大規模放送映像アーカイブの構造化に関する研究

A study on structuring large-scale broadcast video archives

佐藤真一　片山紀生　孟洋　井手一郎

Shin’ichi SATOH　Norio KATAYAMA　Hiroshi MO　Ichiro IDE

何がわかる？

ニュース映像をはじめとする放送映像は、社会情勢をはじめ、流行や事実、そして各種の知識など、人間の知的活動に役立つ数多くの情報を提供しています。この研究プロジェクトでは、単一の番組のみならず、過去から現在までに放送された大量の番組を概観することで、注目度の高い事象や流行の変化、話題の経過を捉えるなど、様々な視点で映像情報の知識化をはかり、人間の知的活動を支援するシステムの実現を目指しています。

どんな研究？

東京地区で見ることのできるテレビ番組を過去１ヶ月分、ニュースなど特定の番組を過去数ヶ月から数年分、巨大なハードディスクに蓄積しています。ここでは「映像」「音声」「文字」の各情報解析技術を複合的に用いることで、これら蓄積された多量の映像情報間の関係を明らかにし、放送映像に内在する様々な情報を発見する技術、及びそれら情報を管理・提供する技術の研究を行っています。

目的＆効果

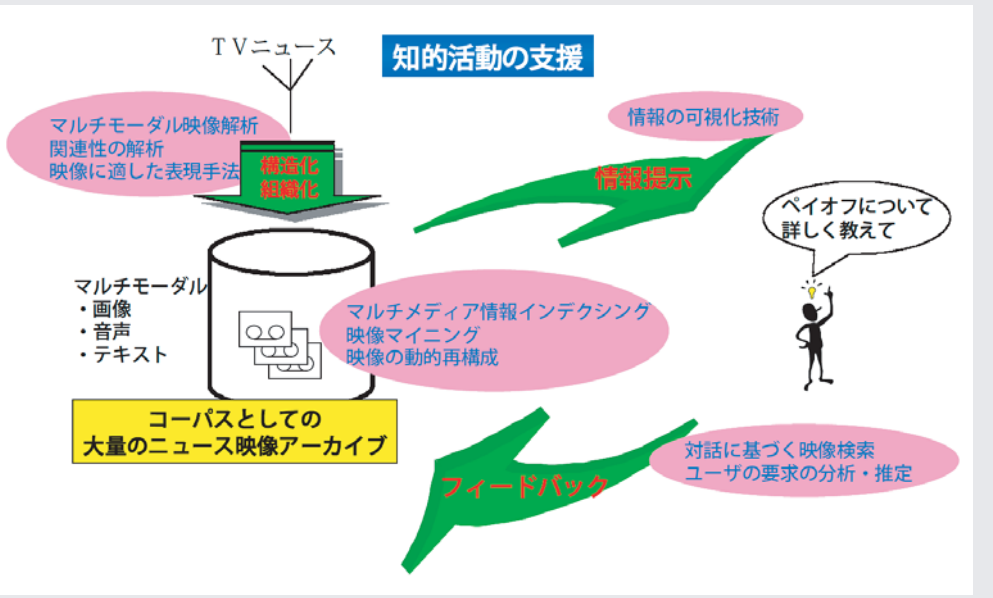
■目的

ニュース映像は、放送という媒体を通じて誰でも容易に取得することができ、映像という視覚的・直接的で理解しやすい形態の情報を提供している上、資料的価値が高く速報性を有しており、人間の知的活動を支援し得る情報を豊富に有していると考えられる。本研究では、実際にニュース映像を大量に獲得し、構造化・組織化して蓄積し、必要な情報を検索・発見し提供することにより、人間の知的活動を高度に支援するシステムの実現を目的とする。

■効果

日々のニュース映像は、それを試聴するだけでも有効な情報を人間に提供してくれる。その上、過去のニュースも含み大量に蓄積し、必要なニュース映像を即座に検索して提供したり、複数のニュース映像間から関連性を抽出したり、情報を利用者の要求に密接に適応した形態で提供したりすることができればニュース映像から人間が利用可能な情報の量も質も格段に高まり、飛躍的に高度な知的活動の支援が可能となる。

また、映像は動画像・音声・テキストなど複数のメディアにより複合的に表現された情報であり、本研究の遂行に当っては、このマルチモーダルな情報の解析、およびデータベース化の検討が必要不可欠である。映像のマルチモーダルな解析手法の検討、および本研究で対象とするような大規模かつ動的なマルチメディア情報によるデータベース構築は、今後技術の確立が切望されている分野であり、学術的な意義も大きい。



知的支援の枠組

放送映像アーカイブシステム
- ニュース（2001年3月～）
- 東京地区地上波（7ch1ヶ月）

技術的課題

■技術的課題

本研究の遂行に当り、以下のような技術的課題を解決していく必要がある。

○大量のニュース映像を解析し、構造化・組織化することにより、その内容情報を抽出するための技術
- マルチモーダル映像解析
- 映像間の関連性の解析
- 映像に適した構造化・組織化のための表現手法

○構造化・組織化された映像を蓄積・管理し、必要な情報への効率よいアクセスを可能とするとともに、大量の映像中の関連性などに基づく新しい知識発見を実現するための技術
- マルチメディア情報インデクシング
- 映像マイニング

○ユーザとの対話を通して、ユーザが必要としている情報を必要な形態で提供するための技術
- 対話に基づく映像検索
- ユーザの要求(意図・視点・興味)の分析・推定
- 映像の動的再構成
- 情報の可視化技術

ソフトウェアプラットフォームとしての映像ブラウザ

NII

連絡先：佐藤　真一（Shin’ichi SATOH）／　国立情報学研究所　コンテンツ科学研究系　教授
TEL：03-4212-2000　FAX：03-3556-1916　Email：satoh@nii.ac.jp

ニュース映像アーカイブの構造化に基づく映像アクセス手法

Intelligent video access based on structuring of a news video archive

孟洋　山岸史典　井手一郎　佐藤真一　坂内正夫

Hiroshi MO　Fuminori YAMAGISHI　Ichiro IDE　Shin’ichi SATOH　Masao SAKAUCHI

何がわかる？

様々な時事問題を扱うニュース映像。日々放送される大量のニュース映像の中から如何に視聴すべきニュース映像を選びだし、わかりやすい形で見せることができるのか。この研究では、過去から現在までの長期間にわたるニュース映像を解析し、「今週おこった大事件は何？」「あの事件のその後はどうなった？」「この映っている人は誰？」など、視聴者の疑問に答えることができるニュース映像閲覧システムの実現を目指しています。

どんな研究？

HDDビデオレコーダの普及に伴う、放送型から蓄積型への視聴環境の変化は、新しい映像情報の活用を可能とします。例えば、長期間にわたるニュース映像を用いると、あるニュースの話題性や経過などの情報を得ることができます。ここでは「映像」「音声」「文字」の各異なる視点からの解析をとおして、ニュース映像間の視覚的、意味的な関連性を抽出するなどして、ニュース映像アーカイブに内在する情報を取り出し、活用する研究を行っています。

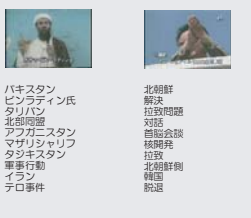
状況設定

■HDDビデオレコーダ、ホームビデオサーバなどの普及
　●長期間にわたる大量のニュース映像の蓄積・視聴が可能に
■蓄積されたニュース映像全てを視聴することは不可能！
　●効率的な視聴を可能とする整理・アクセス技術が必要

■ニュース映像アーカイブの構造化に基づく整理・アクセス技術
　1）多頻出トピックに基づく重要な話題の抽出
　2）トピック間の関連性に基づく話題の流れの追跡
　3）繰り返し映像に基づく代表的な映像ショットの抽出
　4）映像ショットと文字キーワードの共起に基づく索引付け


今週の重要なニュートピックは？

このトピックの経過はどうなってる？

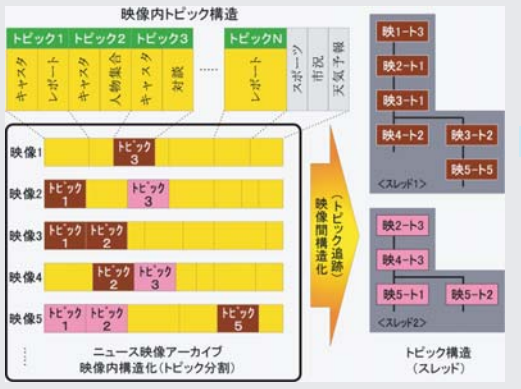


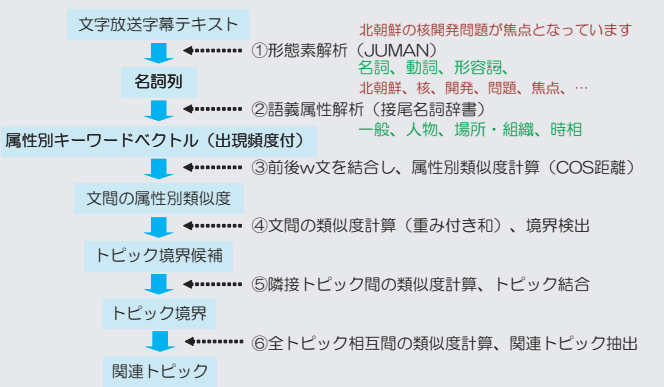
このトピックはどんな内容？

研究状況

トピック構造の抽出（文字情報解析）

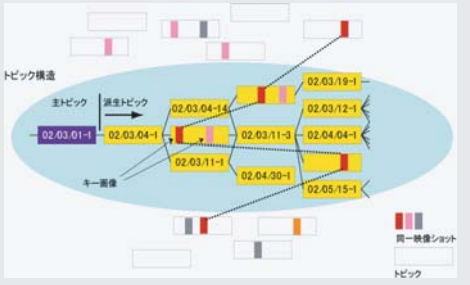
■同一・関連話題内では同一キーワードを多数共有する
　●トピックへの分割
　　ー連続した文間のキーワードの類似性の評価
　●トピックのグルーピング　⇒　重要な話題
　　ートピック間のキーワードの類似性の評価
　●トピックの追跡　⇒　話題の流れ
　　ー類似性と時間順序を考慮した連鎖構造の抽出

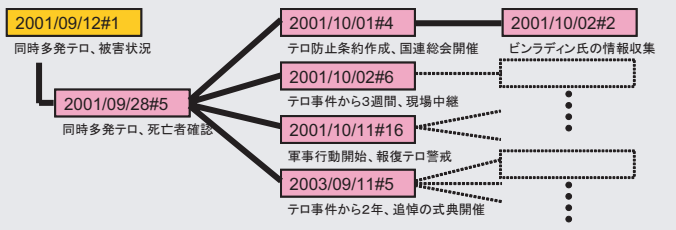




代表的な映像ショット抽出（映像情報解析）

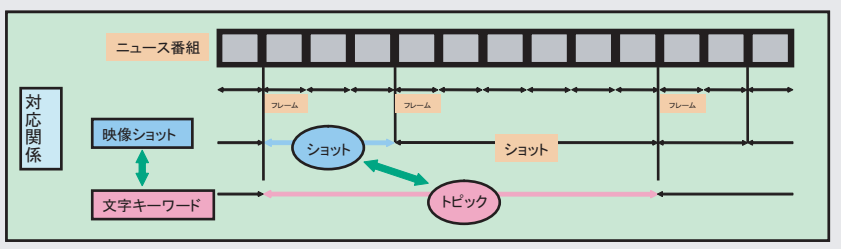
■重要な映像ショットは繰り返し利用される
　●同一映像ショットの抽出
　　ー同一映像フレームを含むショット対の取得
　●話題を代表するキーショットの抽出
　　ー同一トピック構造内の同一映像ショットの取得





映像ショットへの文字キーワードの付与

■同一内容の映像に対し同一のキーワードが出現する
　●共起に基づく映像ショットと文字キーワードの対応付け
　　ー出現分布による共起度の評価




16327 @ 2001/09/15

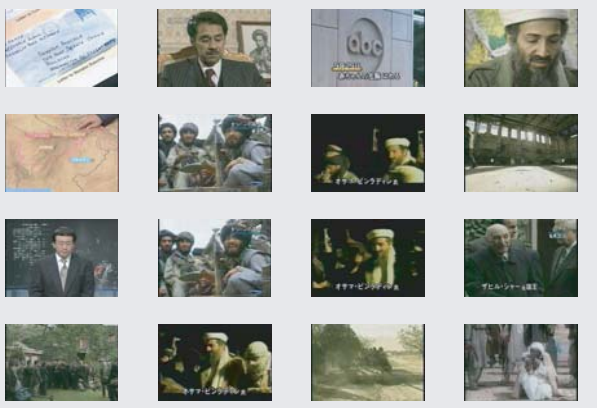
パキスタン	77/899	56.1
ビンラディン氏	56/483	55.9
タリバン	112/1583	54.1
北部同盟	73/1110	46.5
アフガニスタン	94/1656	43.6
マザリシャリフ	28/148	42.3
タジキスタン	28/150	42.1
軍事行動	33/320	38.8
イラン	23/159	34.0
テロ事件	26/258	33.0
テロ組織	21/142	32.1
国境	25/263	31.5
協力	38/736	31.0
ロシア	31/512	30.1
アメリカ軍	40/1022	26.9
アメリカ	166/3327	26.7
報復攻撃	15/91	25.8
準備	22/322	25.8
攻撃	36/946	25.461
テロ	27/567	25.0

（キーワード数：1914）


009923 @ 2002/07/07

北朝鮮	299/2659	77.2
解決	53/454	54.3
拉致問題	35/167	51.1
対話	41/309	48.9
首脳会談	33/197	45.8
核開発	36/275	44.7
拉致	39/351	44.3
北朝鮮側	40/438	41.6
韓国	50/841	37.9
脱退	18/42	37.0
姿勢	40/668	34.3
小泉総理大臣	67/1544	33.1
会談	51/1083	33.0
方々	17/63	32.0
キム・ジョンイル総書記	17/101	28.5
被害者	28/464	28.4
核開発問題	21/218	28.2
IAEA	16/84	28.1
NPT	12/23	27.8
閣僚級会談	13/39	27.1

（キーワード数：2086）



NII

連絡先：孟　洋（Hiroshi MO）／　国立情報学研究所　コンテンツ科学研究系　助手
TEL：03-4212-2000　FAX：03-3556-1916　Email：mo@nii.ac.jp

A General Framework for Large Scale Video Indexing Using Multi-modal Analysis

Duy-Dinh LE — The Graduate University for Advanced Studies

Shin'ichi SATOH — National Institute of Informatics

Michael E. HOULE — National Institute of Informatics

Dat P. T. NGUYEN — The University of Tokyo

Motivation

- Large scale videos need easy, efficient and scalable tools for indexing and retrieving.
- Using human face information for indexing is required since human face is one of the most important objects in video, especially news video.
- Multi-modal analysis can help to bridge the semantic gaps.

Field of Research

VIRMA is a video indexing and retrieval system that can provide:

- Easily access to video contents using faces and names extracted from video.
- Quickly find news stories, video shots related to a person appearing in video.

Framework Overview

- Face extraction and normalization using a robust face and eye detector.
- PCA is used for feature extraction and reduction.
- Named entity recognition using the LingPipe tool.
- Face grouping using RSC-based clustering.
- Name and face association using a machine translation method implemented by GIZA++.

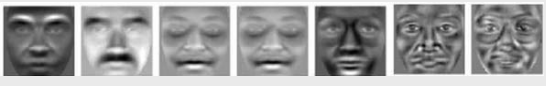
Experiments

1. TRECVID 2003 Dataset

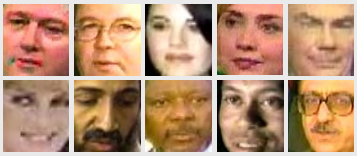
- 133 hour video of CNN and ABC news in 1998.
- Average frames per shot: 100.
- Number of news stories: 4,376.
- Low quality video.

2. Face extraction and normalization

- Normalized size: 52x60.
- Feature reduction using PCA: 786 eigenfaces.
- Number of extracted faces: 28,896.



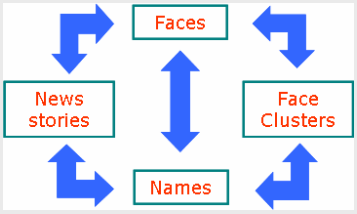
3. Representative faces found by RSC-based clustering model.



4. Name-face association by a machine translation method.

5. Video navigation by faces and names.

- Demo is available at:
- http://satoh-lab.ex.nii.ac.jp/users/ledduy/Demo



NII

Duy-Dinh LE – The Graduate University for Advanced Studies (SOKENDAI)

TEL : 03-4212-2679 FAX : 03-3556-1917 Email : ledduy@grad.nii.ac.jp

Multimodal Video Indexing from Intermediate Concepts

Stéphane Ayache
CLIPS / IMAG
NII Internship student

Shin'ichi Satoh
National Institute of Informatics

Field of Research

- Content-based access of large video libraries
- Search by visual similarity is inefficient because far from users needs

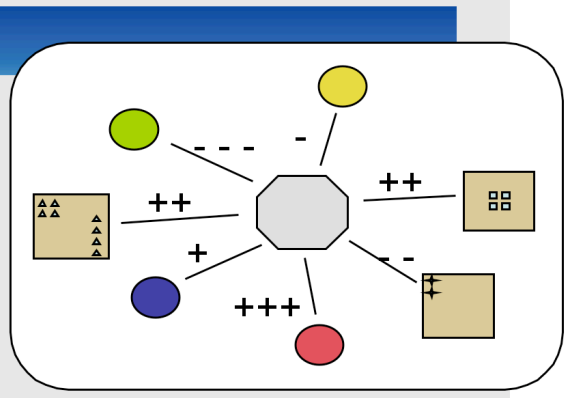
➔ Need to extract semantic from multimodal cues

Aim of Research

- Semantic indexing with a large number of concepts
- Provide a framework to bridge the *semantic gap*
- Merge semantic information from several cues (eg. Image, Text)

Approach

- Detect intermediate concepts from unimodal cues
  - ✓ Visual concepts (Sky, Greenery, Building, Studio setting, Skin)
  - ✓ Topic concepts (Economic, Politic, Sport, Weather)
- Derive High-level semantic concepts from intermediate concepts
- Exploiting implicit relations and contexts between the concepts



Research Description

Visual Concepts:

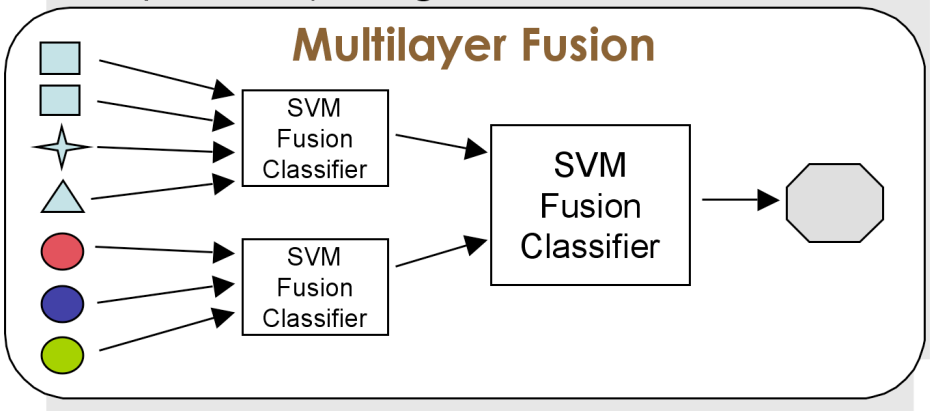
Classification of patches in each keyframes from Low-level features (Color, Texture, Motion). Supervised SVM classifier

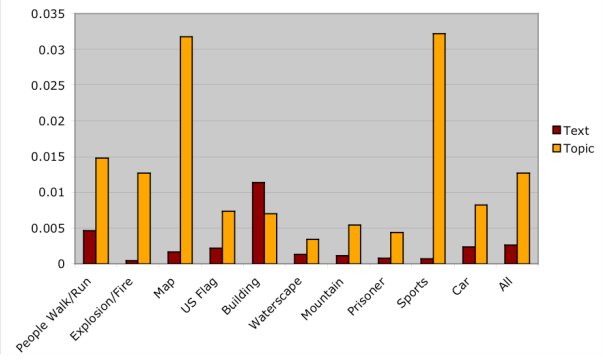
Topic Concepts:

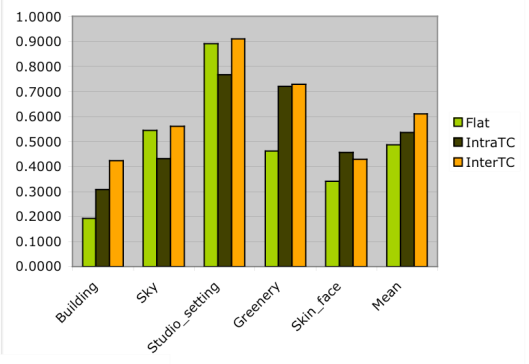
Classification of each speech segment from Automatic Speech Transcription. Trained on Reuters corpora with Rocchio classifier

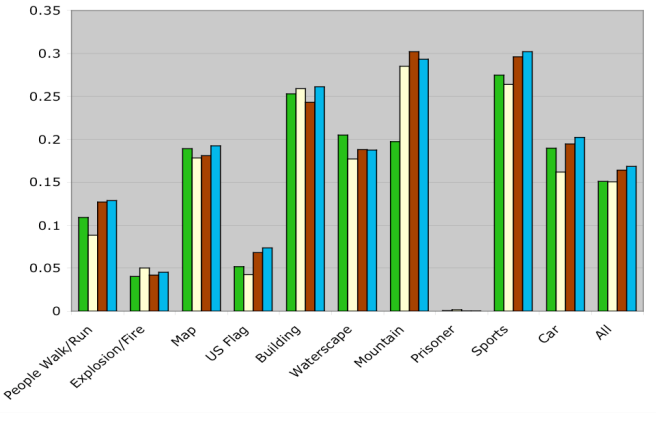
Multimodal Integration:

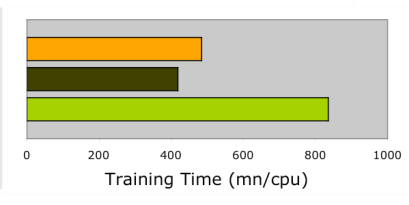
Exploits *topologic* and *semantic contexts* from visual and topic concepts











Experiments on TRECVID 2004 and 2005 corpora

Conclusion - Future works:

- Intermediate concepts help to bridge the semantic gap
- Many possibilities of fusion
- Better use of Topic concepts
- Handle imbalanced input features

NII

Stéphane Ayache
Email: stephane.ayache@imag.fr

Shin'ichi Satoh
Email: satoh@nii.ac.jp

直感的な検索インタフェースに関する検討
On Intuitive Search Interface

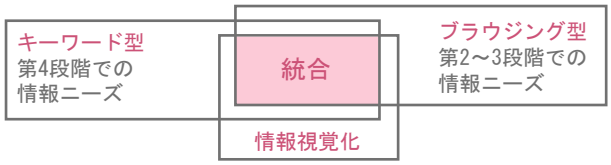
梶山朋子[†,‡]　佐藤真一[‡,†]　[†]総合研究大学院大学情報学専攻　[‡]国立情報学研究所

Information retrieval is a random and drift process. Users' information needs change suddenly, or ambiguous information needs become clear. There are mainly three techniques to support retrieval, keyword search, browsing and information visualization. Each technique has problems, but if they are integrated like 'Concentric Ring View' we proposed, intuitive search can be realized. Intuitive search interface will clarify ambiguous information needs and properly guide novice users to desired information.

情報探しは成り行き任せのプロセス

自分の思考や直感に基づいて、自己責任で進めていくプロセス
人間は情報を探しながら、自らの知識構造を変化させる
知識構造の変化に伴い、情報ニーズも変化する
- 曖昧な情報ニーズが明確化する
- 予想外の情報に出会う

検索とブラウジング



情報ニーズの４段階

(Taylor, 1968)

[第1段階] 心奥のニーズ (visceral need)	漠然と知識が欠けているような感じがする どんな知識が欠けているか分からない どんな情報が必要なのか分からない
[第2段階] 意識したニーズ (conscious need)	どんな情報が欠けているかほぼ分かっている どんな情報が必要か人に説明できない 周囲の人に漠然と質問することで明確化
[第3段階] 具体化したニーズ (formalized need)	欠けている知識が明確に分かっている 論理的な質問を発することができる 答えが欲しい真の疑問が明確となる
[第4段階] 妥協したニーズ (compromised need)	質問する相手の知識構造や理解力を想定 相手が答えられるような質問をする 検索エンジンで探せるようなキーワードを入力

キーワード検索

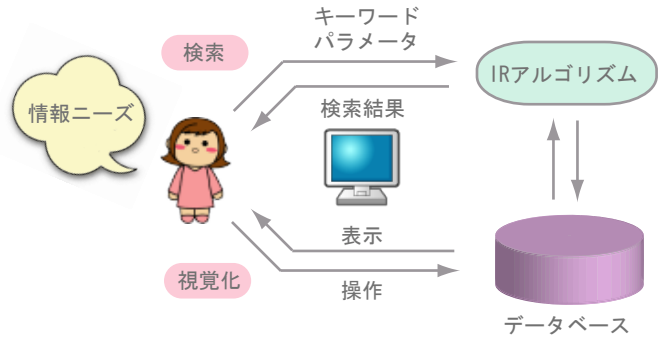
サーチボックスにキーワードを入力して検索することにより、
検索結果が表示される
曖昧な情報ニーズでは、検索を始められない
再検索には、キーワードを改めて考えなくてはならない
言葉で表現しづらいものは、検索が難しい

ディレクトリ型検索

用意された大分類から、求める情報に近いものを選択し、カテゴリの範囲を狭めていくことで、情報にたどり着く
曖昧な情報ニーズでも、分類を選択することにより、検索を始められる
分類の方式がユーザの知識構造と異なる場合、検索が難しい
自分の欲しい情報のない分類へ進んでいることに気づかない

情報視覚化

大量の情報を効果的に表示し、情報を直接理解・操作する
（キーワード空間、ディレクトリ階層、多次元属性）
情報の関連性を見ることができる
関連した情報が線で結ばれたり、全体ー詳細の移動するなど、
表示や操作が複雑である
必要でない情報まで表示されるため、把握に時間がかかる



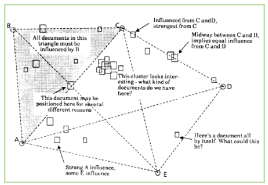
検索とブラウジングの統合

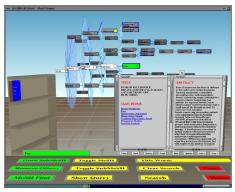
Concentric Ring View

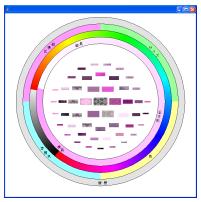
多次元属性情報を対象とする
属性を追加・削除することにより、動的階層構造を生成する
属性値を調節することにより、情報を絞り込む
情報ニーズに適合した候補を選択することで、再検索できる
何かしら属性を選択し操作することにより、検索を始められる
ユーザの思考に合った属性が用意されていない場合、
検索を進めるのが難しい


(http://www.google.co.jp)


(http://shopping.yahoo.co.jp)


(R.Robert 他, VIVE, 1991)


(M.Hearst 他, Cat-a-Cone, 1997)


(T,Kajiyama 他, 2005)

誰でも利用できる検索インタフェースを目指して

曖昧な情報ニーズを明確化し、欲しい情報にたどり着かせるためには、

検索を始めることができる
- とっかかりを用意する

ユーザの思考を妨げない
- システムでユーザの選択を固定しない
- 複雑な操作や表示をしない

検索を進めることができる
- 悩むことなく、クエリを修正できる
- 直接データを操作できる
- システムとの対話が簡単に行える

予想外の情報に出会う機会を与える
- 関連する情報を表示する
- 様々な側面から情報を見られる

ユーザがシステムの特性に合わせることなく、情報を眺めることに集中できるようにさせる

平成18年度 国立情報学研究所オープンハウス　6/8-6/9　於 学術総合センター/一橋記念講堂

Semantic extraction through multimodal analysis from large scale news video archives

Jean Martinet Shin'ichi Satoh
National Institute of Informatics

Research domain

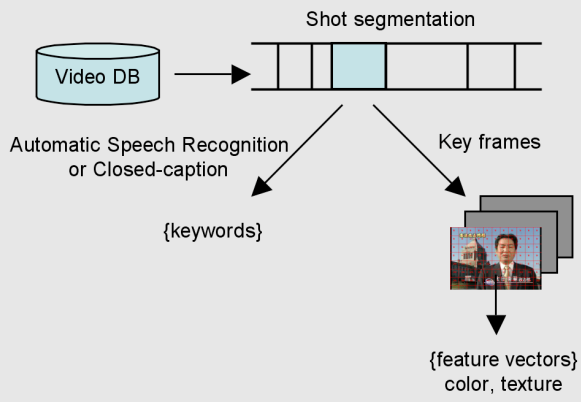
- **Digital content analysis**
- **Semantic understanding and automatic description of multimedia documents**
- **Inter-media data mining**

Applications

- **Image annotation / region naming**
  - **Automatic indexing of image / video databases for retrieval and browsing**
- **Text illustration**
  - **Find images or any visual content related to a given text**

Approach

- **General approach of inter-media data mining:**
  - **Take advantage of information redundancy across several modalities (visual, textual, audio)**
  - **Associate information across media (e.g. visual data and words)**
- **Semantic extraction from video news archives:**
  - **Correlation of visual and textual information**
- **Unsupervised methods using probabilities theory and statistic tools**



Description

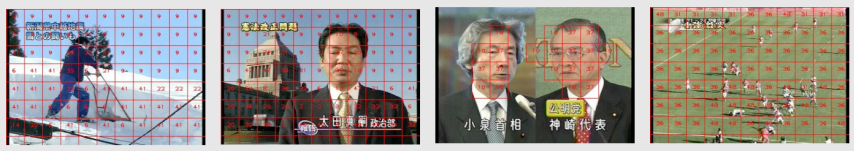
Data pre-processing

Text

- **Filter ASR / closed-caption keywords with standard text processing techniques (stop-word filtering, stemming)**

Video

- **Extract feature vectors from tessellated key-frames**
- **Cluster feature vectors into visually similar categories (using K-means)**



Explore statistical correlation between visual and textual data (TRECVID, NHK)

- **Co-occurrence of blocks and keywords**
- **Block position, context of occurrence**
- **Mutual info, block / keyword entropy**

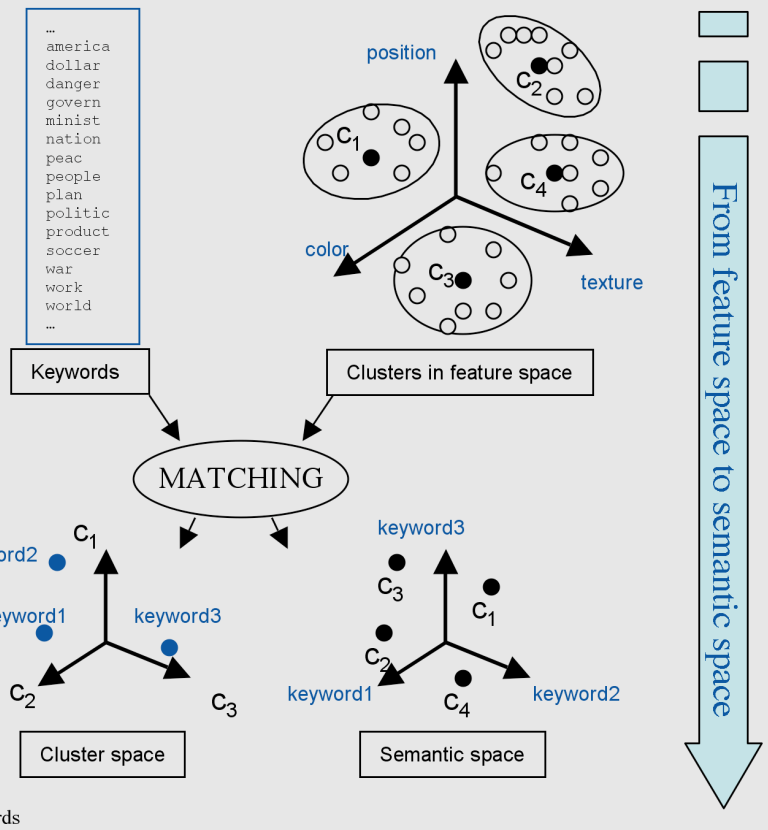
$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)},$$

From feature space to semantics space

- **Keyword signature in feature space**
  - **Block = vector of keywords**
  - **Keyword = vector of blocks**

On-going works

- **Feature selection**
- **Weight regions and keywords (size and position of regions, tf.idf)**
- **Structure keywords in ontology**



NII
Jean Martinet
jean@nii.ac.jp
Shin'ichi Satoh
satoh@nii.ac.jp