

マルチメディアと検索技術

～キーボードを使わずに検索するには？～

国立情報学研究所
コンテンツ科学研究系

片山 紀生

キーボードを使わないとは？

タッチパネル？



Softbank

音声入力？



NEC

脳波入力？



Guger Technologies

今回は、入力装置（入力デバイス）の
話ではありません。

では、どのような話でしょう？

マルチメディア情報と検索技術

マルチメディア情報:

複数の媒体(メディア)から成る情報

例) 映像、画像、音声

(写真、ホームビデオ、テレビ放送)

マルチメディア情報も、キーワードなどの言語情報が付与されていれば、Web ページなどと同様に、キーワードによる検索が可能。

例) Google 画像検索, Flickr, YouTube

しかし、それではテキスト検索と変わりません。

マルチメディア情報の内容による検索

例えば、テレビ放送を検索する時:

- 番組名による検索
- 出演者名による検索
- 解説文による検索

⇒これらはテキスト検索

今回のテーマ

- 野球場の映像の検索
- 記者会見の映像の検索
- ○○(人・物)が映っている映像の検索

⇒これらは内容解析を必要とする検索

あらまし

1. 背景：マルチメディア情報があふれる時代
2. 概論：マルチメディア情報の検索
3. 主題：特徴量による類似検索
4. 応用：テレビ放送からの類似ショットの検出

マルチメディア情報があふれる時代

ペタバイト(テラバイトの千倍)規模の画像や映像

- Web ページの画像 (Google 画像検索など)
- 写真投稿サイト (Flickr など)
- 動画投稿サイト (YouTube など)
- 放送映像アーカイブ (NHK アーカイブス)

学術研究用放送映像アーカイブ

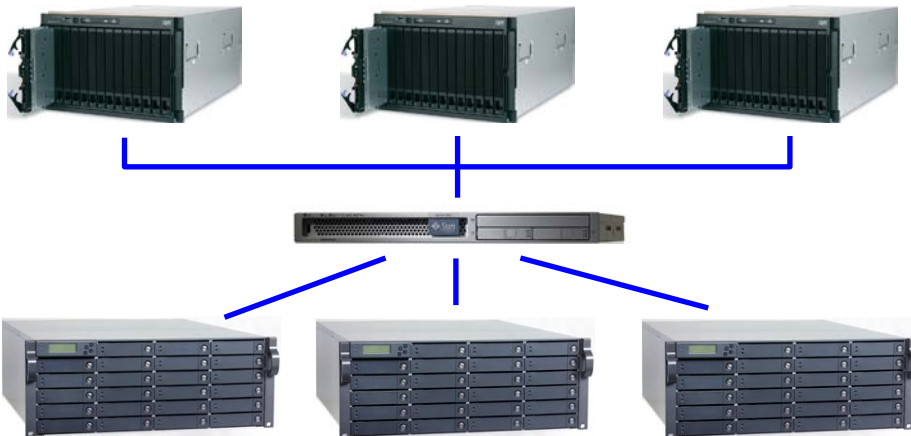
- Informedia (米国カーネギーメロン大学)
- Inamedia (仏国 INA 研究所)
- NII TV-RECS (国立情報学研究所)

マルチメディア情報検索は 情報学のフロンティアのひとつ

- ようやく大量に蓄積できるようになった段階。
- 内容解析などの処理は、現在の高速なプロセッサを用いても挑戦的な課題。

NII TV-RECS

東京地区7チャンネルを昨年8月より蓄積。
1日分だけでも168時間。累計8万時間以上。



高密度サーバ (400コア)

簡単な処理でも、1日分に
1時間、1ヶ月分に丸1日！

大容量ディスクアレイ (200TB)

あらまし

1. 背景：マルチメディア情報があふれる時代
2. 概論：マルチメディア情報の検索
3. 主題：特徴量による類似検索
4. 応用：テレビ放送からの類似ショットの検出

マルチメディア情報の検索

● 画像検索

- 報道写真
- 博物館などの写真・
絵画
- 医療画像
- Web画像
- 航空写真
- 衛星写真

● 映像検索

- 放送映像
- ホームビデオ
- 監視カメラの映像

● 音声検索

- 音声言語
(≒テキスト検索)
- 音楽

(1) メタデータによる検索

(2) 内容の解析による検索

マルチメディア情報の検索

- メタデータによる検索(属性情報による検索)

キーワード、注釈、分類情報、時刻情報、内容記述

- 内容解析による検索(Content-Based Retrieval)

- 特徴量による類似検索

- 画像・映像:

色特徴、領域特徴、形状特徴、テクスチャ特徴
物体単位での特徴量抽出(顔、建物、背景等)

- 音楽: 周波数特徴

- 内容解析による属性情報の抽出

- 文字認識による文字情報の抽出

- 音声認識による書き下しテキストの抽出

メタデータ vs 内容解析

(1) メタデータ

- 基本的にテキストベース。
- テキストを対象とした手法が応用可能。
- 内容解析に比べ処理コストが小さい。
- 高い検索精度を期待できる。実用性が高い。

(2) 内容解析

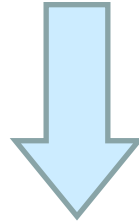
- 処理コストが大きい。
- 計算機性能が上がり、精度も向上すれば、
 - (a) メタデータの自動付与に応用できる。
 - (b) メタデータにはない情報発掘が可能。

あらまし

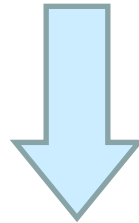
1. 背景：マルチメディア情報があふれる時代
2. 概論：マルチメディア情報の検索
3. 主題：特徴量による類似検索
4. 応用：テレビ放送からの類似ショットの検出

特徴量による類似検索

映像や音声のデジタル化(デジタル記録)

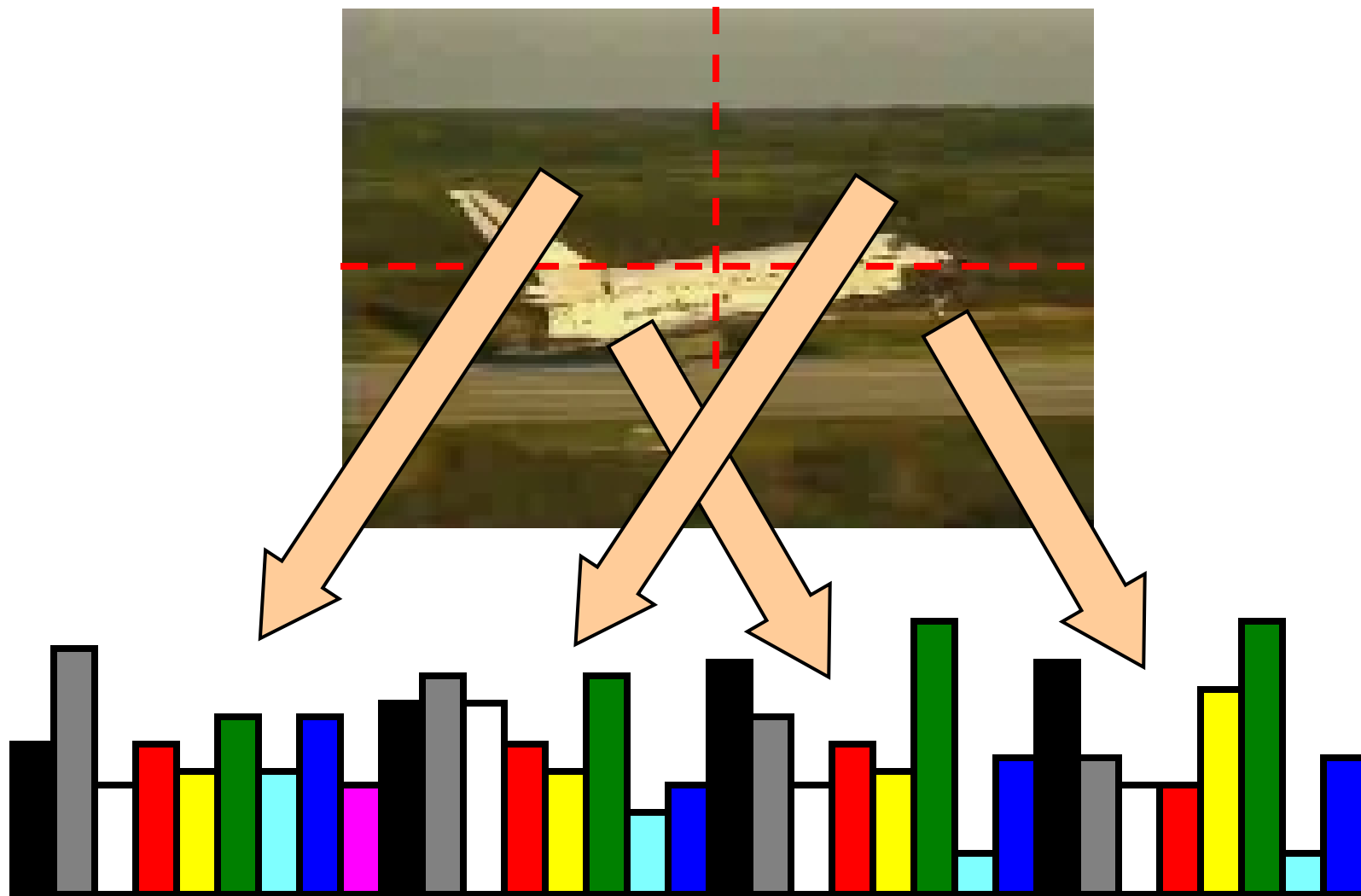


デジタル化された映像・音声からの特徴量抽出

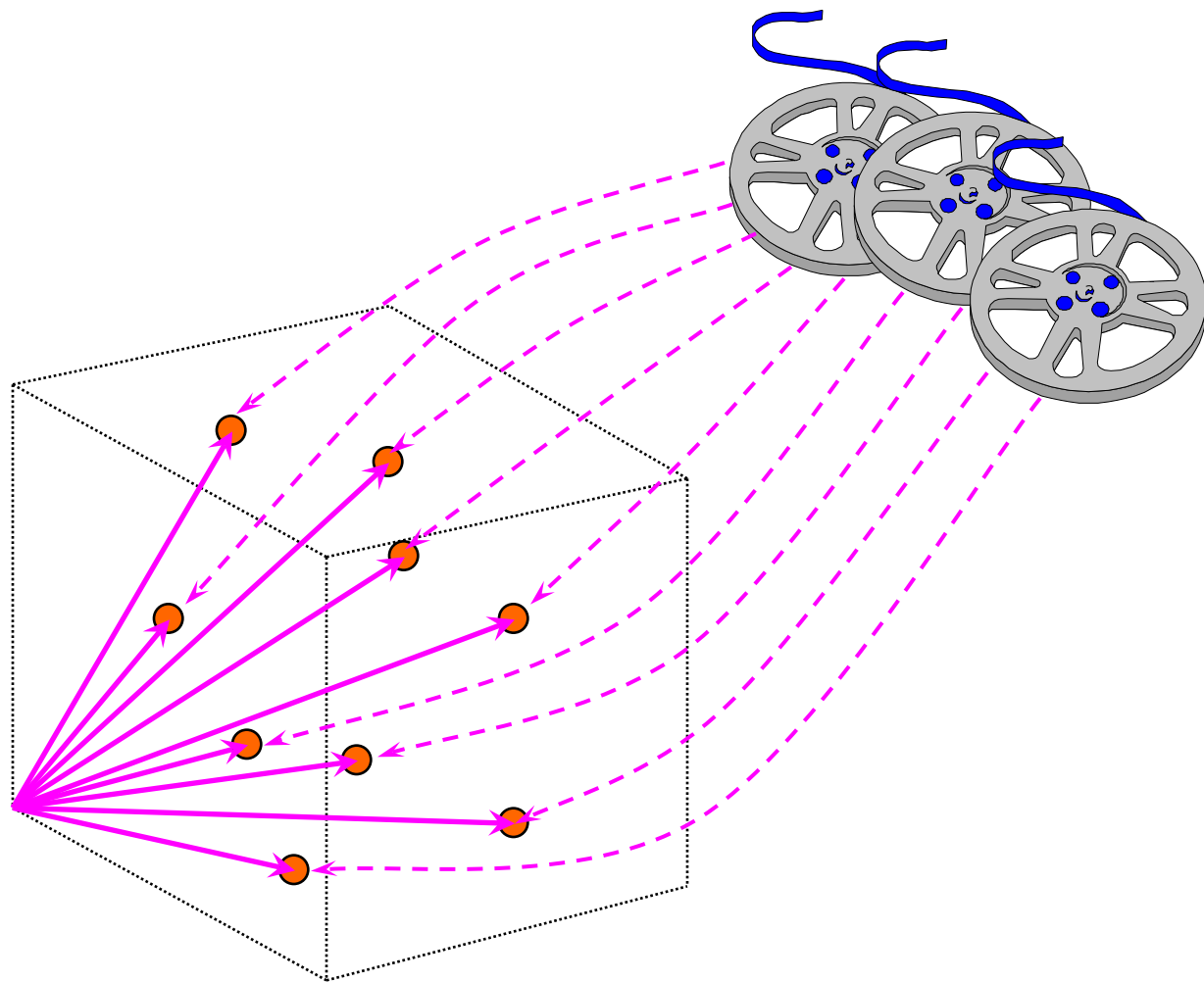
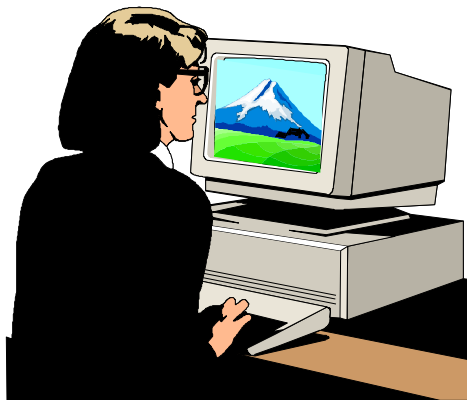


特徴量の照合による類似検索

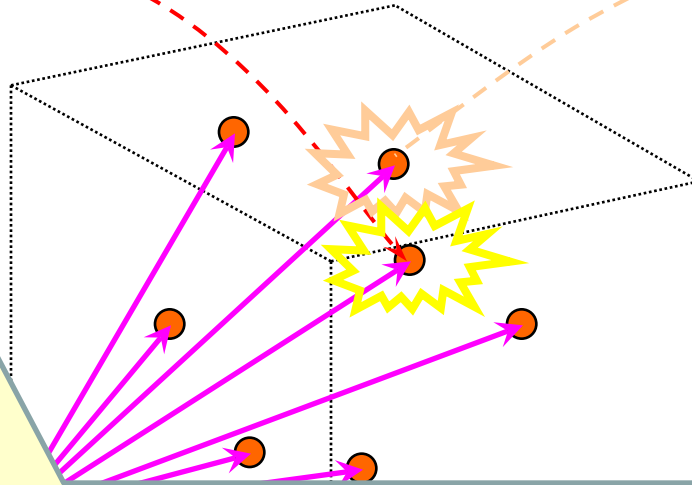
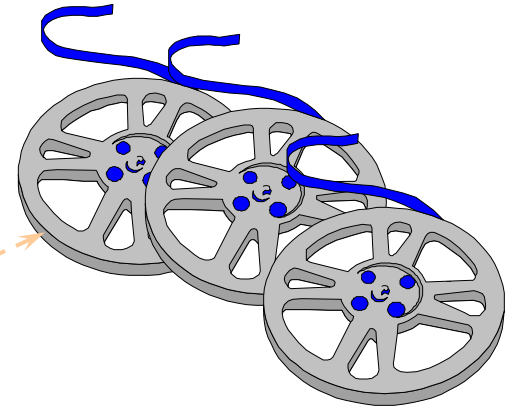
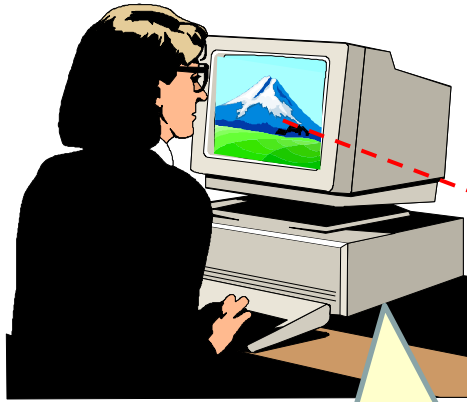
色特徴の例：カラーヒストグラム



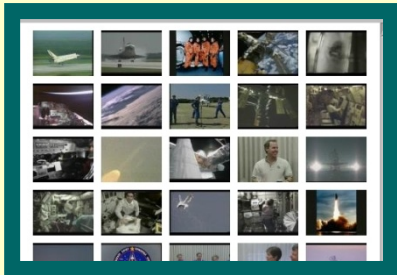
画像・映像から特徴量を抽出



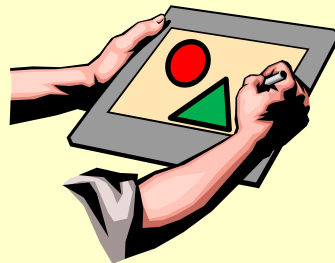
特徴量の照合による類似検索



見本から選択



スケッチ



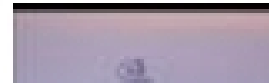
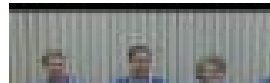
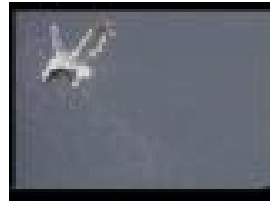
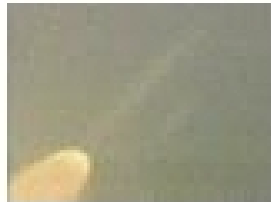
撮影



録音



検索例



検索例

SR-Tree

CPU 時間

0.213 秒

ディスク I/O

232 キロバイト



0.000000



0.104250



0.178399



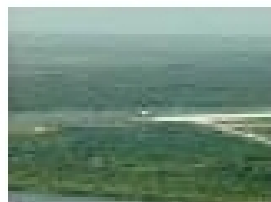
0.200112



0.218683



0.241580



0.300035



0.319319

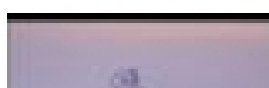
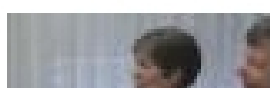
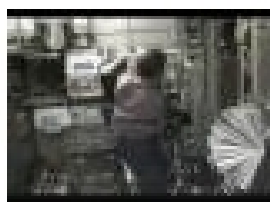
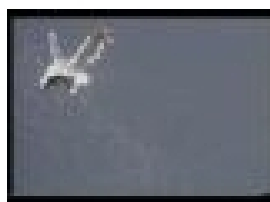
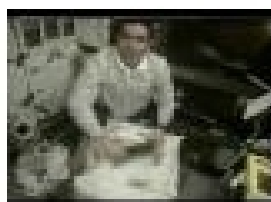
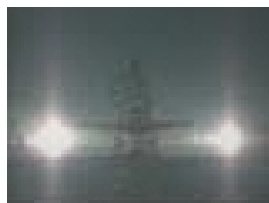
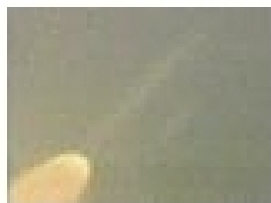
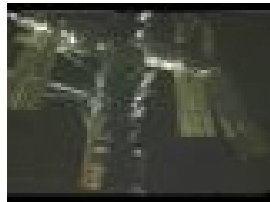


0.357990



0.366802

検索例



検索例

SR-Tree

CPU 時間



ディスク I/O



0.000000



0.202575



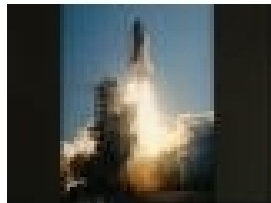
0.224234



0.241305



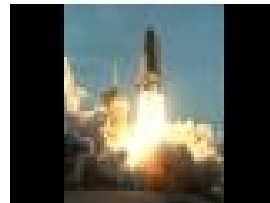
0.244123



0.268849



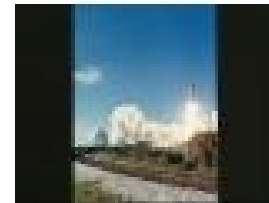
0.272393



0.277421



0.286772



0.292795

類似検索における研究課題

映

完全に同一なもの同士の照合なら簡単。
軽微な差異（回転、縮尺、歪みなど）があっても、漏れや誤りが少ない特徴量。

デジタル化された映像・音声からの特徴量抽出

大量のデータに適用可能な
高速な照合アルゴリズム

特徴量の照合による類似検索

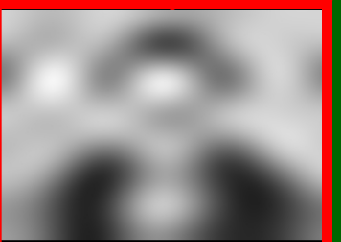
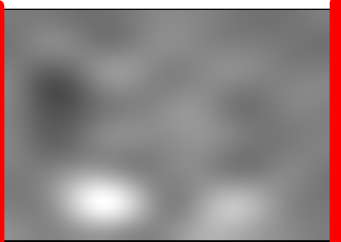
特徴量の例

ぼかす(高周波成分を捨てる)
ことで、主要な差異に絞る。

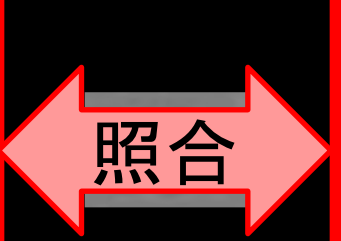
原画像



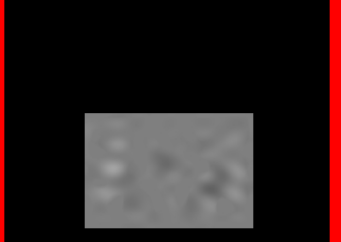
輝度成分



線成分
(水平)



線成分
(垂直)



照合

画像処理
結果

DCT特徴量

DCT特徴量

画像処理
結果

あらまし

1. 背景：マルチメディア情報があふれる時代
2. 概論：マルチメディア情報の検索
3. 主題：特徴量による類似検索
4. 応用：テレビ放送からの類似ショットの検出

テレビ放送からの類似ショットの検出

NHK



日テレ



TBS



フジ



テレ朝



テレ東



- 同一ショットの検出が、高精度に可能なことを活かし、2つ以上のチャンネルに現れる類似ショットを検出。
- CM が混ざって検出されるという問題があるが除去可能。

出現チャンネル数での重み付けによる ニュースショットのカレンダー表示

Date	News Shot Cloud
2009_09_01	
2009_09_02	<p data-bbox="463 592 1535 921">類似検索の仕組みにより 効率的な一覧表示を実現</p> 
2009_09_03	

出現チャンネル数が多いショットに絞り、
かつ、多いものほど大きく表示

まとめ

1. 背景：マルチメディア情報があふれる時代
2. 概論：マルチメディア情報の検索
3. 主題：特徴量による類似検索
4. 応用：テレビ放送からの類似ショットの検出

今後期待される展開

- 汎用かつ詳細な内容解析法の構築
- 高精度かつ高速な類似検索法の実現

この分野の今後の発展にご期待を！

関連書籍

『石頭なコンピュータの眼を鍛える —コーパスで人間の視覚にどこ まで迫れるか—』

佐藤 真一（国立情報学研究所・教授）、
齋藤 淳（サイエンスライター） 著

情報研シリーズ12

丸善ライブラリー刊（2009年12月）

ISBN 978-4-621-05378-2

コンピュータによる映像理解技術の
最近10年の研究動向を紹介。

