NII-KAORI™
A Semantic Video Search Engine

Duy-Dinh LE, Thanh Duc NGO, and Shin'ichi SATOH
National Institute of Informatics

Motivation

❖ Current video retrieval systems rely on meta data (e.g. genre, date, producer, etc) and user-generated tags.
❖ Advanced video search engines should rely on the actual content of the video data such as objects, motions, people and events.
❖ Building such systems requires to solve the problem of bridging the semantic gap.

Proposal

❖ We introduce a semantic video search engine for content-based video retrieval.
❖ The video content information is automatically extracted using semantic video analysis.
❖ The demo enables users to search and explore the content of large video archives with several thousand hours of video.

Feature Highlights

1. Large and diverse video datasets

❖ TRECVID 2004-2008: news and documentary programs in languages such as English, Arabic, Chinese and Dutch.
❖ NHKNews7 2001-2009: news in Japanese.

2. Automatic semantic concept extraction

❖ Airplane, Building, Animal, Sitting, etc.

3. People search

❖ Search by names.

4. Region search

5. Support both targeted search and exploratory search

❖ Search by keywords and names.
❖ Browse by similar images.

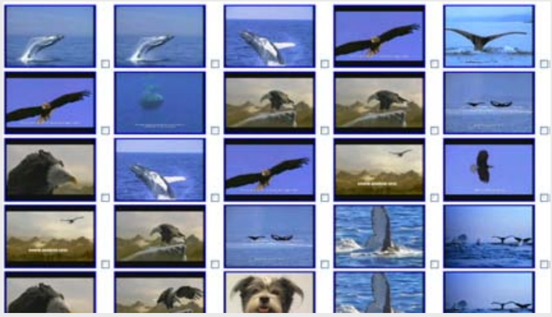
Technical Details

1. Video datasets

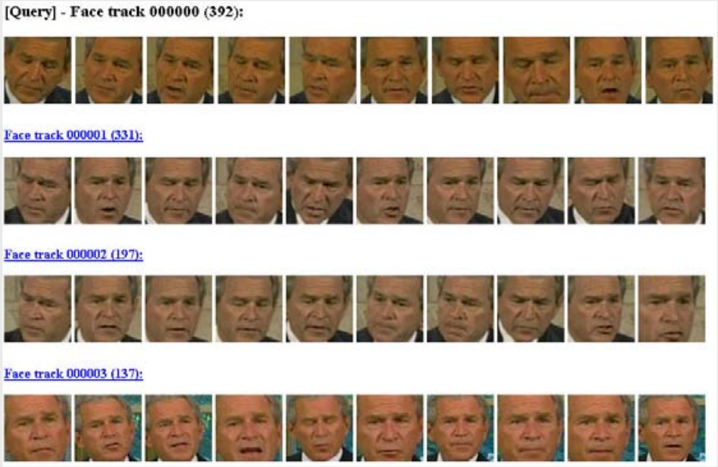
❖ TRECVID 2004-2008: 658 hours, 659,322 keyframes.
❖ NHKNews7 2001-2009: 1,413 hours, 747,529 keyframes.
❖ JP News and Documentary Video, 2005-2009: 11,335 hours, 680,106 keyframes.
❖ Total: 13,406 hours, 2,086,957 keyframes.

2. Semantic concepts

❖ LSCOM (http://www.lscom.org/) : 374 concepts.



Query keyword: animal



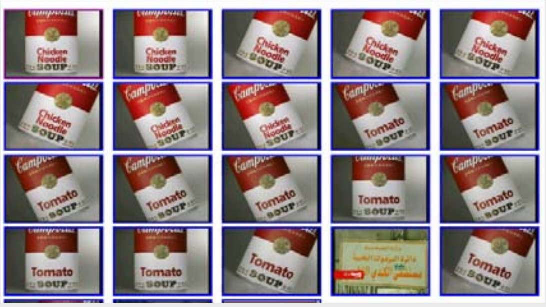
[Query] - Face track 000000 (392):

Face track 000001 (331):

Face track 000002 (197):

Face track 000003 (137):

Query name: Bush



Query region: "Chicken Noodle Soup"

Future Goals

1. Core Services for Video Analysis

❖ *Sharing and Collaboration.*

2. Applications for Knowledge Discovery

❖ *Video recommendation, video mining, social analysis.*

NII Duy-Dinh LE, Thanh Duc NGO, and Shin'ichi SATOH, National Institute of Informatics
TEL : 03-4212-2583 FAX : 03-3556-1916 Email : {ledduy,ndthanh,satoh}@nii.ac.jp

Web Image Search in TV Video Archive

Sebastien POULLOT and Shin’ichi SATOH

National Institute of Informatics

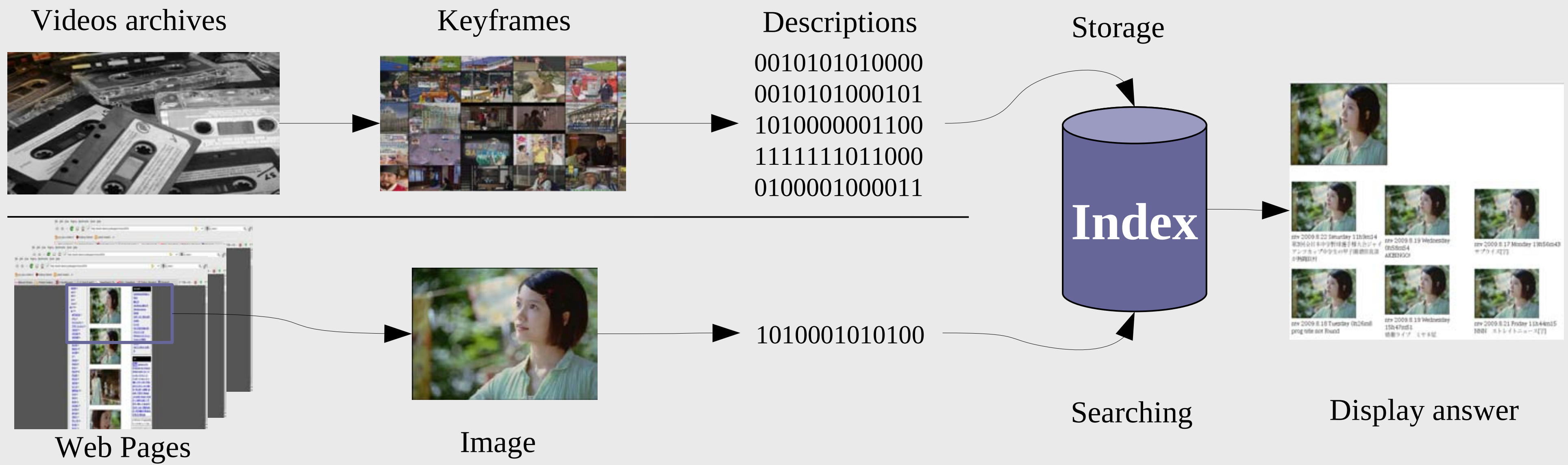
Motivation

❖ **Link Internet resources to Archives:** The Internet is a massive source of detailed informations, by automatically linking pictures (and videos) from web pages to precise position in the videos, the surrounding texts can be used for automatic or semi automatic annotation (high granularity and precise descriptions).

Proposal

❖ We mainly have to face the **very large size** of the existing video databases. The focus is on a **scalable** solution. However it should be capable to handle various transformations between images and videos. For now, only images that are **copies** (screen shots for example) can be linked to a video source.

Framework Overview



Videos archives

Keyframes

Descriptions

0010101010000
0010101000101
1010000001100
1111111011000
0100001000011

Storage

Index

Web Pages

Image

1010001010100

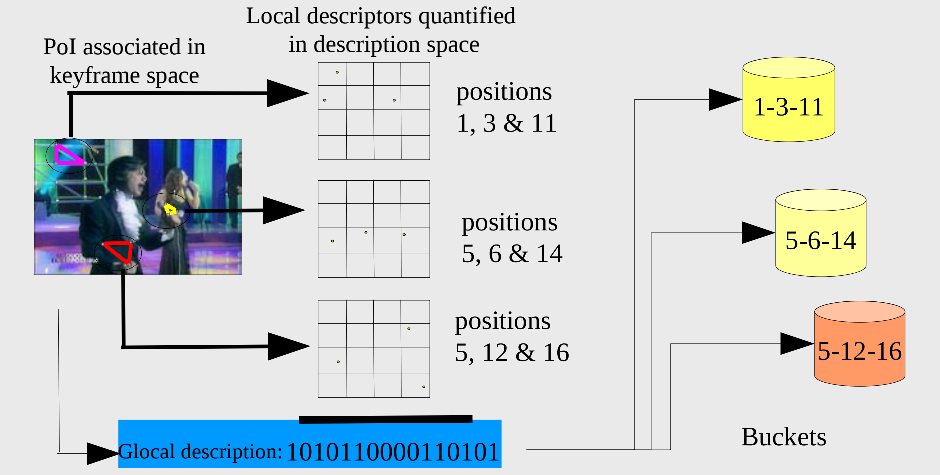
Searching

Display answer

Experiments and perspectives

1. Method

- ❖ **Keyframe** extraction based on global luminance changes (about 1KF/second)
- ❖ **PoI** detection and **local** description extraction (DoG and SIFT).
- ❖ **Frame level** description using quantization (similar to BoF approach).
- ❖ Indexing based on **locality** of PoI in the frames.



PoI associated in keyframe space

Local descriptors quantified in description space

positions 1, 3 & 11

positions 5, 6 & 14

positions 5, 12 & 16

1-3-11

5-6-14

5-12-16

Buckets

Glocal description: 1010110000110101

2. Video Datasets and Server

- ❖ Video recordings on 7 channels for 6 weeks: 8000 hours, 27 millions of keyframes.
- ❖ Database size: 38Gb.
- ❖ Server: 228Gb of main memory.
- ❖ A 40,000 hours database is feasible.

3. Applications

- ❖ **Fast** answer for searching pictures or videos in the reference database.
- ❖ **Mining** based on occurences in a database.
- ❖ Auto **propagation** of annotations.
- ❖ Video clustering.
- ❖ Automatic and semi automatic processes.

4. Next steps

- ❖ Test with larger and larger databases, one year of recordings: **60,000h**.
- ❖ **Parallel** implementation.
- ❖ Automatic update of the database: investigate the **dynamic** aspects.
- ❖ Link with an automatic annotation **propagation** application → some precision improvements are necessary.

5. 10 years after

- ❖ Instant browsing in video space.
- ❖ Dynamic behavior of the video space.
- ❖ New content based video links: people, objects, places…
- ❖ Fusion of multi modal views: **real multimedia**.

NII

Sebastien POULLOT and Shin’ichi SATOH, National Institute of Informatics

TEL : 03-4212-2583 FAX : 03-3556-1916 Email : {spoullot,satoh}@nii.ac.jp

Celebrity Search in Large Video Archives

Duy-Dinh LE, Thanh Duc NGO, and Shin'ichi SATOH

National Institute of Informatics

Motivation

❖A large volume of search queries in image search engines is involved to celebrities.

❖Most of the current image and video retrieval systems rely on text descriptions for searching persons rather than identifying them.

❖Few studies on organizing very large video archives using faces.

Proposal

❖We introduce a framework for person search in large video archives.

❖Huge number of faces in videos are extracted, tracked, and grouped into clusters.

❖Semi-supervised techniques are used to annotated face clusters in efficient and effective ways.

Feature Highlights

1. Large Scale Video Dataset

❖NHKNews7 2001-2008:

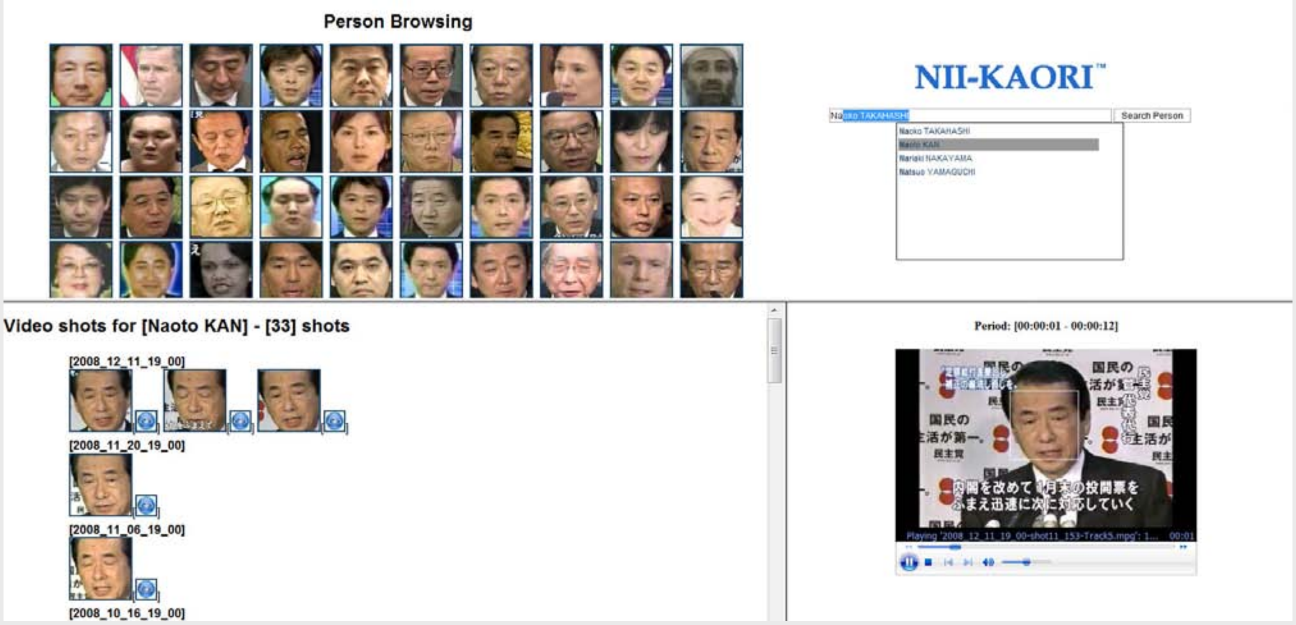
- ❖ ~ 1,500 hours
- ❖ ~ 35 million faces

2. Many Celebrities

❖118 persons:

- ❖ government leaders
- ❖ athletics
- ❖movie stars.

People search by names and images



Future Goals

1. Person Search Engine

❖*Search all appearances of all people appearing in videos.*

2. Applications

❖*Robust face identification for security, video conferencing, and human computer interaction.*

NII

Duy-Dinh LE, Thanh Duc NGO, and Shin'ichi SATOH, National Institute of Informatics
TEL : 03-4212-2583 FAX : 03-3556-1916 Email : {ledduy,ndthanh,satoh}@nii.ac.jp

Large-Scale News Topic Tracking and Key-scene Ranking with Video Near-Duplicate Constraints

Xiaomeng WU[1] Ichiro IDE[2] Shin’ichi SATOH[1]

1 National Institute of Informatics 2 Nagoya University

Motivation

- ❖ Topic tracking is a fundamental step for news browsing, retrieval, topic threading, and summarization.
- ❖ Providing summarization and visualization of large-scale news videos is an important task, which aims at saving time for general audiences in searching and reading news of interest.

展望

- ❖ デジタル放送の普及につれて、意味的に構造化された映像・メタデータの効率的流通について考える必要があり，これらのコンテンツの制作や再利用を行うコンテンツ提供者，コンテンツを伝送若しくは視聴するための端末を開発する企業や大学，コンテンツを楽しむ視聴者，皆が有意義に参加できるプロジェクトを設定し，具体的目標を定めて連携して進めていく予定である．

NEWS TOPIC TRACKING WITH NEAR-DUPLICATE CONSTRAINTS

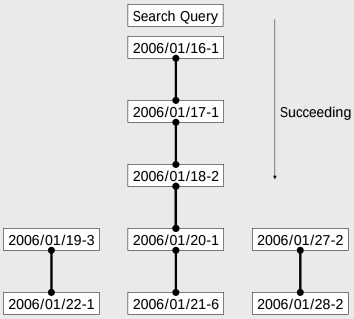
- ❖ **News Story Tracking Based on Textual Information**

- ❖ Topic Segmentation
- ❖ Topic Threading

- ❖ This procedure forms a simple story link tree starting from the story of interest.

- ❖ Children stories are defined as news stories related to a parent, under the condition that the time stamps of the children stories always chronologically succeed their parent.

- ❖ The link tree can also be considered a set of candidate news stories that is similar to the search query.



- ❖ **Video Near-Duplicate Detection**
- ❖ LIP-IS+OOS [9]

FNN SPEAK
2006/11/06

FNN SPEAK
2006/12/27

FNN SPEAK
2006/12/28

NHK NEWS 7
2006/11/05

NHK NEWS 7
2006/12/27

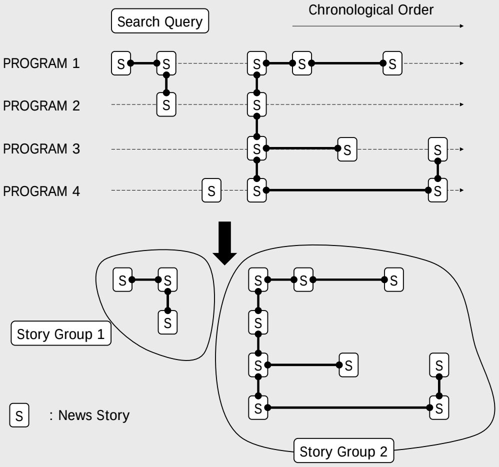
NHK NEWS 7
2006/12/29

- ❖ **Near-Duplicate Constraints**

- ❖ Assumption 1: Most stories in the same story group depict the same topic.
- ❖ Assumption 2: The largest story group depicts the same topic as the query.

- ❖ Based on these assumptions, the largest story group is chosen as the expanded query to represent the characteristics of the corresponding news topic that the query depicts.

- ❖ **Keyword Vector**
- ❖ General / Personal / Locational (Organizational)



- ❖ **TF-IDF Weighting Based on Expanded Query**
- ❖ Different from traditional tf-idf weighting schemes, we consider one story group as one document.

- ❖ **TF-IDF Similarity Based on Expanded Query**
- ❖ The similarity between each story and the search query is re-defined as the cosine similarity between each story and the expanded query.

KEY-SCENE RANKING AND EXPERIMENTS

- ❖ **n_{PR}**
- ❖ number of programs across which the corresponding scene is broadcasted

- ❖ **n_{ST}**
- ❖ number of stories across which the corresponding scene is broadcasted

- ❖ **n_{SH}**
- ❖ number of shots in the corresponding key-scene group

$n_{PR} = 3$
$n_{ST} = 8$
$n_{SH} = 13$

$n_{PR} = 3$
$n_{ST} = 6$
$n_{SH} = 8$

$n_{PR} = 3$
$n_{ST} = 5$
$n_{SH} = 8$

$n_{PR} = 1$
$n_{ST} = 3$
$n_{SH} = 8$

$n_{PR} = 1$
$n_{ST} = 3$
$n_{SH} = 8$

$n_{PR} = 1$
$n_{ST} = 2$
$n_{SH} = 8$

- ❖ Observation 1: Important scenes are normally broadcasted across different news programs (e.g. $n_{PR} \geq 2$), while scenes of reporters are only used by one single program ($n_{PR} = 1$).

- ❖ Observation 2: Within one story, scenes of reporters are normally used more frequently than important scenes (i.e. scenes of reporters tend to have lower value of n_{ST}/n_{SH} than important scenes).

- ❖ Observation 3: Key-scenes broadcasted across important stories are also important

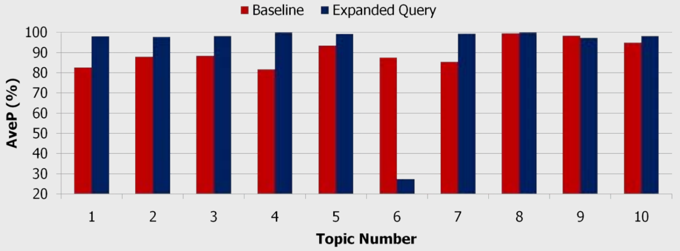
$$sr_{PR}(k_i) = \frac{n_{PR}(k_i)}{N_{PR}}$$

$$sr_{ST}(k_i) = \frac{n_{ST}(k_i)}{n_{SH}(k_i)}$$

$$sr_{R}(k_i) = \frac{\sum_{s \in \Gamma_i} r(\Sigma, s)}{n_{ST}(k_i)}$$

$$\Gamma_i = \{s : k_i \in s\}$$

$$sr = (w \times sr_{PR} + (1 - w) \times sr_{ST}) \times sr_R$$



Rank: 1/10
sr = 0.32

Rank: 2/10
sr = 0.31

Rank: 3/10
sr = 0.29

Rank: 4/10
sr = 0.29

Rank: 5/10
sr = 0.28

Rank: 6/10
sr = 0.24

Rank: 7/10
sr = 0.23

Rank: 8/10
sr = 0.08

Rank: 9/10
sr = 0.04

Rank: 10/10
sr = 0.02

NII
Xiaomeng WU ／ National Institute of Informatics
TEL：+81-3-4212-2662 Email：wxmeng@nii.ac.jp