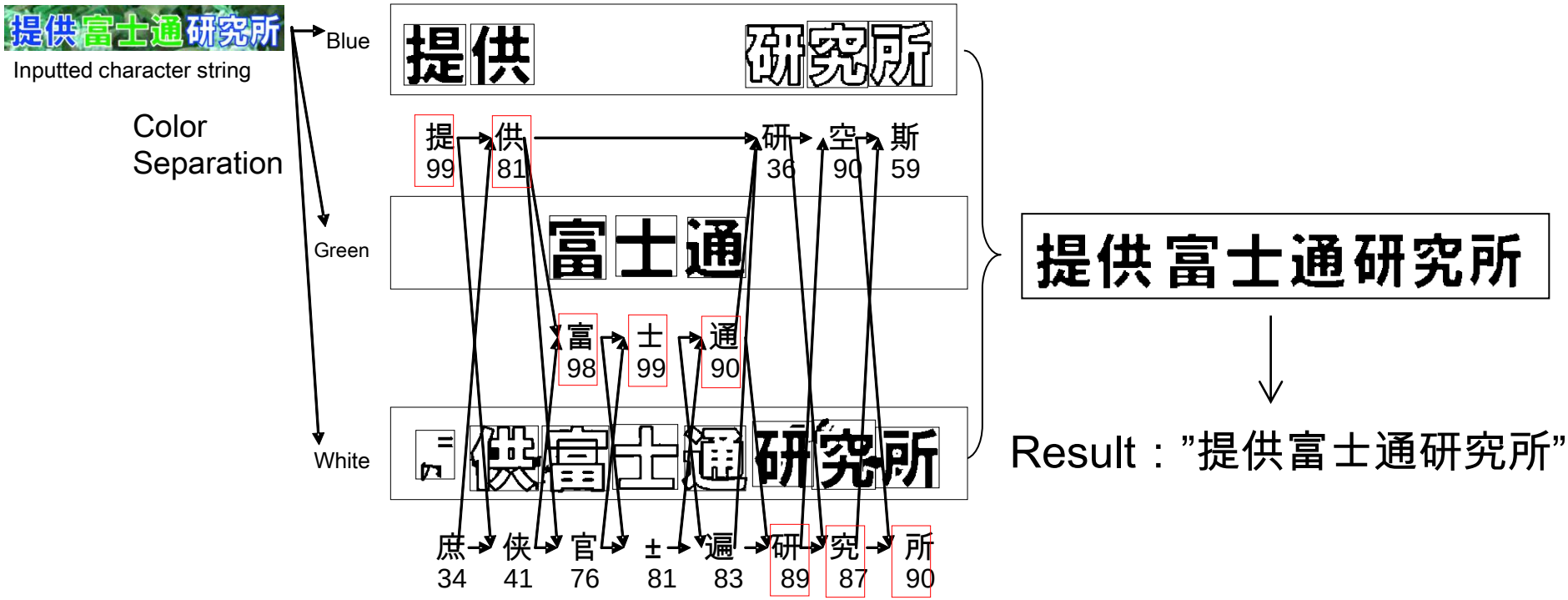


A recognition result of single color plane include only part of correct texts.

Copyright 2012 FUJITSU LABORATORIES LTD.

Multi-colored caption recognition (proposed)

- Decompose an input color image and extract character region candidate
- Recognition reliability and continuity of rectangles and colors are evaluated by Dynamic programming



Copyright 2012 FUJITSU LABORATORIES LTD.

Recognition result

Result： “気になる費用は”

Result： “問題ここはどこ”

Copyright 2012 FUJITSU LABORATORIES LTD.

2. Digest Video Generation

FUJITSU

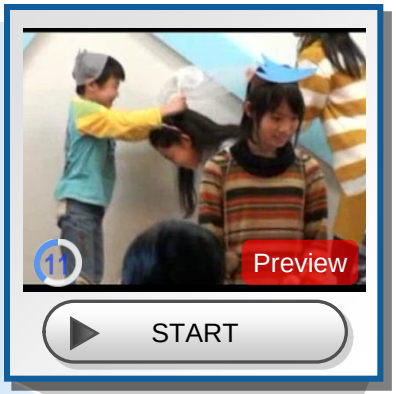
Automatic generation of digest videos from TV program, movie, surveillance / homemade videos etc.

Application

Sports program

Homemade videos

Digest Video Generation



Copyright 2012 FUJITSU LABORATORIES LTD.

Scene analysis for digest video generation

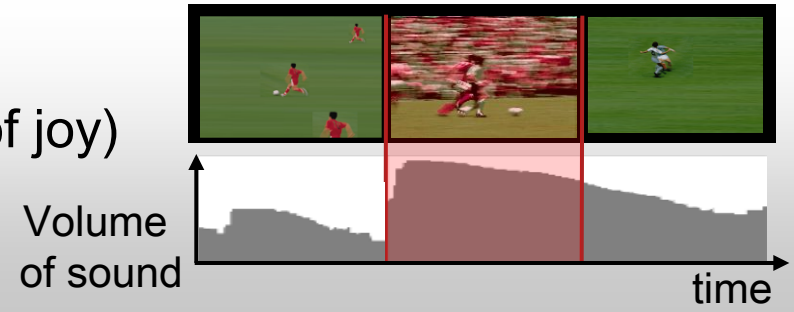
FUJITSU

- Specific scenes in a video are detected by analyzing video/sound

Examples of scene analysis

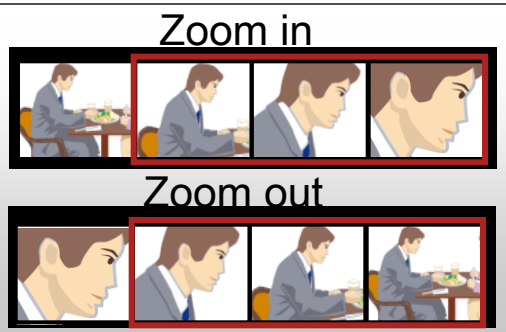
Highlight detection

Analyze the highlight of sound(a shout of joy) of sports program and detect a score scene



Camera motion estimation

Analyze camera motion such as zoom-in, zoom-out and hand-shake in video and then detect the motion change scene



Speech/Sound detection

Analyze sound features and detect a conversation scene or music scene

Detect only song scene from music show

Copyright 2012 FUJITSU LABORATORIES LTD.

How to utilize sound feature

FUJITSU

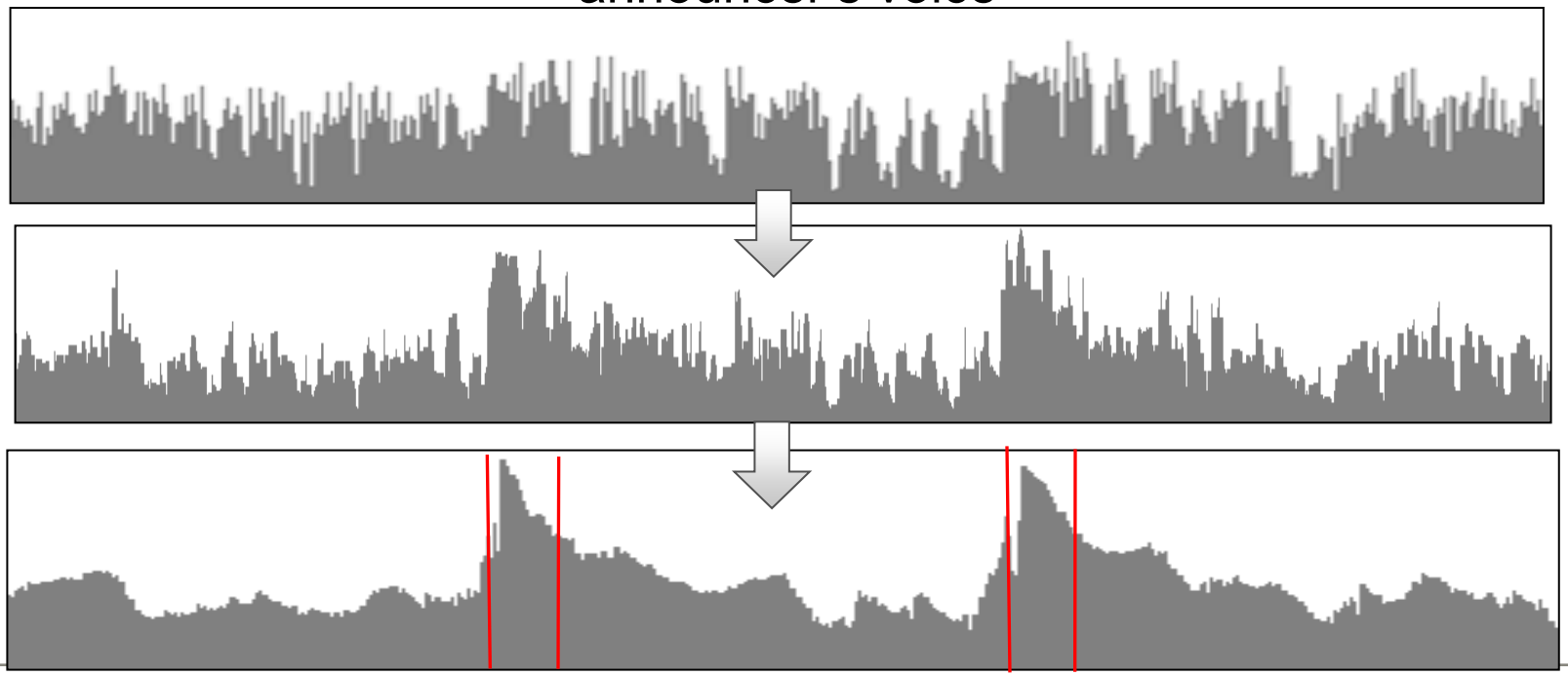
Original video

Ideal detection

Features

Volume of sound + Removal of announcer's voice + Sudden sound change

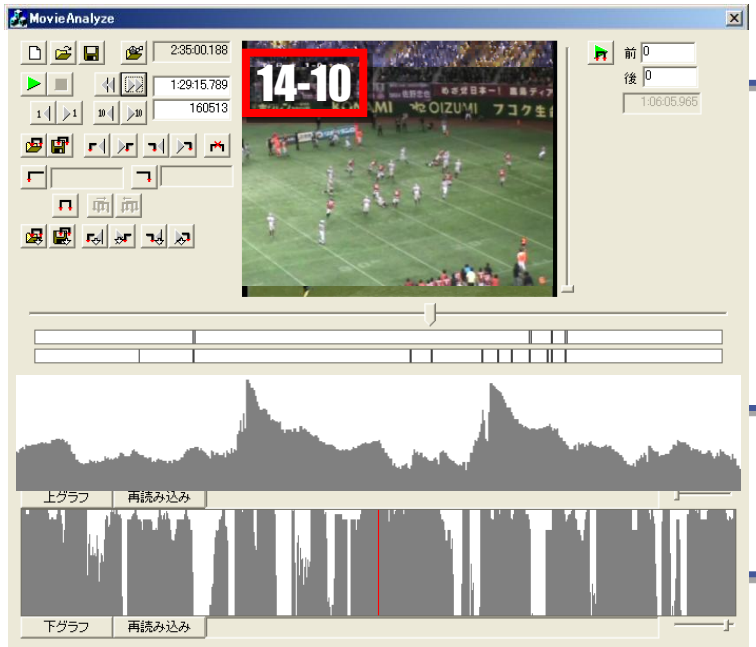
Extracted value

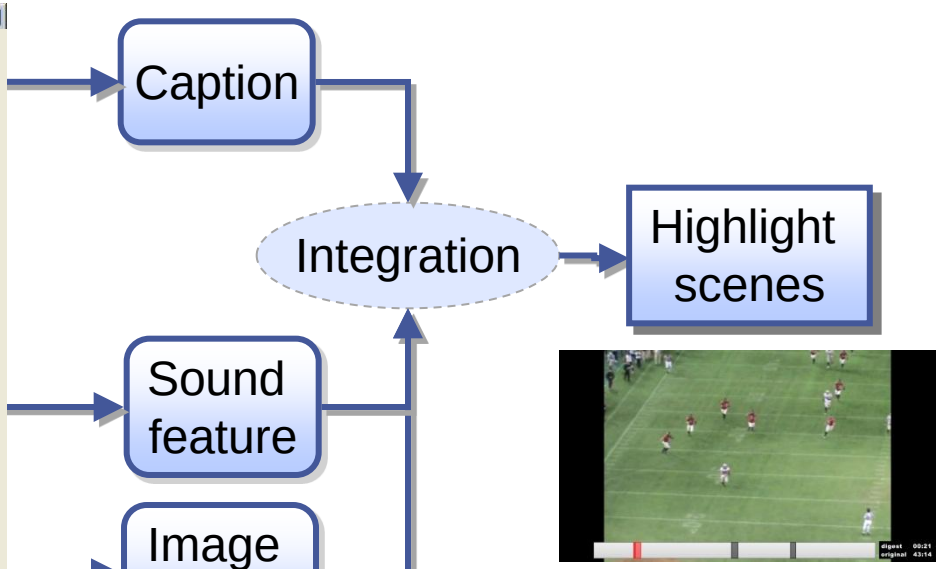


Copyright 2012 FUJITSU LABORATORIES LTD.

Example1: Sports video retrieval

- Extracting various kinds of scenes for digest video creation
  - Caption features for score, sound features for a shout of joy , image features for ground color are integrated.





Highlight scene extraction

Copyright 2012 FUJITSU LABORATORIES LTD.

Example2: Illegal dumping scene retrieval

- What is “Illegal dumping” ?
  → Someone illegally dumps waste.
- Illegal dumping is a social issue.
- Officials in local governments check surveillance videos to retrieve illegal dumping scenes.

Image feature (Frame difference for human detection)

Image feature (Frame difference for waste detection

Integration

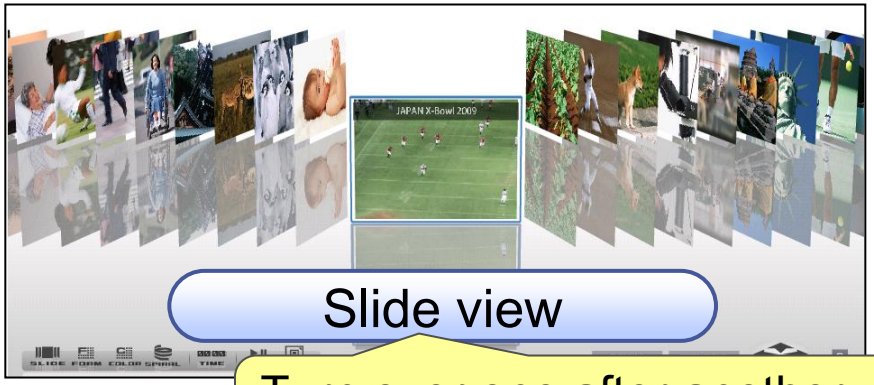
Highlight scenes

Copyright 2012 FUJITSU LABORATORIES LTD.

3. Visual User Interface

FUJITSU

- Easier video search with various GUIs

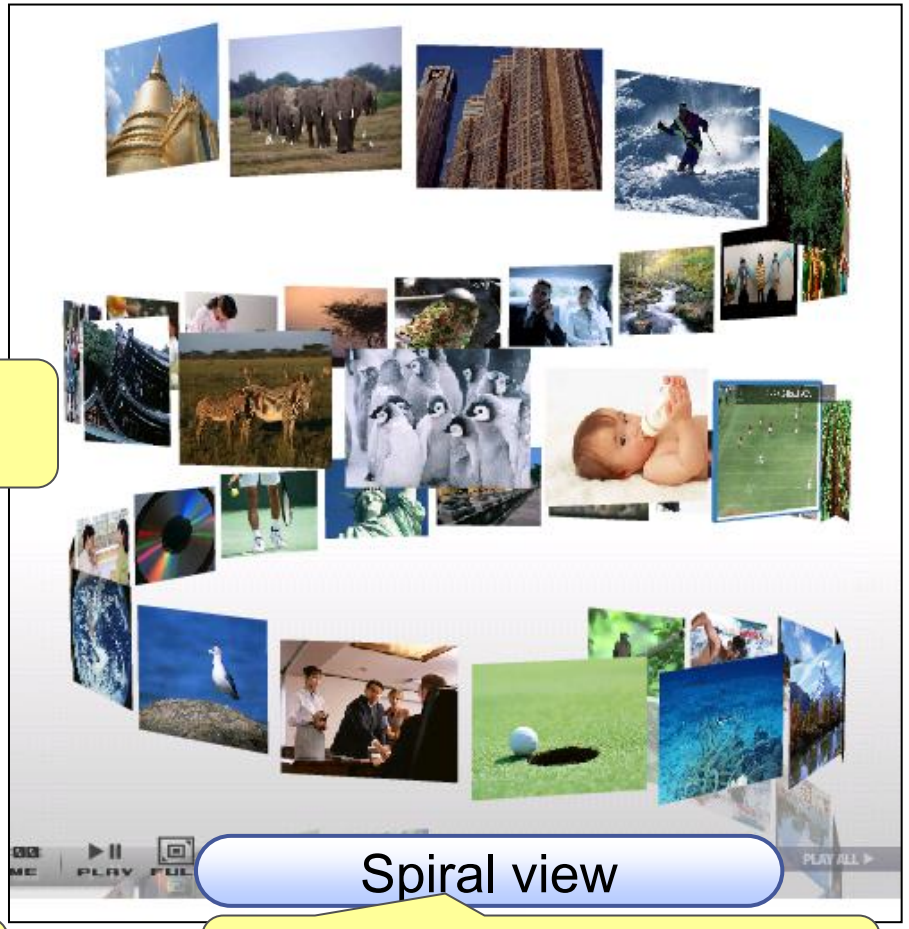


Slide view

Turn over one after another like the shuffle of cards

Color/shape based view

Arrange similar color or shape images closely



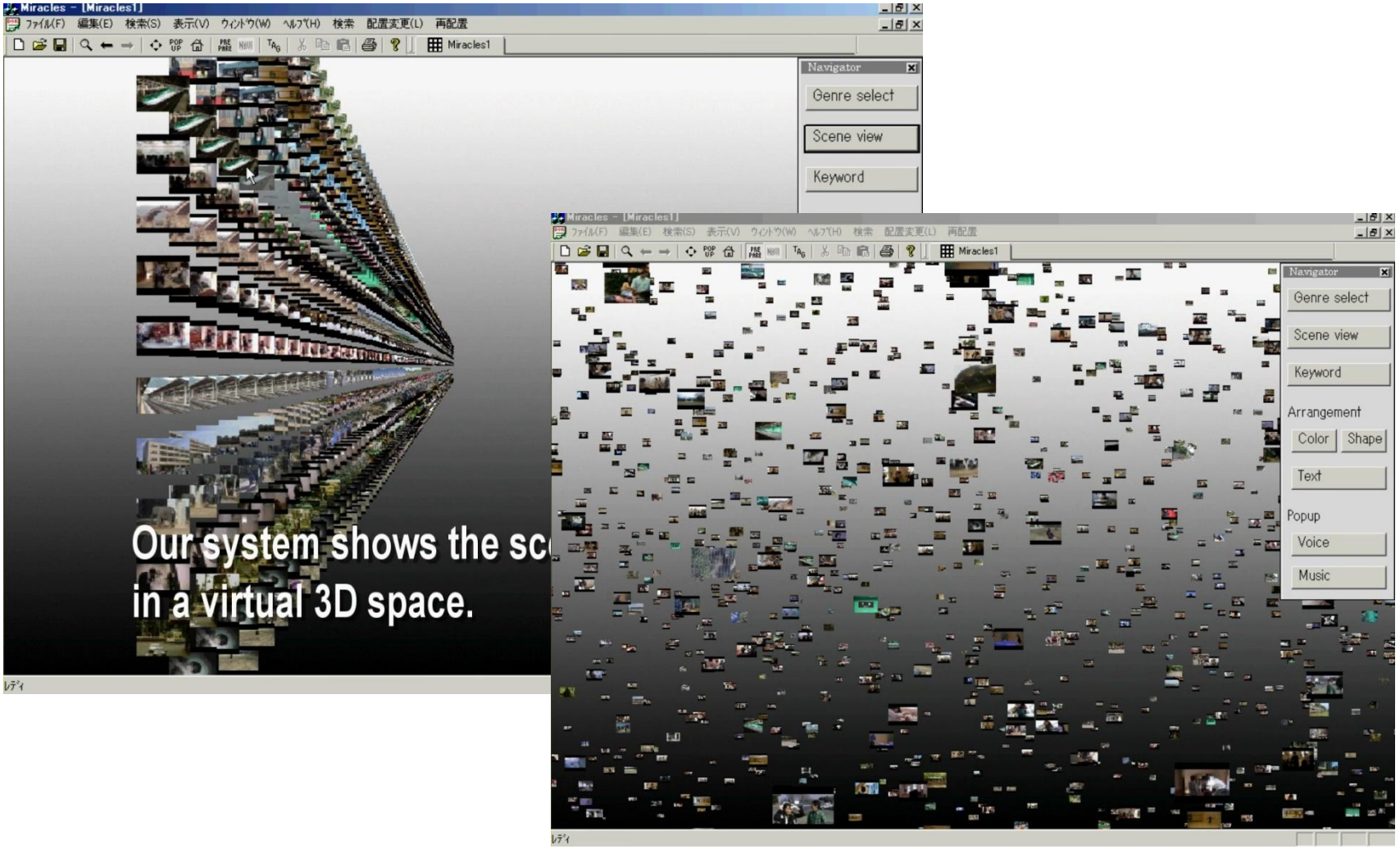
Spiral view

Simultaneous reproduction of images on a thumbnail

Copyright 2012 FUJITSU LABORATORIES LTD.

Demo of visual user interface

- Reduce retrieval cost by combination of keyword search and visual user interface



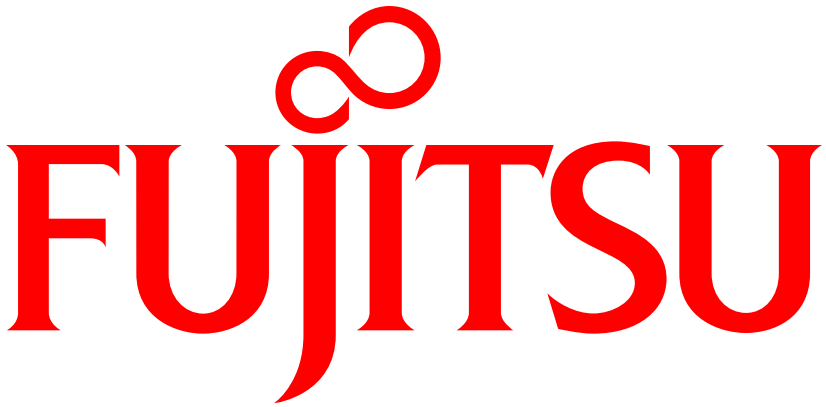
Copyright 2012 FUJITSU LABORATORIES LTD.

4. Summary

FUJITSU

- Several video retrieval technologies were introduced.
  - Caption extraction and recognition
  - Digest video generation
  - Visual user interface
- In next 5-10 years
  - Various types of metadata such as peoples name, object name (ex. Landmarks) can be added with progress of object recognition
  - Web information are dynamically linked with contents of video
    ex. Peoples face->Facebook,
    Landmark -> Wikipedia
    Restaurant -> Reputation site..

Copyright 2012 FUJITSU LABORATORIES LTD.



shaping tomorrow with you