

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 (092107005)

Gesture-based Video Archive Retrieval Utilizing Mixed Reality Technique

研究代表者

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）

Kimiaki Shirahama Graduate School of Economics, Kobe University (Muroran Institute of Technology)

研究期間 平成 21 年度～平成 23 年度

概要

映像検索では、検索モデルを構築する際、クエリに適合する映像のサンプルを収集するのに多大なコストを要する点が大きな問題となってくる。そこで、本研究では、映像検索に“複合現実感技術”を導入して、ユーザ自身が任意のクエリに関する“仮想サンプル”を生成し検索を行う“Query-By-Virtual-Example (QBVE)”という革新的な映像検索システムを開発した。概要としては、カメラの前でのユーザの演技、3DCG、背景画像を合成して仮想サンプルを生成し、仮想サンプルに類似した特徴量を含む映像を検索する。

Abstract

One of the biggest problems in video retrieval is the expensive cost of collecting examples, which are used to build a retrieval model for a query. To overcome this, a novel video retrieval system, called *Query-By-Virtual-Example* (QBVE), was developed where ‘virtual’ examples for a query are created based on *mixed reality technique*. Roughly speaking, this system creates virtual examples by synthesizing a user’s gesture in front of the video camera, 3D object and background images, and retrieves videos which are visually similar to virtual examples.

1. まえがき

YouTube やニコニコ動画に代表される動画共有サイトの普及によって、現在の映像配信形態は、従来のブロードキャスティングやナローキャスティングよりも、はるかに多くの映像を選択・視聴できるようになったという意味で“シンキャスト”と呼ばれている。こうした状況の中で、映像の内容を自動解析し、クエリに適合する映像を効率的に検索する“映像検索技術”に多大な期待が寄せられている。

映像検索では、色、エッジ、動きといった特徴量に基づいて、クエリに適合する映像を特定するための検索モデルを構築する。言い換えると、検索モデルの構築には、どのような映像がクエリに適合するかを教示するための“サンプル”が必要になってくる。しかしながら、多種多様なクエリに関するサンプルを収集するには多大な人的コストを要する（もしくは、非現実的である）。特に、「山中で刀を振っている」、「ベンチで本を読んでいる」といった“特殊”なクエリに関するサンプルは容易に見つからない。また、ユーザから与えられたサンプルと、類似した特徴量を含む映像を検索する“Query-By-Example (QBE)”という代表的な手法があるが、サンプルが手元になれば検索自体が行えない。そこで、本研究では、“複合現実感技術”を導入して、ユーザ自身が任意のクエリに関する“仮想サンプル”を生成し検索を行う“Query-By-Virtual-Example (QBVE)”という革新的なシステムを開発した。

2. 研究内容及び成果

図 1 に、開発した QBVE の概要を示す。まず、図 1(a)に示すように、カメラの前でのユーザの演技、3DCG、背景画像を合成して、クエリに関する仮想サンプルを生成する。そして、図 1(b)のように、仮想サンプルから検索モデルを生成し、データベースからクエリに適合する映像を検索する。以下では、QBVE の開発に当たって取り組んだ 3 つの研究テーマについて概説する。

1. 仮想サンプル生成システムの開発：多種多様なクエリに関する仮想サンプルを生成するには、ユーザの演技に同

期して 3DCG を合成する必要がある。そこで、図 1(a)に示すように、“マーカトラッキング”技術を用いて、トラッキングされたマーカの位置に 3DCG を合成している。特に、演技中に、カメラ映像からマーカが隠れてもトラッキングできるように、磁気式のマーカトラッキングシステムを採用している。最終的に、クロマキー合成を行ってユーザと 3DCG の領域を切り出し、背景画像と合成して、仮想サンプルを生成している。

2. ラフ集合理論を用いた映像検索手法の開発：仮想サンプルを用いて検索モデルを構築するに当たって、「撮影技法・状況によって、同一のクエリに適合する映像でも、見た目（特徴量）が大きく異なってくる」という映像の“多様性”が問題になってくる。そこで、“ラフ集合理論”を用いて、クエリに適合する多様な映像を検索可能な手法を開発した。ラフ集合理論とは、集合論に基づいて、曖昧、不確実なデータから、クラスに対する「粗い」ルールを抽出する分類手法である。ここで、粗いとは、クラスに属するサンプルの全集合ではなく、部分集合を用いるという意味である。すなわち、図 1(b)のように、仮想サンプルの部分集合を特定可能な複数の検索ルールを抽出して、多様な映像を網羅できるようにしている。

3. 部分教師つき学習を用いた反サンプル収集手法の開発：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とみなせる。つまり、高精度な検索を実現するには、クエリに適合する映像の見本となる仮想サンプルだけでなく、適合しない映像の見本となる“反サンプル”が必要になってくる。そこで、仮想サンプルに基づいて、データベース中の映像から反サンプルを自動収集する手法を開発した。概要としては、“部分教師つき学習”という、対象クラスに関するサンプルとクラスラベルを付与されていないサンプルから分類モデルを導出する手法を拡張して、反サンプルを収集している。特に、仮想サンプルと映像間での特徴量とオブジェクト認識結果の類似性から、図 1(b) 下部の黒枠の映像のような、「仮想サンプルに類似しているが、クエリには適合しない反サンプル」を収

a) 仮想サンプル生成

→ ユーザの演技、3DCG、背景画像を合成



b) 映像検索

□ ラフ集合理論を用いて、撮影技術・状況の異なる多様な映像を網羅可能な検索ルールを抽出

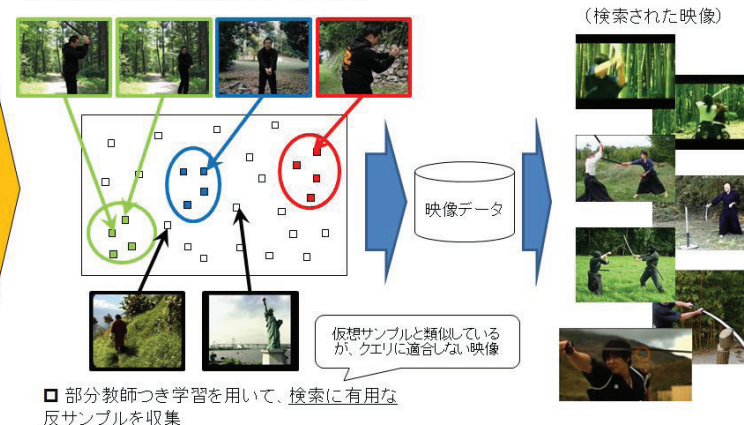


図 1：開発した QBVE の概要

集している。このような反サンプルは、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ために有用である。

以下では、開発した QBVE の性能評価結果について概説する。図 2 に、QBVE の検索結果を示す。各クエリに関して、左側に生成した 4 つの仮想サンプル、右側に検索された上位 10 個の映像を示している。ここで、青色の背景の映像が、実際にクエリに適合していた映像である。図 2 から、実映像と比べると、見た目が多少不自然な仮想サンプルを用いても、高精度な検索が行えていることが分かる。ただし、仮想サンプルには、カメラワーク、周囲の人物やオブジェクトの動きが含まれておらず、動きに関する情報が少ないということが原因で、実映像からのサンプルに比べて、検索性能が落ちてしまうことが分かっている（実映像のサンプルを用いれば、図 2 よりもさらに高精度な検索が行える）。この点に関して、最先端の複合現実感技術を用いても、カメラワークや周囲のオブジェクトの動きを自動で付与することは困難である。そこで、仮想サンプルに含まれる少量の動きだけから高精度な検索を行うために、ユーザと 3DCG の動きに類似した映像中の動領域を検出して検索を行う手法を現在開発中である。

◆ クエリ1: 刀を振っている (平均精度: 0.093)



◆ クエリ2: スーツケースを引いている (平均精度: 0.272)



図 2：仮想サンプルによる検索結果の例

図 3 に、ラフ集合理論を用いた映像検索手法と、TRECVID 2009 という世界的な映像検索コンテストで開発された手法との性能の比較結果を示す(実映像からのサンプルを使用)。図 3 の各グラフでは、右から左へと性能の高い順に検索手法がランク付けされており、矢印で示す本手法は、世界でも上位の検索性能を実現できている。

図 4 に、開発した手法と他手法で収集された反サンプルを用いた場合の検索性能の比較結果を示す。図 4 から、開

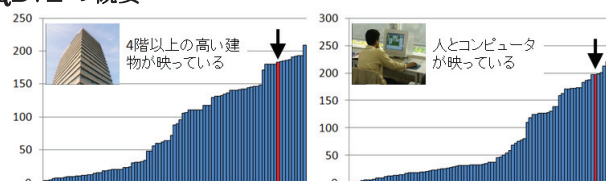


図 3：TRECVID 2009 で開発された検索手法との比較

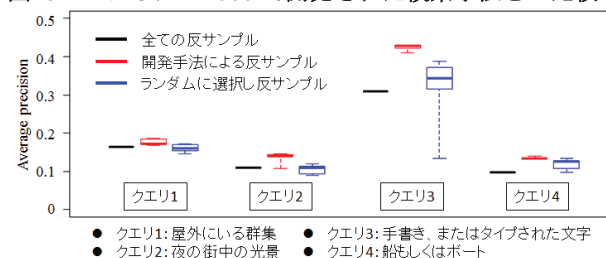


図 4：開発手法による反サンプルによる検索性能の評価
開発手法によって収集された反サンプルを用いた方が、検索性能が高いことが分かる。特に、世界最先端の検索手法の多くは、ランダムに選択した映像を反サンプルとしており（図 4 の青で示されている手法）、これよりも高い検索性能を実現できている点に非常に意義がある。

3. むすび

本研究では、映像検索に複合現実感技術を導入して、ユーザ自身がクエリに関する仮想サンプルを生成し検索を行う QBVE を開発した。今後は、QBVE の実用化を目指して、更なる検索精度の改善、及び仮想サンプル生成の効率化に取り組む予定である。

【誌上発表リスト】

- [1] Kimiaki Shirahama, Yuta Matsuoka and Kuniaki Uehara, "Hybrid Negative Example Selection Using Visual and Conceptual Features", Multimedia Tools and Applications, (Accepted)
- [2] 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 (2012)
- [3] Kimiaki Shirahama, Yuta Matsuoka and Kuniaki Uehara, "Constructing and Utilizing Video Ontology for Accurate and Fast Retrieval", International Journal of Multimedia Data Engineering and Management, Vol. 2, No. 4, pp. 59-75 (2011)