

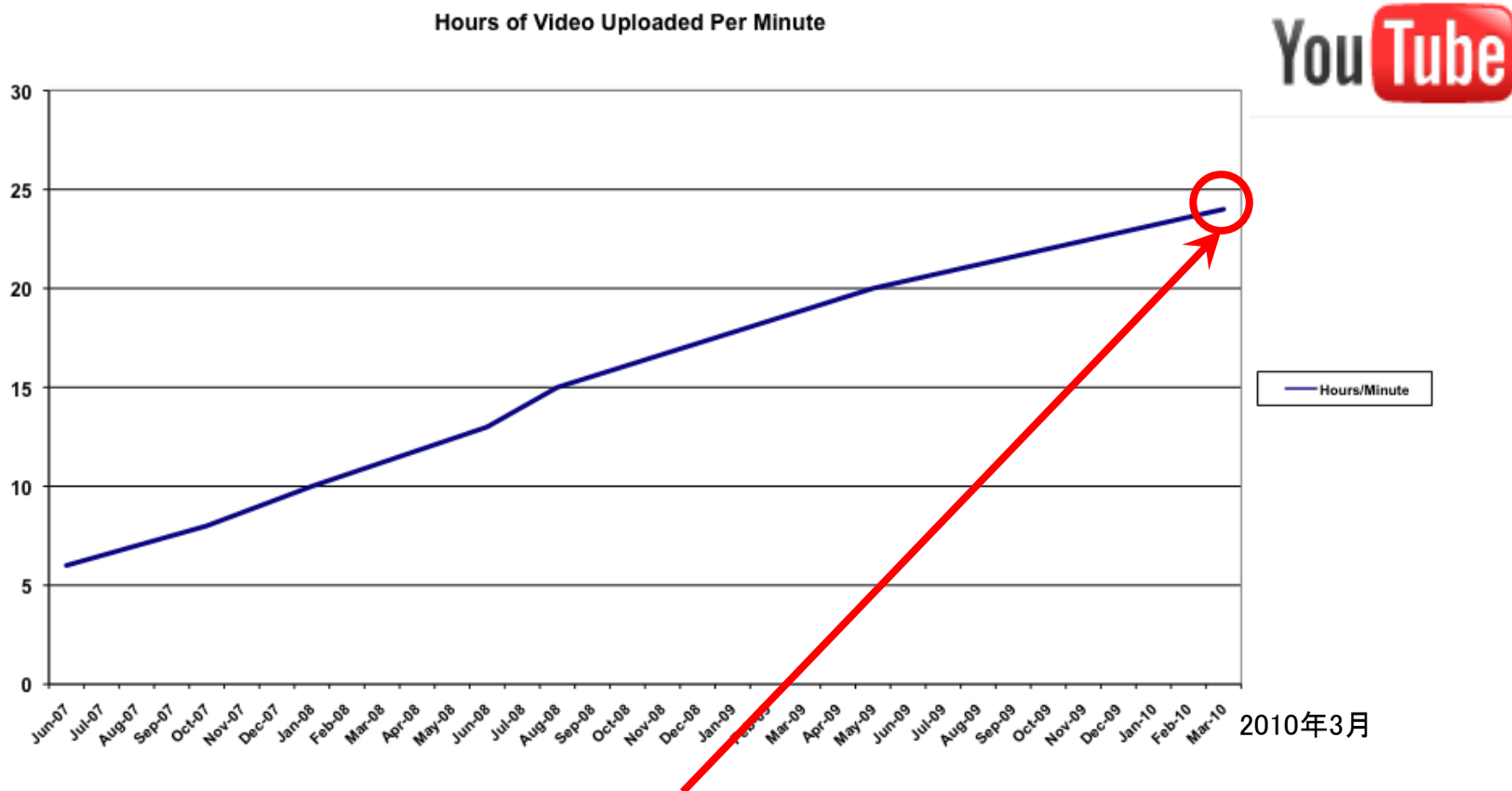
複合現実感技術との融合によるユーザの 演技に基づく映像アーカイブ検索システム の研究開発

白浜 公章

神戸大学大学院経済学研究科

(現在、室蘭工業大学しくみ情報系領域)

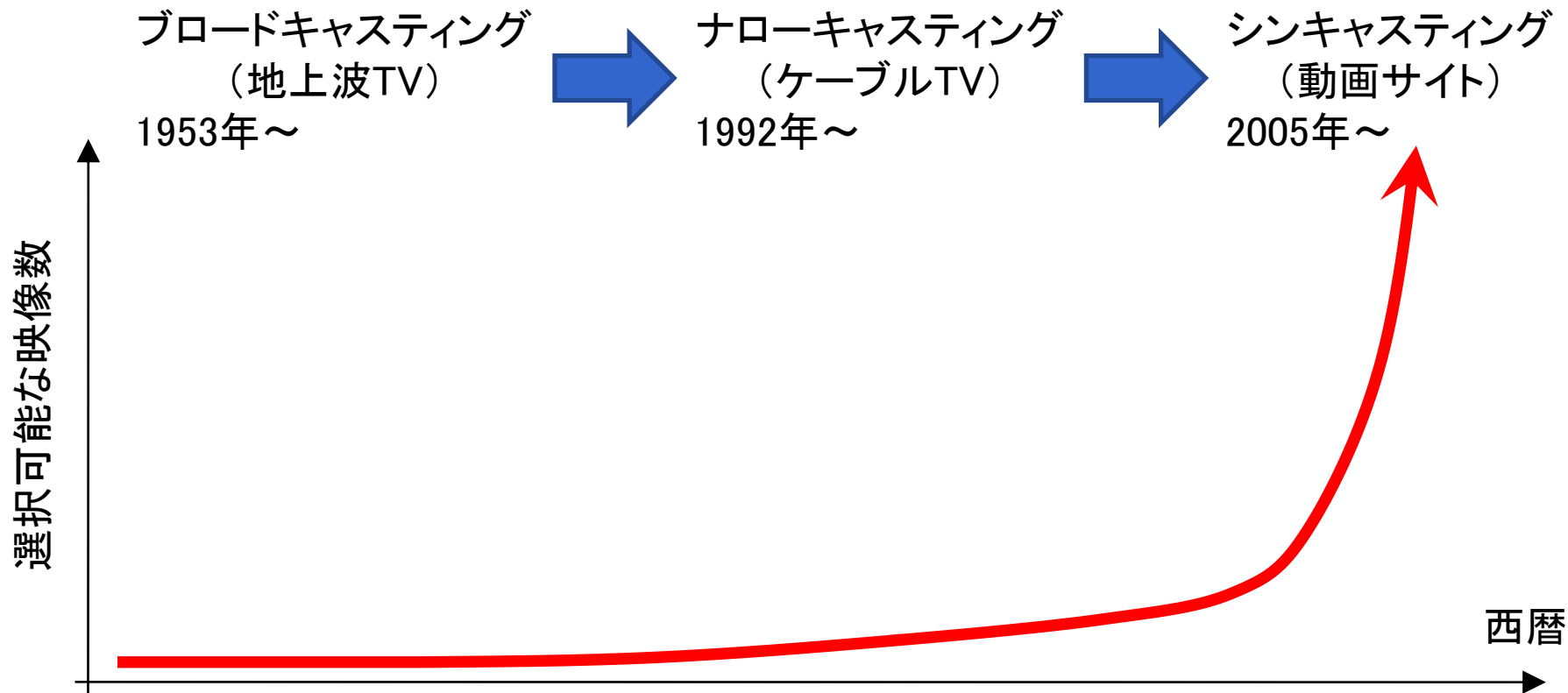
増加し続ける映像データ



1分間に24時間分の映像が新規にアップロードされる！

60日間にYouTubeにアップロードされる映像数 > 米国3大ネットワークが過去60年間に制作した番組数

映像配信形態の変化



大量の映像の中から、必要とする映像を効率的に探し出す映像検索技術の重要性！

キーワード検索: Query By Keyword (QBK)

侍 刀

検索

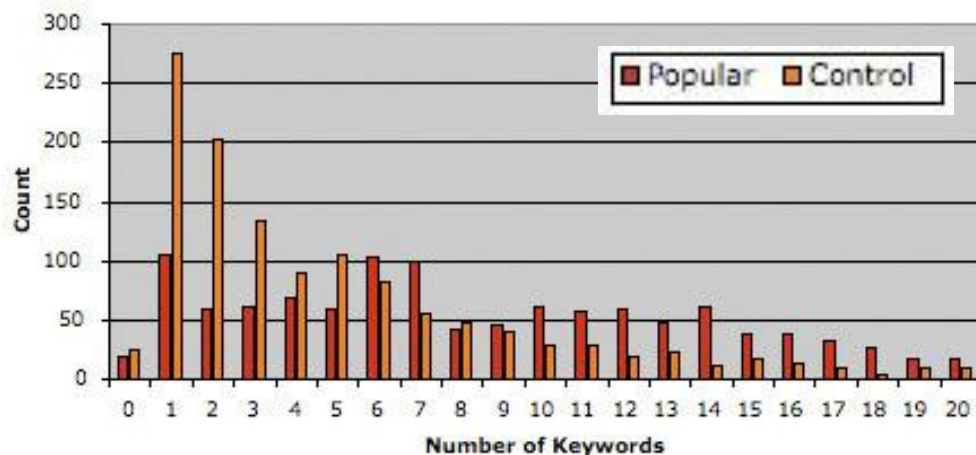


(検索したい映像)



- × 言語の多義性や主観による表記ゆれ
- × 注釈づけの労力

Number of Keywords for YouTube Videos



- 人気のある映像で1～14個程度
 - 通常の映像では1～6個程度
- **明らかに不完全！**

映像例示型検索： Query By Example (QBE)

検索したい映像の**サンプル**を与える



特徴量(色、エッジ、動き
など)に関する類似性

(検索したい映像)



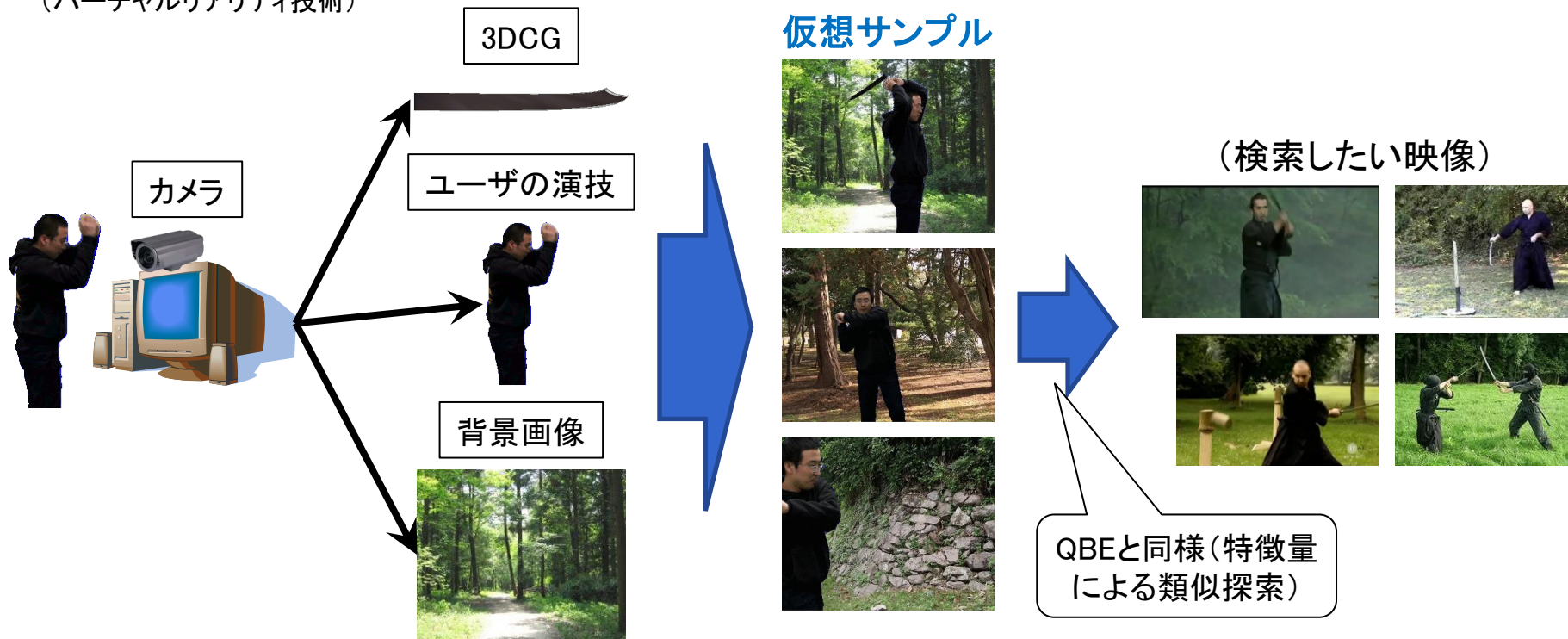
- 言語の多義性や主観による表記ゆれがない
- 特徴量は自動抽出可能なので、映像を注釈づけする必要がない

× サンプルがなければ、そもそも検索を行えない！

仮想映像例示型検索

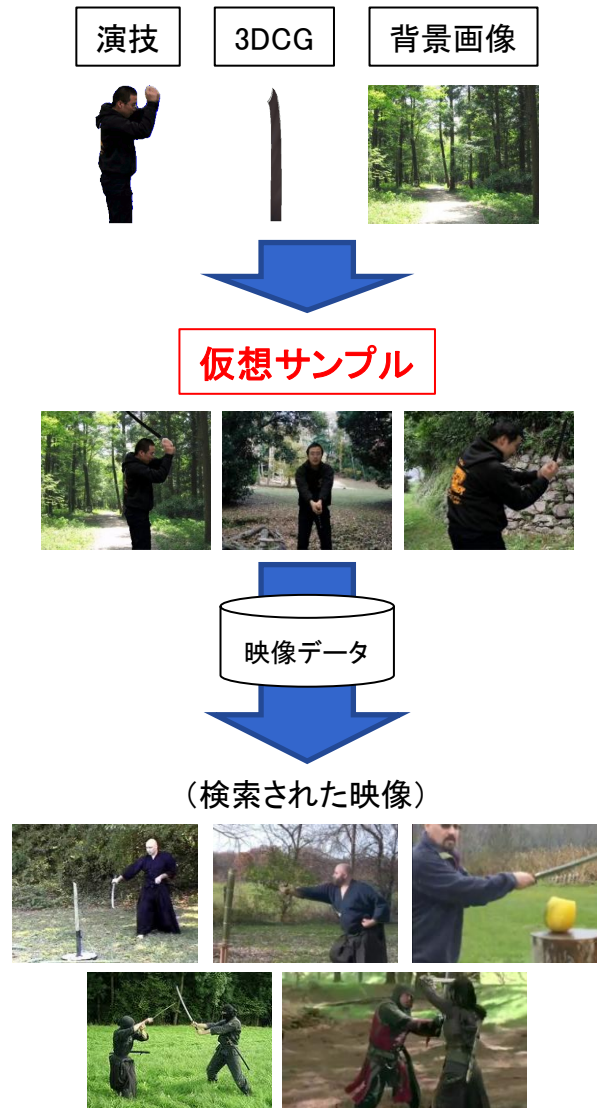
Query By Virtual Example (QBVE)

複合現実感技術を用いて、ユーザ自身が仮想的なサンプルを作成し検索を行う
(バーチャルリアリティ技術)



- あらゆるクエリに対する検索が可能(特に、特殊なクエリに有効)
- ジェスチャーによる映像データへの問い合わせ
→ 監視カメラ映像からの不審動作検出、動作障害を呈している人物検出
- 効率的なサンプルの生成(画像・映像分類、物体認識など)

取り組んだ3つの課題



1. 多種多様なクエリに関するリアリティのある
仮想サンプルの生成

2. 撮影技法・状況によって、見た目(特徴量)が
多様に異なる映像の網羅的検索

3. 反サンプルの自動収集

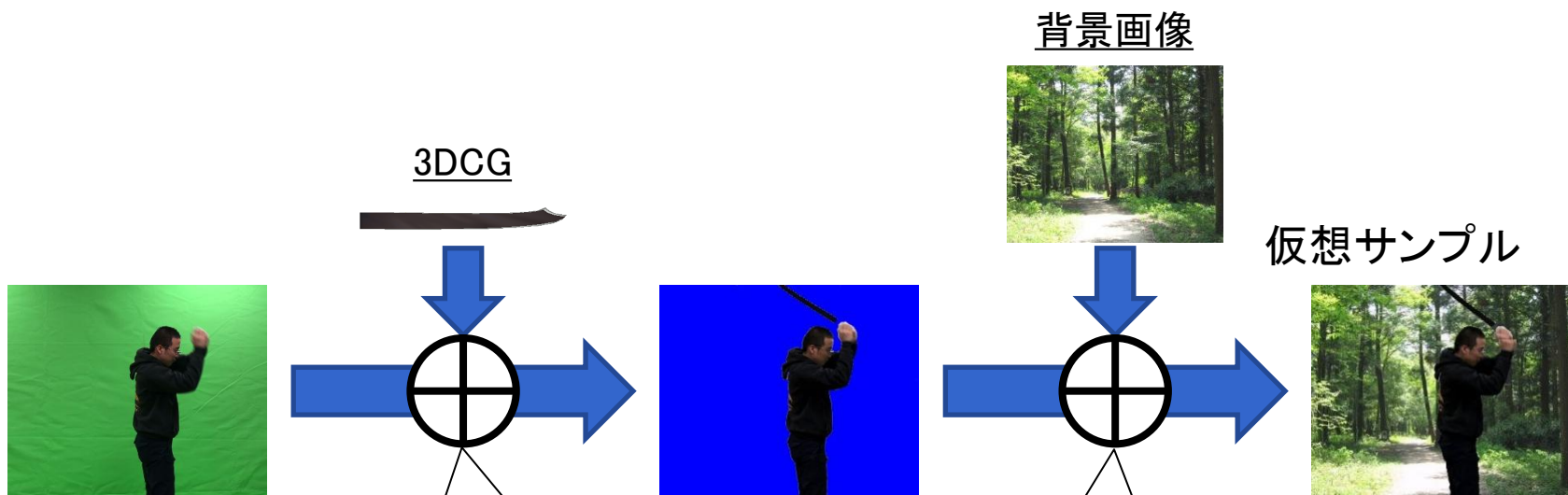
映像検索は、クエリに適合する映像としない映像を識別する分類問題

仮想サンプル

反サンプル

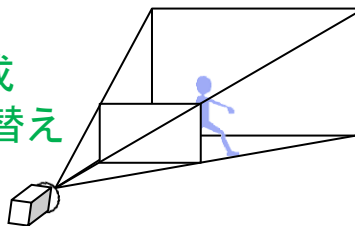
(与えられていない！)

仮想サンプル生成システム



磁気式マーカを用いた
マーカトラッキング

3DCGの合成
平面の切り替え



クロマキー合成



単純な背景差分だと、演技時に
生じる影の影響でうまくいかない

取り組んだ3つの課題



1. 多種多様なクエリに関するリアリティのある
仮想サンプルの生成

2. 撮影技法・状況によって、見た目(特徴量)が
多様に異なる映像の網羅的検索

3. 反サンプルの自動収集

映像検索は、クエリに適合する映像としない映像を識別する分類問題

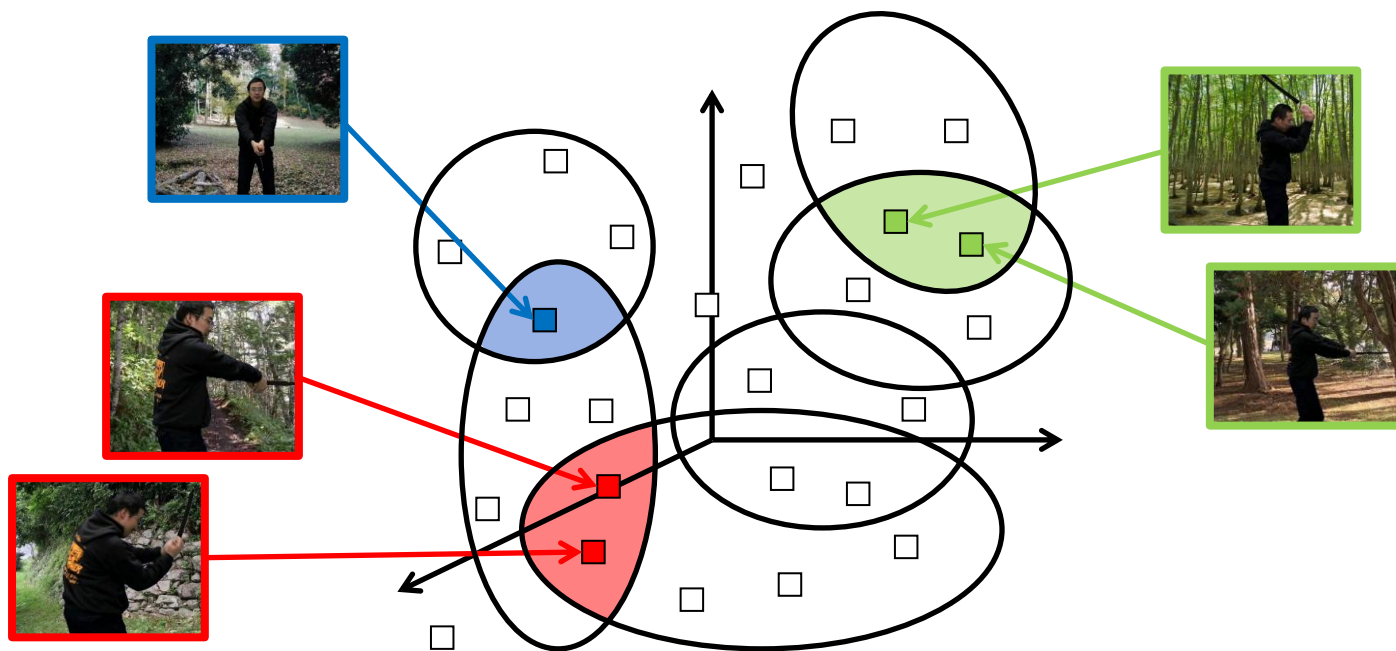
仮想サンプル

反サンプル

(与えられていない！)

ラフ集合理論を用いた検索ルール抽出

集合論に基づいて、曖昧、不確実なデータから、クラスを部分的に定義するルールを抽出



一旦、多数の分類器を構築しておいて、仮想サンプルの部分集合を特定可能な分類器の組み合わせを検索ルールとして抽出する

→ 精度を落とさずに、多様な映像を網羅できる

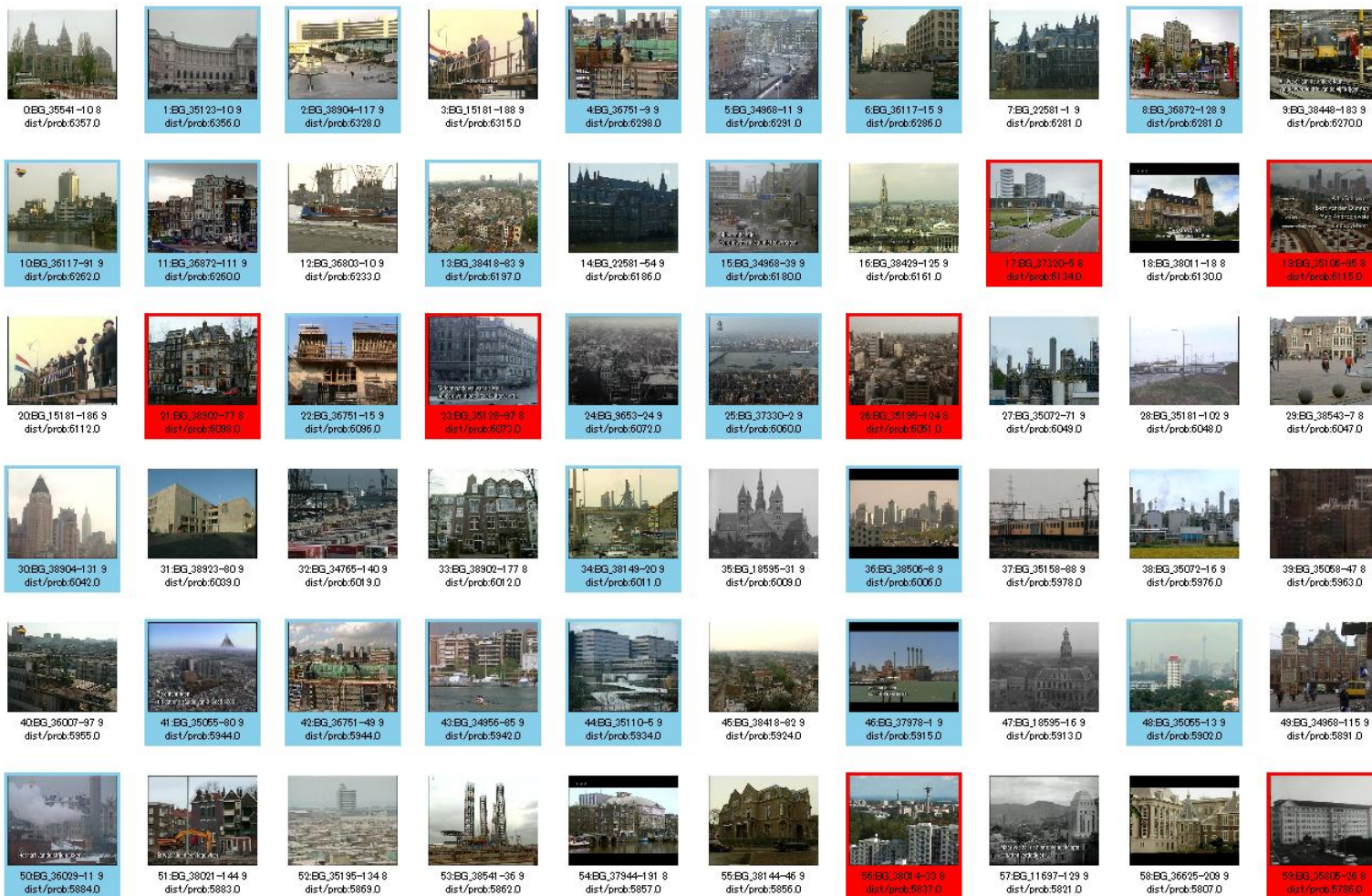
検索結果の例

NIST(米国標準技術局)後援の世界的な映像検索コンテスト

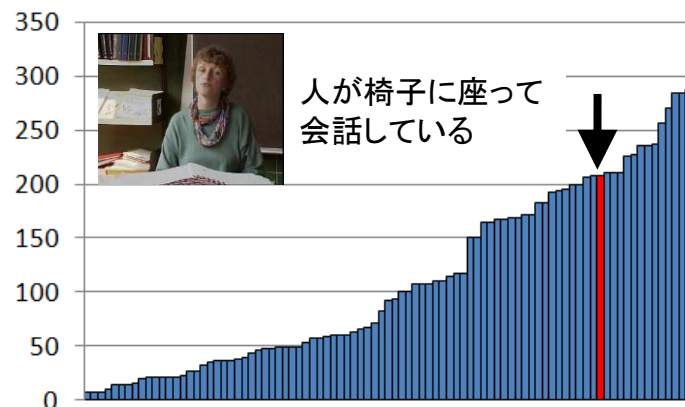
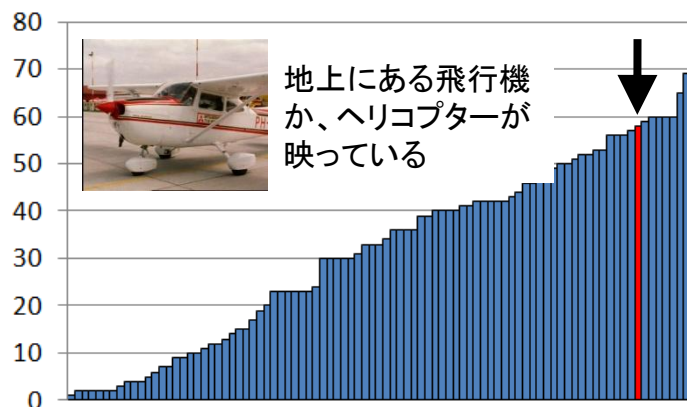
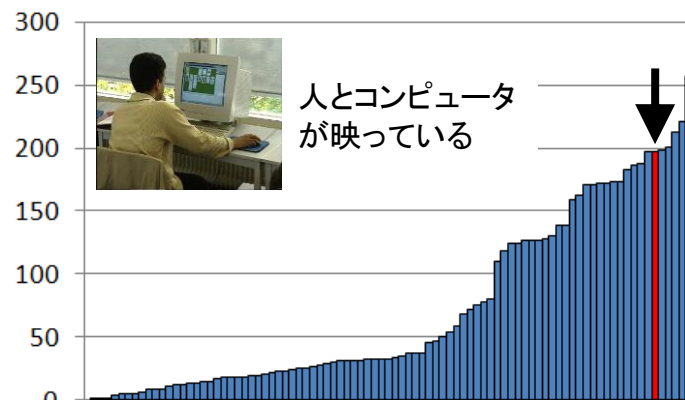
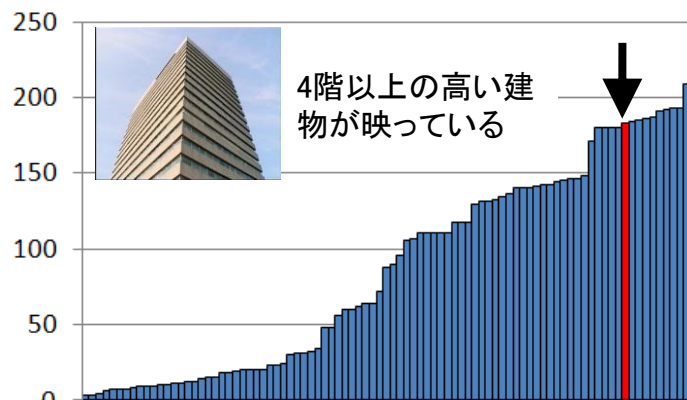
TRECVID 2009から提供された大規模ベンチマークデータ

- 36,106個の訓練用映像から、サンプルを選択
- 97,150個のテスト用映像から、クエリに適合すると判定された1,000映像を検索

クエリ: 4階建て以上の建物が映っている

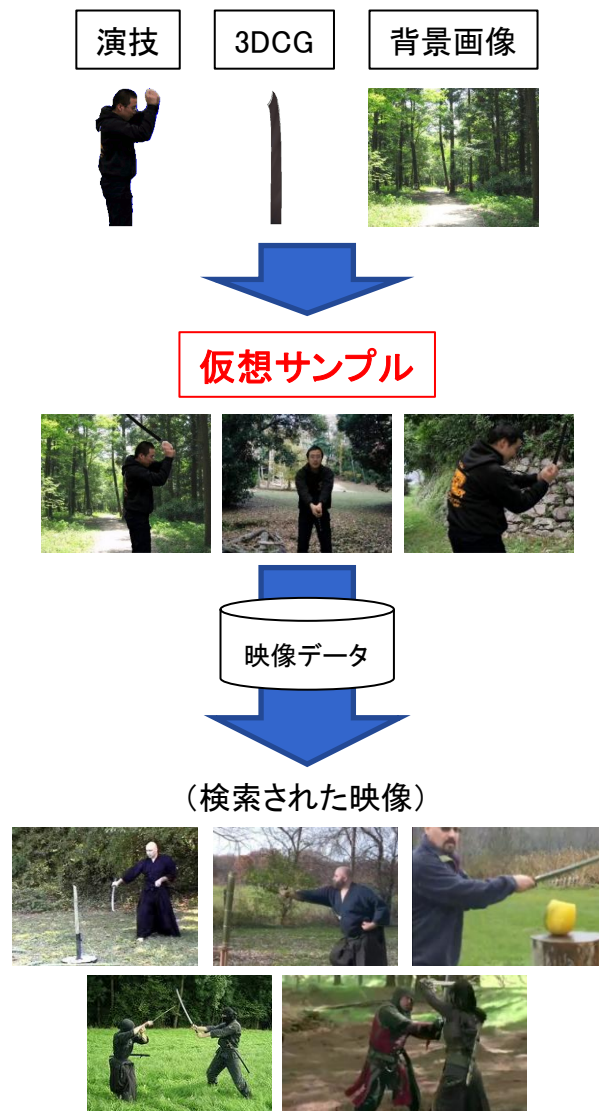


検索手法の性能評価



検索された上位1,000映像に含まれる正解数による検索手法のランキング
→ **世界でも上位に入る検索性能を実現！**

取り組んだ3つの課題



1. 多種多様なクエリに関するリアリティのある
仮想サンプルの生成

2. 撮影技法・状況によって、見た目(特徴量)が
多様に異なる映像の網羅的検索

3. 反サンプルの自動収集

映像検索は、クエリに適合する映像としない映像を識別する分類問題

仮想サンプル

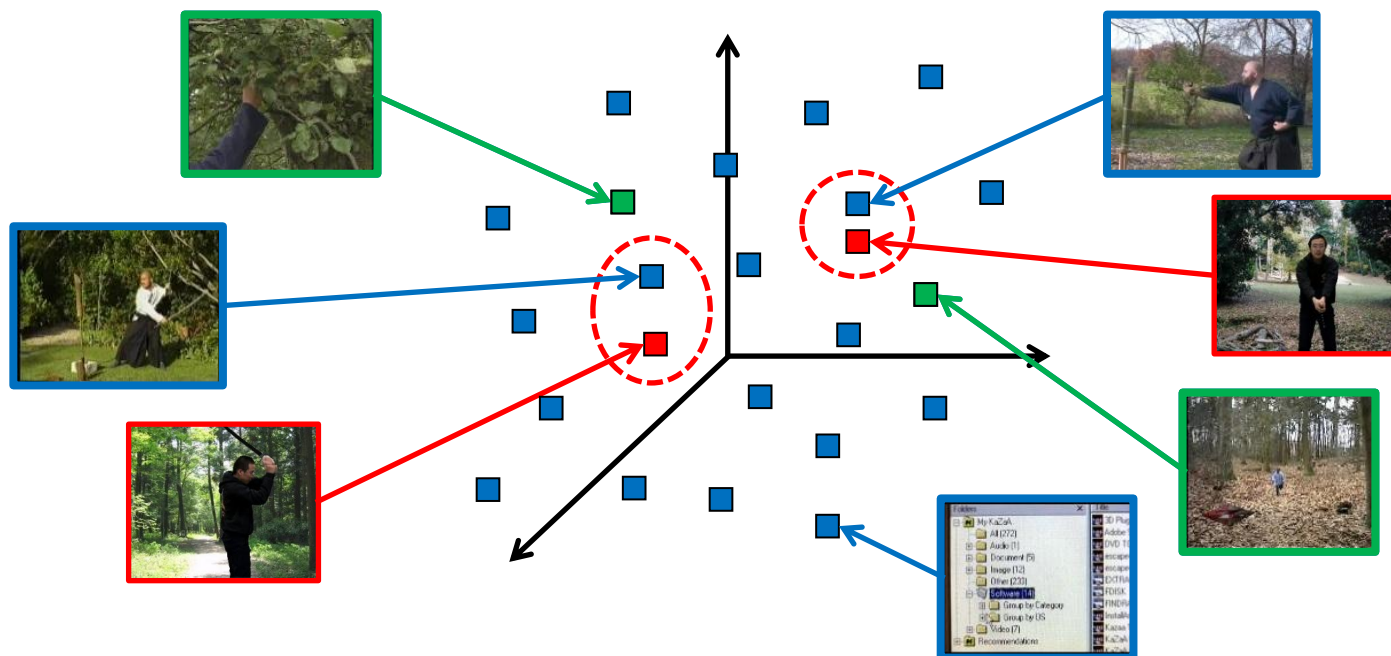
反サンプル

(与えられていない！)

部分教師つき学習を用いた 反サンプル収集

クラスが既知のサンプルとクラスが未知のサンプルから分類器を構築

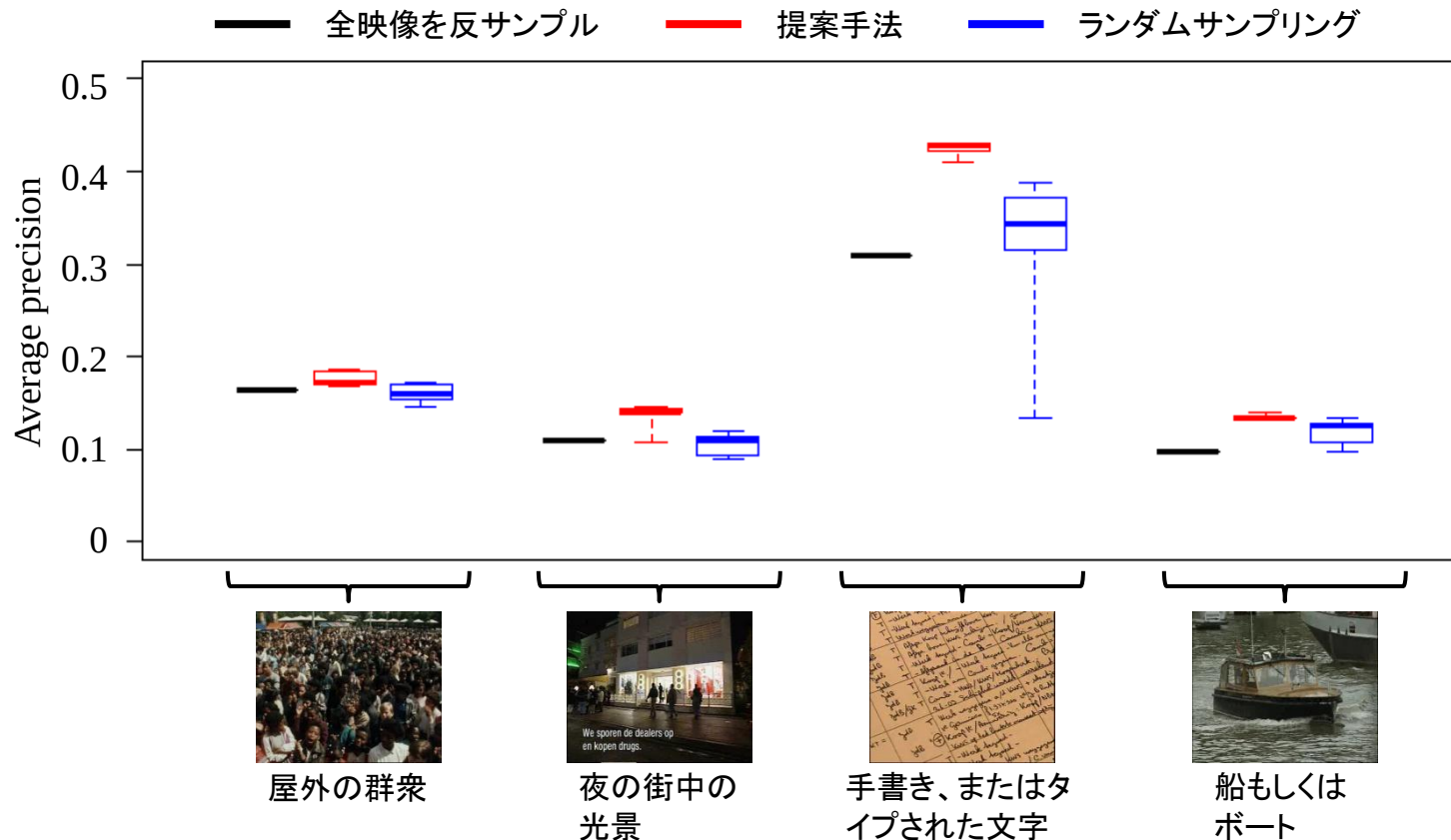
- 仮想サンプル: クエリに適合する(**クラス既知**)
- データベースの映像: クエリに適合するかもしれない(クラス未知)



特徴量とオブジェクト認識を併用して、「仮想サンプルと類似しているが、クエリには適合しない映像」を反サンプルとして収集

→ クエリに適合する映像としない映像の境界を定義するために有効

反サンプル収集手法の性能



現在、世界最先端の手法の多くで採用されているランダムサンプリングに勝る検索性能を実現！

オブジェクト認識用サンプル収集のためのオンラインゲーム

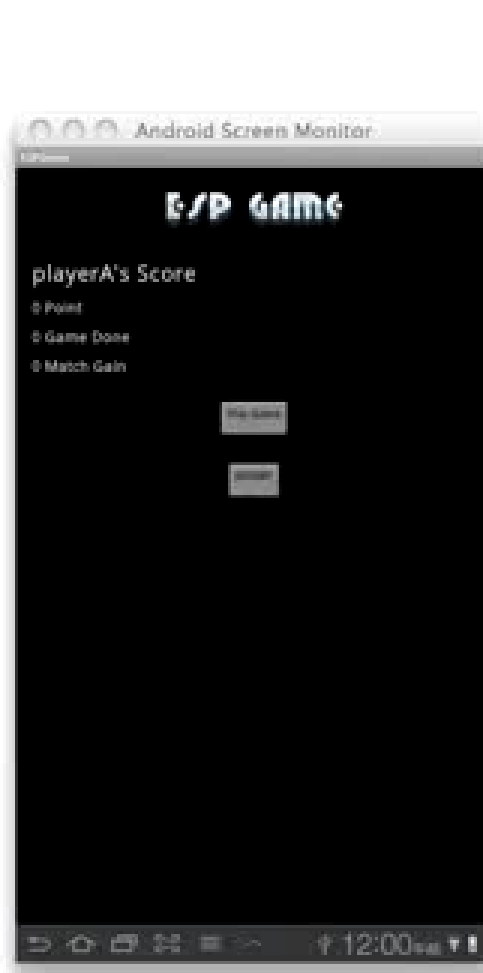
反サンプル手法の精度を向上させるためには、
オブジェクト認識の精度を上げることが重要！



各オブジェクトに関する大規模なサンプルが必要



インターネット上のユーザの助けを借りて、効率的
に大規模なサンプルを収集するための**オンライン
ゲーム**を開発(ゲームをクリアするために入力した
データが映像の注釈になる)



プラットフォーム: Android 1.6以上

仮想映像例示型検索の性能評価

□ 映像データ

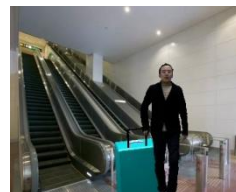
24,029映像 (TRECVID 2009 + YouTubeから収集した映像)

□ 検索クエリ

クエリ1: 人が刀を振っている(正解映像数123)



クエリ2: 人がスーツケースを引いている(正解映像数75)

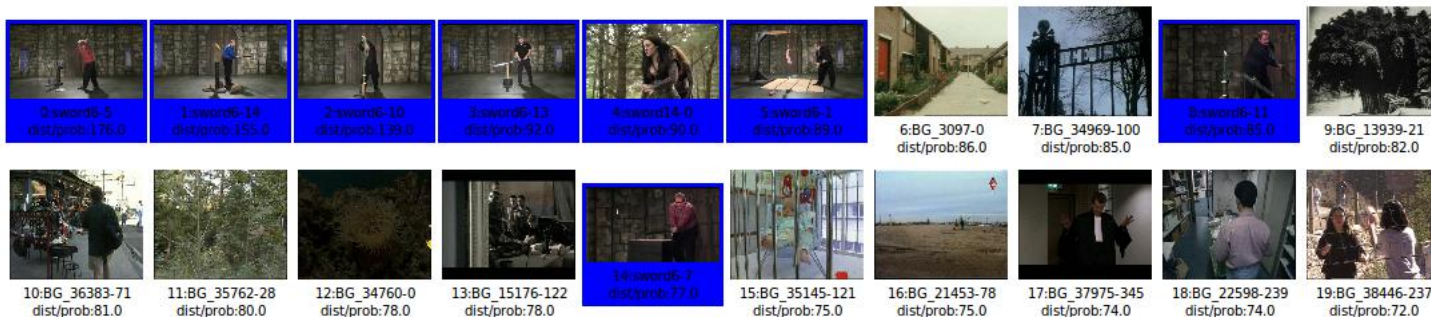


クエリ3: 人が電話している(正解映像数251)

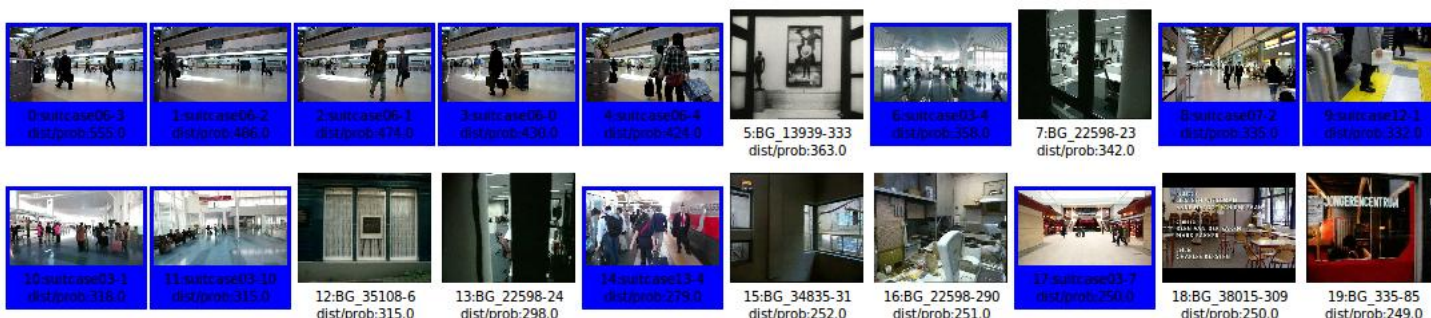


検索結果の例

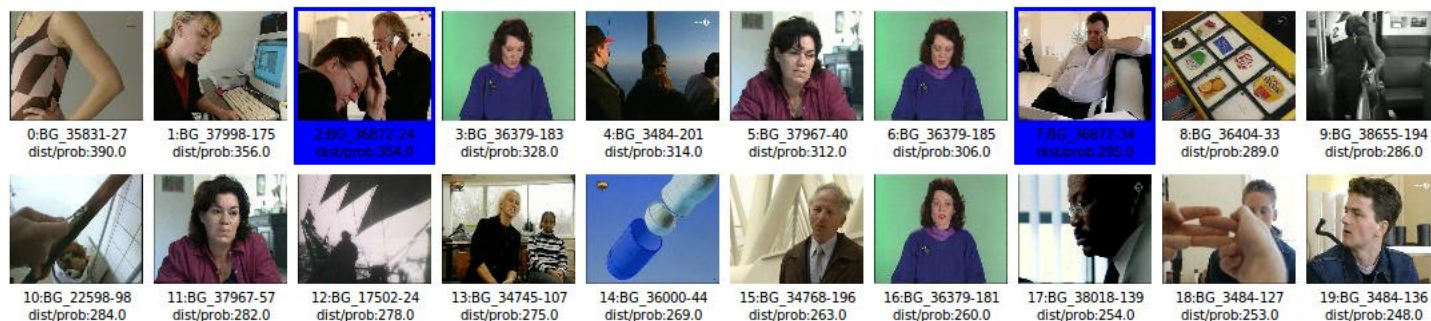
クエリ1:
人が刀を振っている
平均精度: 0.0929



クエリ2:
人がスーツケースを
引いている
平均精度: 0.2716

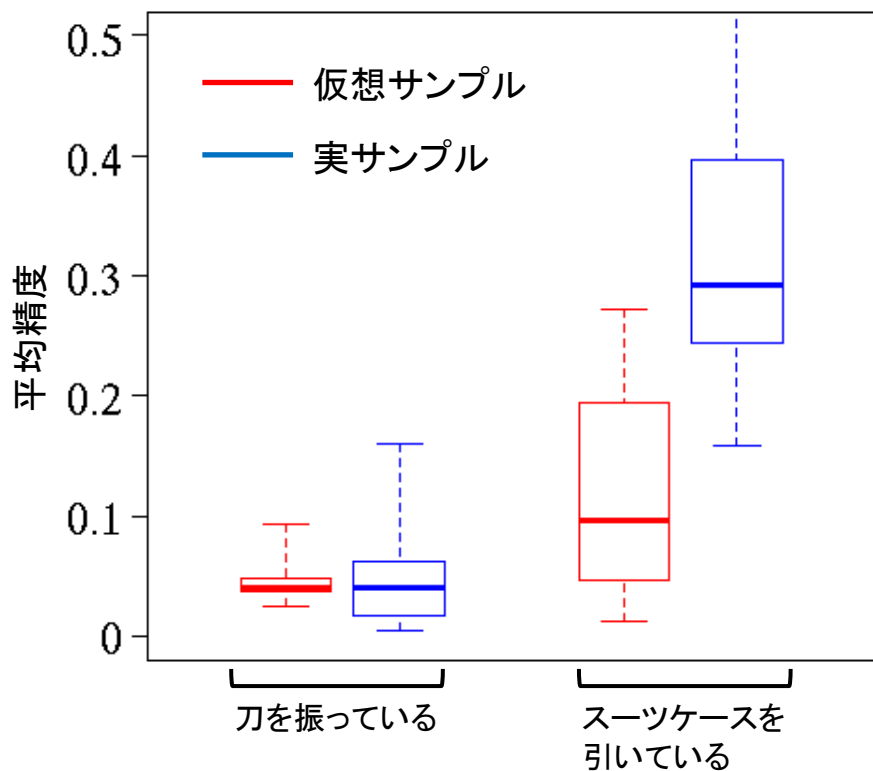


クエリ3:
人が電話している
平均精度: 0.0062



クエリ3以外では、精度の良い検索が行えている

実サンプルとの性能比較



実サンプルと比べると、性能が落ちている

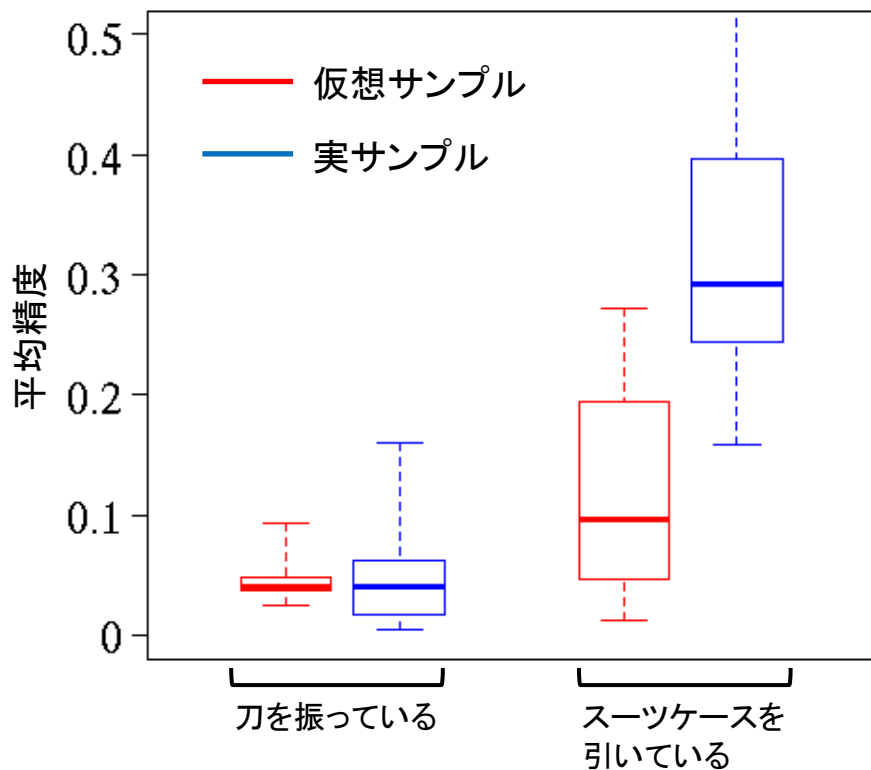
仮想サンプルにおける動きの少数性



- 仮想サンプルにおける軌跡数の平均: 5022.811
- 実サンプルにおける軌跡数の平均: 43150.54

カメラワーク、周囲の人物やオブジェクトの動きがない

実サンプルとの性能比較



実サンプルと比べると、性能が落ちている

仮想サンプルにおける動きの少数性

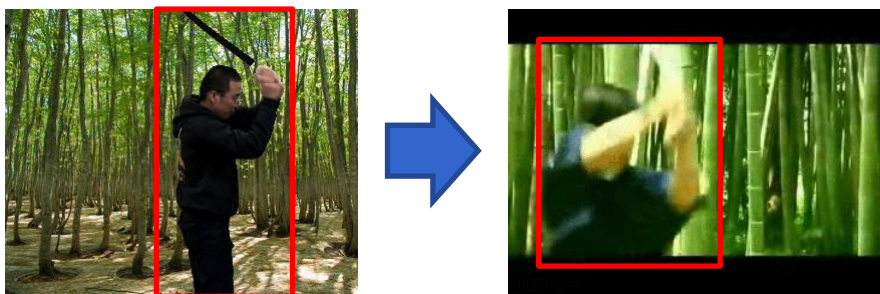


カメラの自動制御、他の人物との
インタラクションの自動生成は困難

少量の動きだけからの検索 (予備検討)

全領域から抽出された特徴量を考慮していることが問題

→ ユーザと3DCGの動きに類似したショット中の動領域を検出



「人が刀を振っている」に関する2,117映像を対象とした動領域検出による平均精度の比較

	DISP	HOG	MBH
全領域	0.036	0.061	0.067
動領域	0.082	0.129	0.224

111%～234%の
精度向上！

特徴量の統合、仮想・反サンプルの使用方法について、さらに検討が必要

まとめ

仮想映像例示型検索: ユーザ自身が仮想サンプルを作成し検索を行う
→ 実サンプルの入手困難さ・コストを解決し、あらゆるクエリに関する検索が行える

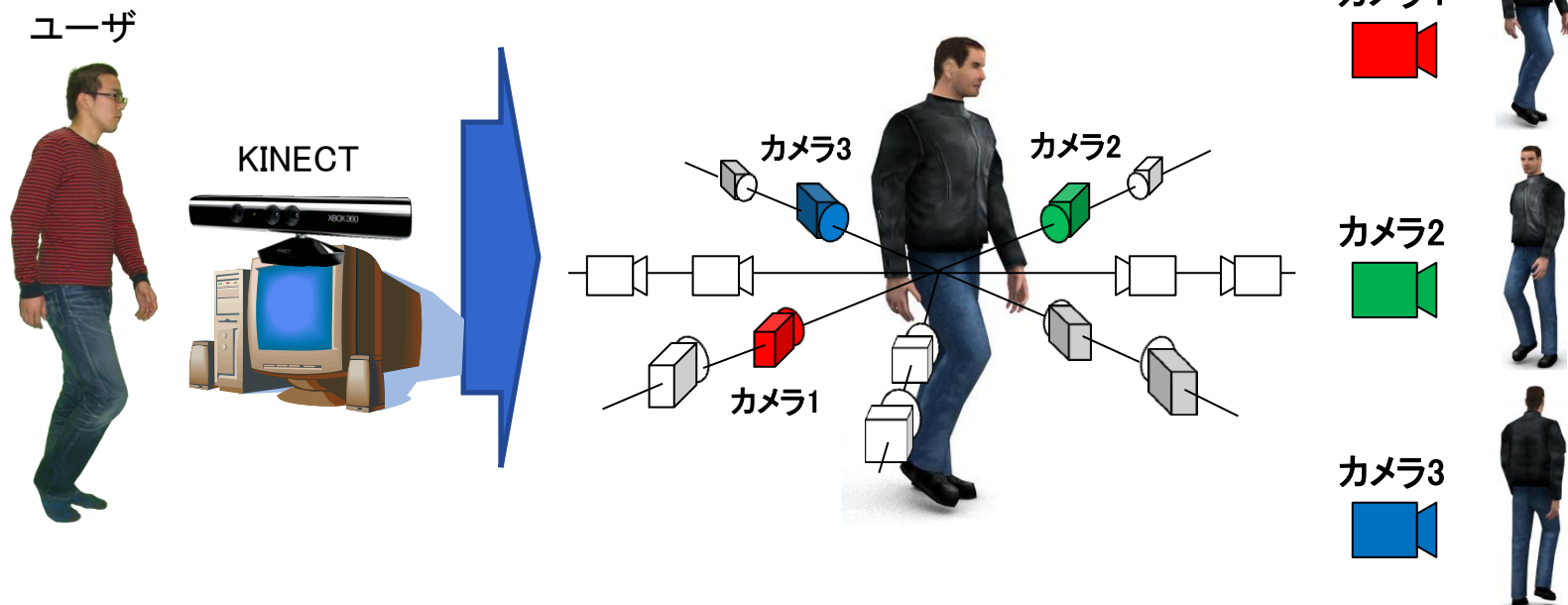
1. **仮想サンプル生成システム** (マーカトラッキング、合成平面、クロマキー合成)
多種多様なクエリに関するリアリティのある仮想サンプルを可能
2. **ラフ集合理論を用いた検索手法**
撮影技法・状況の異なる多様な映像を検索可能
3. **部分教師つき学習を用いた反サンプル収集手法**
クエリに適合する映像としない映像の境界を特徴づける反サンプルを収集可能

実現した点

- ◆ 仮想サンプルを用いた精度の良い検索 (改善の余地あり)
- ◆ 世界最先端の性能を有する検索手法
- ◆ 業界標準よりも、高精度な検索を可能とする反サンプルの収集

今後の課題

1. 動領域検出による少量の動きだけからの高精度な検索手法
2. 仮想サンプル生成システムの効率化
(仮想・現実空間の位置合わせ、クロマキー合成のための大規模なセット)
→ KINECTによる3Dモデルアニメーションを用いた仮想サンプル生成
 - 1回の演技で複数視点から仮想サンプルを生成可能
 - 仮想・現実空間の位置合わせは不要



研究成果(1/2)

(SCOPEの委託研究に基づく成果、及び関連する成果を全て含む)

学術論文 4件

- [1] Kimiaki Shirahama, Yuta Matsuoka and Kuniaki Uehara: “Hybrid Negative Example Selection Using Visual and Conceptual Features”, Multimedia Tools and Applications, (採録決定)
- [2] 白浜公章, 松岡悠太, 上原邦昭: “ラフ集合理論を用いたクエリの帰納的定義に基づく例示映像検索”, 映像情報メディア学会誌, Vol. 66, No. 5, pp. J124 – J135 (2012)
- [3] Kimiaki Shirahama and Kuniaki Uehara: “Constructing and Utilizing Video Ontology for Accurate and Fast Retrieval”, International Journal of Multimedia Data Engineering and Management (IJMDEM), Vol. 2, No. 4, pp. 59–75 (2011)
- [4] Kimiaki Shirahama, Yuta Matsuoka and Kuniaki Uehara: “Event Retrieval in Video Archives Using Rough Set Theory and Partially Supervised Learning”, Multimedia Tools and Applications, Vol. 57, No. 1, pp. 145–173 (2011)

国際会議 15件

- [1] Kimiaki Shirahama, Kuniaki Uehara and Marcin Grzegorzec: “Examining the Applicability of Virtual Reality Technique for Video Retrieval”, Proc. of the 10th Workshop on Content-Based Multimedia Indexing (CBMI 2012), pp. 211–216 (2012)
- [2] Kimiaki Shirahama, Lin Yanpeng and Kuniaki Uehara: “Kobe University at TRECVID 2011 Semantic Indexing and Multimedia Event Detection”, Proc. of TREC Video Retrieval Evaluation (TRECVID) 2011 Workshop, pp. 243–249 (2011)
- [3] Kimiaki Shirahama and Kuniaki Uehara: “Utilizing Video Ontology for Fast and Accurate Query-by-Example Retrieval”, Proc. of the IEEE International Workshop on Semantic Multimedia (ICSC-SMM 2011), pp. 395–402 (2011)
- [4] Kimiaki Shirahama and Kuniaki Uehara: “Effectiveness of Video Ontology in Query by Example Approach”, Proc. of the 2011 International Conference on Active Media Technology (AMT 2011), pp. 49–58 (2011)
- [5] Kimiaki Shirahama and Kuniaki Uehara: “Query by Virtual Example: Video Retrieval Using Example Shots Created by Virtual Reality Techniques”, Proc. of the 6th International Conference on Image and Graphics (ICIG 2011), pp. 829–834 (2011)
- [6] Kimiaki Shirahama, Yuta Matsuoka and Kuniaki Uehara: “Video Event Retrieval from a Small Number of Examples Using Rough Set Theory”, Proc. of the 17th International Conference on Multimedia Modeling (MMM 2011), pp. 96–106 (2011)
- [7] Kimiaki Shirahama, Lin Yanpeng, Yuta Matsuoka and Kuniaki Uehara: “Query by Example for Large-Scale Video Data by Parallelizing Rough Set Theory Based on MapReduce”, Proc. of the 2010 International Conference on Science and Social Research (CSSR 2010), pp. 390–395 (2010)
- [8] Kimiaki Shirahama, Yuta Matsuoka and Kuniaki Uehara: “Query by Few Video Examples Using Rough Set Theory and Partially Supervised Learning”, Proc. of the 5th International Conference on Semantic and Digital Media Technologies (SAMT 2010), pp. 190–191 (2010)
- [9] Kimiaki Shirahama and Kuniaki Uehara: “Video Retrieval from Few Examples Using Ontology and Rough Set Theory”, Proc. of the 2nd Workshop on Semantic Multimedia Database Technologies (SMDT 2010), pp. 5–16 (2010)
- [10] Kimiaki Shirahama and Kuniaki Uehara: “Example-based Event Retrieval in Video Archive Using Rough Set Theory and Video Ontology”, Proc. of the 10th International Workshop on Multimedia Data Mining (MDM/KDD 2010), pp. 42–48 (2010)
- [11] Kimiaki Shirahama, Chieri Sugihara, Yuta Matsuoka and Kuniaki Uehara: “Query-based Video Event Definition Using Rough Set Theory and Video Prototypes”, Proc. of IS&T/SPIE Electronic Imaging Multimedia Content Access: Algorithms and Systems IV, 7540B–41 (2010)

研究成果(2/2)

(SCOPEの委託研究に基づく成果、及び関連する成果を全て含む)

- [12] Kimiaki Shirahama, Chieri Sugihara and Kuniaki Uehara: “Query-based Video Event Definition Using Rough Set Theory and High-dimensional Representation”, Proc. of the 16-th International Conference on Multimedia Modeling (MMM 2010), pp. 358–369 (2010)
- [13] Kimiaki Shirahama, Chieri Sugihara, Kana Matsumura, Yuta Matsuoka and Kuniaki Uehara: “Mining Event Definitions from Queries for Video Retrieval on the Internet”, Proc. of the First International Workshop on Internet Multimedia Mining (IMM 2009), pp. 176–183 (2009)
- [14] Kimiaki Shirahama, Chieri Sugihara, Yuta Matsuoka, Kana Matsumura and Kuniaki Uehara: “Kobe University at TRECVID 2009 Search Task”, Proc. of TREC Video Retrieval Evaluation (TRECVID) 2009 Workshop, pp. 76–84 (2009)
- [15] Kimiaki Shirahama, Chieri Sugihara, Yuta Matsuoka and Kuniaki Uehara: “Query-based Video Event Definition Using Rough Set Theory”, Proc. of the First ACM International Workshop on Events in Multimedia (EiMM 2009), pp. 9–15 (2009)

国内研究会 5件

- [1] 白浜公章, 河村智美, 上原邦昭: “バーチャルリアリティ技術を用いた映像例示型検索の性能評価”, 電子情報通信学会技術研究報告, Vol. 111, No. 499, pp. 183–188 (PRMU2011–269) (2012)
- [2] 白浜公章, 上原邦昭: “Query by Virtual Example: 映像検索モデル構築のためのバーチャルリアリティ技術による仮想サンプル映像の生成”, 電子情報通信学会技術研究報告, Vol. 111, No. 193, pp. 139–144 (PRMU2011–73) (2011)
- [3] 白浜公章, 松岡悠太, 上原邦昭: “ラフ集合理論に基づく少数のサンプル映像からの多様性に対応可能な検索ルール抽出手法”, 電子情報通信学会技術研究報告, Vol. 110, No. 414, pp. 129–134 (PRMU2010–228) (2011)
- [4] 松岡悠太, 白浜公章, 上原邦昭: “映像例示型検索における高次元特徴量に対する負例選択”, 電子情報通信学会技術研究報告, Vol. 110, No. 414, pp. 123–128 (PRMU2010–227) (2011)
- [5] 杉原ちえり, 白浜公章, 上原邦昭: “ラフ集合理論を用いたクエリー映像からのイベント検索モデルの導出”, 平成21年度 情報処理学会関西支部大会, C-06 (2009)

関連するWebページ

- [1] http://www.ai.cs.scitec.kobe-u.ac.jp/~kimi/image/samt_poster.png
ラフ集合理論と部分教師つき学習を用いた映像検索手法に関するポスター
- [2] <http://www-nlpir.nist.gov/projects/tvpubs/tv9.slides/kobe.slides.pdf>
TRECVID 2009における口頭発表で使用したスライド



ご清聴、ありがとうございました