

研究報告書

「映像分析による知識の抽出と、その利用による新たな映像合成」

研究タイプ: 通常型

研究期間: 平成 21 年 10 月～平成 25 年 3 月

研究者: 岡部 誠

1. 研究のねらい

本研究の目的は、大量の映像データを統計的に分析して知識を抽出し、ユーザがその知識を利用することで容易な、直感的な映像製作を可能にする環境を作ることである。現在、撮影技術の進歩とインターネットの普及により、大量の映像データに容易にアクセスできる。また、テレビ局、CG プロダクションでも、独自に映像データベースを作っており、それらの加工・編集を通して、独自に新たな映像を作っている。しかし、そこで実際に行われている作業を見てみると、タグ付けされた言葉による映像の検索、見つけた映像の切り貼りと言った、技術的にはごく簡単なことしか行われていない。

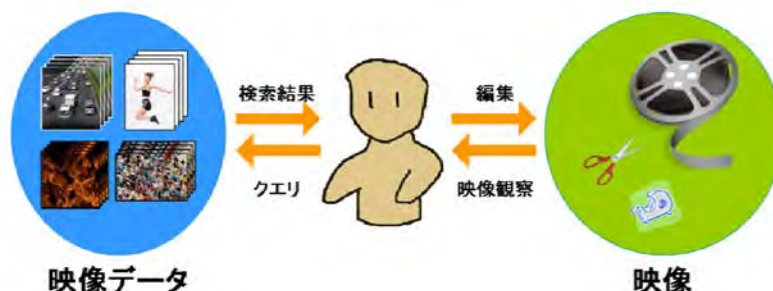


図: 現在の映像データベースの使われ方。ユーザは編集中の映像を観察し、そこに加えたい内容の映像を言葉によって検索する。システムはタグ情報に基づき検索、結果を返し、ユーザはその映像を切り貼りして編集を続ける。ユーザは、以上のループを繰り返す。

本研究では、人手によるタグ付け作業に頼らず、大量の映像データの中から抽象的な知識を自動的に抽出し、それらを効率よく利用するためのユーザインタフェースの提案を行う。今までの切り貼りを越えた、映像同士の混ぜ合わせを可能にすることで、より高度な映像製作を可能にする。映像製作技術と「大量の映像データベースの創造的な再利用」の間の橋渡しをめざす。

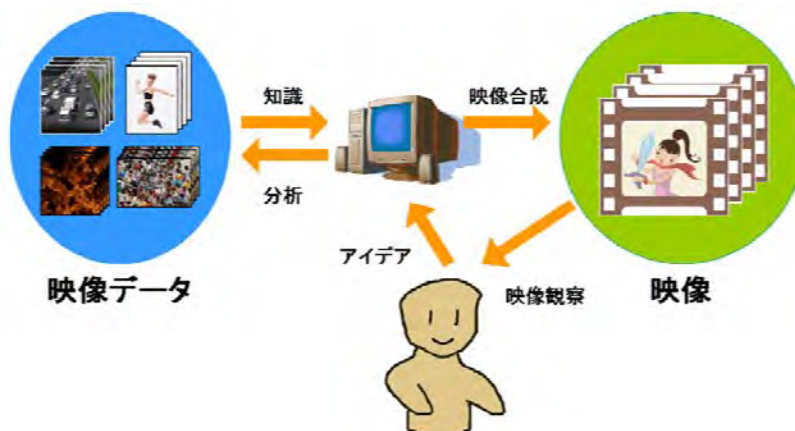


図: 提案手法による映像データベースの使われ方。コンピュータはあらかじめ、大量の映像データを分析し抽象的な知識の抽出を行う。ユーザは編集中の映像を観察し、そこに欲しい内容を、見た目や

動きの指定による、直感的なユーザインタフェースでコンピュータに指定する。コンピュータは抽出済みの知識に基づき映像合成をし、結果を更に編集する。

2. 研究成果

(1) 概要

大量の映像データを統計的に分析して知識を抽出し、ユーザが直感的な映像製作を可能にする環境を作ることがテーマである。このテーマに沿って、「ビデオデータベースを用いて流体画像をアニメーションさせる手法」、「動画鑑賞のためのスケッチインタフェース」、「入力音楽をあたかもバンド演奏しているかのように見える動画の生成技術」、「アマチュア演技者の動作を格好良く見せるように動画を編集する手法」、「勾配画像処理に基づく動画からの流体抽出」等の基本的な技術の研究開発を行った。

「ビデオデータベースを用いて流体画像をアニメーションさせる手法」は局所的な動画検索である。ユーザが絵と簡単なアノテーションという入力力でコンピュータにアイデアを与えると、コンピュータが自動的にアニメーションを制作してくれる。

「動画鑑賞のためのスケッチインタフェース」は画像検索と軌跡検索の組み合わせで、新しい映像閲覧のためのユーザインタフェースの研究である。ユーザはスケッチというシンプルなアノテーションを与えるだけで、予め動画を解析したコンピュータがユーザの見た動画の即座に検索してくる事ができる。

「入力音楽をあたかもバンド演奏しているかのように見える動画の生成技術」は動画の編集技術の提案である。映像が音楽の楽しみ方に与える影響について、引き続き分析と実験を行っている。

「アマチュア演技者の動作を格好良く見せるように動画を編集する手法」は、アマチュア動画とプロ動画の両方の動作を分析して抽出し、前者を後者に合わせて編集する技術である。

「勾配画像処理に基づく動画からの流体抽出」は爆発や煙、炎等の半透明領域が広範囲に渡る動画において、その流体を抽出する技術である。リッチな動画データベースを構築するのに助ける技術である。

2次元動画データベースの解析と合成の研究に従事したと共に、2次元動画の限界も見えている。今後、上記の2次元流体動画の応用研究を、3次元流体動画に拡張していく。

(2) 詳細

① . ビデオデータベースを用いて流体画像をアニメーションさせる手法

「論文 1, 学会発表 1, 11, 受賞 1, 2」

下図(a)はキャンプファイアの写真, (b)は 400 年くらい前に描かれた油絵である. 1枚の絵であれば, (b)の油絵のように, アーティストは独自のタッチで詳細に描き込むことができる. しかし, そのようなシーンのアニメーションを作ろうとすると, 流体シミュレーションを走らせて3次元 CG で再現するのも難しいし, かと言って何枚も同じような絵を描いてパラパラアニメ化するのもとても大変である. そこで, 1枚の絵の持つ雰囲気, 質感をなるべく損なわないようにアニメーションを作るため, 我々はビデオ検索技術と画像合成技術を組み合わせて技術を開発した.



図. (a)キャンプファイアの写真. (b)川の急流を描いた油絵.

システムは下図(a)に示すように, 流体ビデオのデータベースを持っている. ユーザが指定するのは3つで, (b)の流体画像とその上の大体の流れ方向を指定するオレンジの矢印, (c)の流体領域である. 後は自動的にシステムが(d)の流体アニメーションを作る.

<http://130.153.152.4/~okabe/FluidAnimeVideoDB/index-j.htm> にアクセス頂くか、「岡部 ビデオデータベース」で検索するとデモ動画が見られるのでご覧頂きたい.

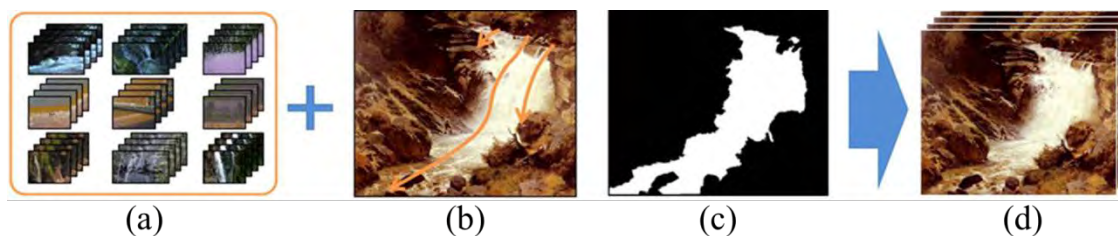


図. システム概要.

この技術の肝は, 局所的な動画検索である. 下図は, 3つのビデオデータを用いて小川の写真をアニメーションさせようとしている. この3つの小川のビデオは写真の小川とは全く異なるものだ. しかし, 局所的に見ると似ている部分もある. 例えば, 黄色, オレンジ, 青の丸印で囲んだ領域では, それぞれビデオと写真の間で水しぶきの調子が似ている. このような類似性にもとづいて, 各丸で囲んだ部分のビデオを写真上へコピーする, というのがアイデアである. 写真上の全ての場所へ, いろんなビデオから部分的なコピーをしつつ, 画像合成技術を使って切れ目が見えないように合成することで最終的なアニメーションを得ることができる.

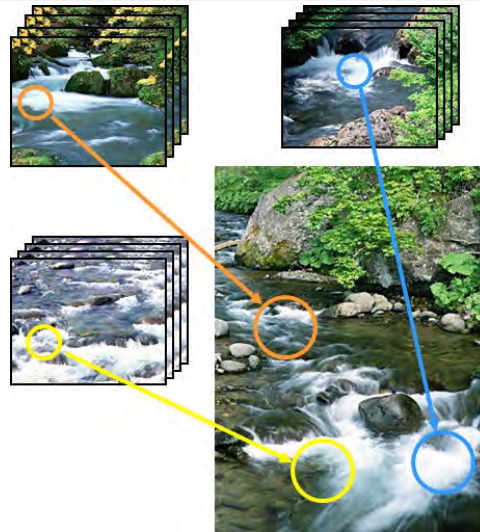


図. ビデオの部分的なコピーと画像合成によって小川のアニメーションを作る.

② 動画鑑賞のためのスケッチインターフェース

「論文 2, 特許 1, 学会発表 2, 9」

今, 下図(a)のように, 車が左から右へ走るシーンを見ているとする. この時「この車が手前へ走ってくる, (b)みたいなシーンがあったなあ, あれを見たいなあ」と思うことがある. 普通だと, 早送り, 巻戻しボタンや再生バーを操作してシーンを探す. それでも見つからない場合は YouTube へ行って検索するかも知れない. しかし, ここではもっと迅速に目的のシーンへ移動する手法を提案する.



図. (a)左から右へ走る車を鑑賞中に, (b)のような手前に走ってくる車のシーンを見たくなった. (c)前後のウイングをつまんでクルッと回すイメージで緑の矢印を描くと, (d)の目的のシーンが検索される.

ユーザの見たい(b)のシーンとはつまり, 今見ている(a)の車を, クルッと回して前を向かせたようなシーンのことである. そこで, このクルッと回す感じを(c)のように緑の矢印でもって, ユーザにスケッチしてもらおう. 「回転して, 前後のウイングが丁度この辺に来ると, 私の見たいシーンになるなあ」という感じである. すると, システムが即座に目的のシーンを検索してくれる. デモ動画は「川手 岡部 尾内」で検索すると川手氏のホームページが出てくるので, そちらでご覧頂きたい.

この技術の肝は, 画像検索と軌跡検索を組み合わせたところにある. 画像検索だけだと, (a)の画像が与えられても (b)の車を直接検索することは難しい. この様子を下図に示す. 画像検索でよく用いられる, 局所的な画像特徴のマッチングを試みた例である. (a)に示す車は左右反転

しているものの、車自体は似たような姿勢をしていてマッチングがたくさん採れるので検索することができる。一方、今ユーザが見たいと思っている(b)は、局所的なマッチングは3つしか採れなかったのが通常は検索結果とならない。

しかし、この(a)と(b)の関係に注目する。実は (b)の車が手前に走ってきて左側へカーブしていくのが(a)、つまり(a)と(b)は一続きのシーンなのだ。そこで我々のシステムは、まず局所マッチングに基づいて(a)を検索し、次に(a)を起点に早送り、巻戻しを自動で行いながら「ユーザがスケッチした緑の矢印みたいな軌跡を描いて前後のウイングが動いたようなシーンはないかしら？」とチェックする。すると、(a)を少し巻戻したところに(b)が見つかるというわけである。

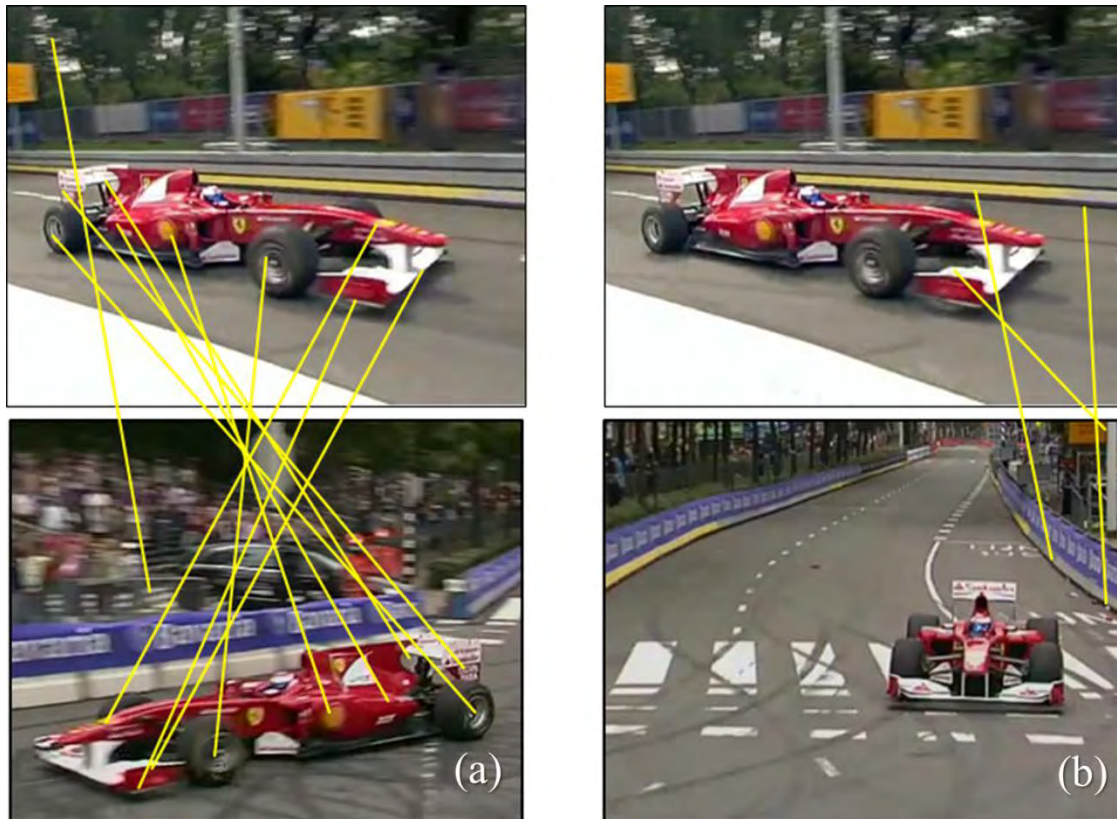


図. 画像検索で用いられる局所的な画像特徴のマッチング.

③ 入力音楽をあたかもバンド演奏しているかのように見える動画の生成技術

「論文 5, 学会発表 3, 5, 8, 12, 受賞 4」

入力音楽に合わせて、その音楽の演奏動画を半自動的に生成し、演奏を耳と目の両方で楽しむことができる。この手法ではピアノ、ギター、ベース、ドラムなどの演奏動画を収集して動画データベースを作っている。下図の Database に示すように、動画のオーディオトラックを予め分析することで、図中の(b)などの音符の発音タイミングを抽出しておく。この発音タイミングは実際に演奏者が演奏しているであろう楽譜上の音符の位置に相当するようなものだ。今、下図の Input Music に示すような入力音楽が与えられたとすると、システムは同様のアルゴリズムで図中の(a)に示すように発音タイミングを抽出する。この発音タイミングを特徴量として動画データベースを検索することで、類似した発音タイミングを持つ動画を検索して下図中の「??」部

分に割り当てる。ここでは(a)と(b)は類似しているのので、(b)の動画を割り当てれば、(a)を演奏しているように見えるに違いない、というわけである。この作業を入力音楽の全てのパートに繰り返すことで、演奏動画を生成する。デモ動画は「yamamoto okabe onai」などと検索すると山本氏のホームページが出てくるので、そちらでご覧頂きたい。

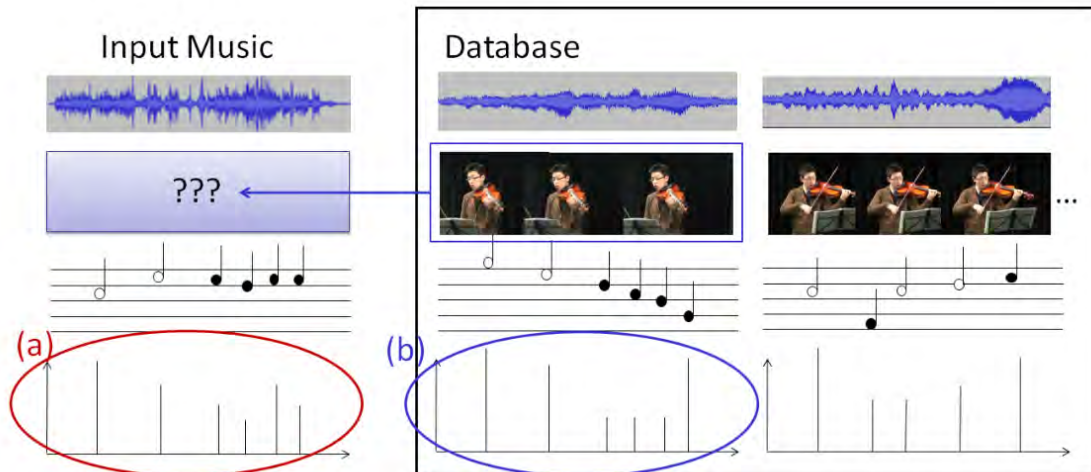
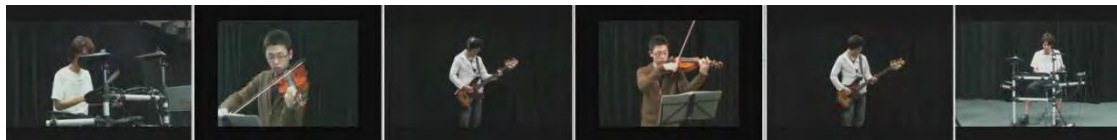


図. 入力音楽とデータベース. 両者において、音符の発音タイミングを自動的に抽出して検索を行う。

下図に合成動画のフレームをいくつか示す。音楽をただ聴くだけだと、それほど注意して聴いていなかった低いベースの音も、映像中のベース奏者の指の動きと一緒に聴くとより鮮明に聴こえてくる。映像（偽モノですが）が音楽の楽しみ方に与える影響について、引き続き分析と実験を行っている。



下図. ここでは、ビートルズの「レットイットビー」を入力音楽としてバンド演奏を合成している。

④ アマチュア演技者の動作を格好良く見せるように動画を編集する手法

「論文 4, 学会発表 4, 7」

プロの演技者はダンサーでもアクションスターでも非常に格好良い。体は柔らかく筋力があり動きにキレがある。ところが、我々がそれに憧れて真似ようとしても上手くいかない。体は硬く筋力はなく動きにキレが無いのである。そこで、コンピュータに手伝ってもらい、せめて動画の中でだけでも格好良くなろう、という願いに基づいた研究である。

下図(a)では、アマチュアが回し蹴りに挑戦している。しかし、(b)のプロのように足は高く上がらないし、動きも遅くて格好悪い。そこで、このプロの動作を抽出してアマチュアの動画に適用することで編集したものが(c)である。これで、足は高く上がっているし、動きも速く格好良くなった。「mizui okabe onai」で検索すると水井氏のホームページがあるので、是非デモ動画をチェッ

くして頂きたい。



図. プロの動作を抽出してアマチュアの動画に適用することでアマチュアを格好良くする。

技術的には、アマチュア動画とプロ動画の両方の動作を分析して抽出(下図(a)と(b))して、前者を後者に合わせて編集する(下図(c))。動作の分析は図のように、人物の動きをスケルトンで追跡することで行う。この際、全自動でコンピュータが動きを完璧に追跡することは難しいので、間違いが起こった際はユーザがマウスでスケルトンを編集できるようなツールを作っている。スケルトンが抽出できたら、(a)のスケルトンが(b)のスケルトンに重なるように(a)の画像を変形する。すると(c)ができ、足が高く上がったように見える(背景もそれに合わせて歪んでいるが、人物のみを自動で切り出して最終結果を合成する)。

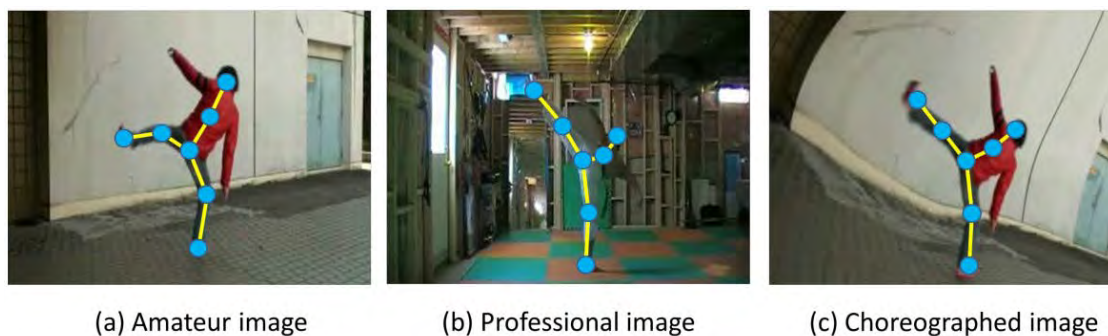


図. スケルトンを用いた動作の追跡と変形への適用。

⑤ . 勾配画像処理に基づく動画からの流体抽出

「論文 3, 学会発表 1, 6」

動画から爆発や煙等の半透明部分が広範囲に現れるような流体を抽出する手法を提案した。例えば下図では、ビル前で起こった爆発の映像から、ビルを消し、爆発のみを抽出できれば、後でこの爆発を動画制作に利用できる。しかし、流体の背後にはこのビルのような複雑な背景がうっすらと見えるのだが、これと流体を分離するのは時間と手間の掛かる作業となる。

我々は全自動でこの問題を解決する手法を提案した。ここで動画を撮影したカメラは静止しているものとし、我々はまず背景画像を推定し、次に流体の前景とアルファマツを推定する。本手法では、動きの少ないピクセルで背景が見えることを用いて背景を推定する。次に推定された背景との差分としてアルファマツを推定する。この差分によるアルファマツには背景模様が出現してしまう。そこでこの模様を消すために勾配画像処理を用いる。勾配画像中で背景模様に関する勾配を弱め、その上で積分することで精製されたアルファマツを得る。前景画像も同様の勾配画像処理を用いて得られる。

本手法を用いて爆発、煙、炎等の半透明領域が広範囲に渡る動画において、その流体の抽出

ができることを示す。また、下図に見えるように、既存の Trimap を用いた手法と比較しても、提案手法がクオリティの高い流体抽出ができることを示した。

<http://makotookabe.com/FluidExtraction/index.htm> にて動画及びソフトウェアを公開中。

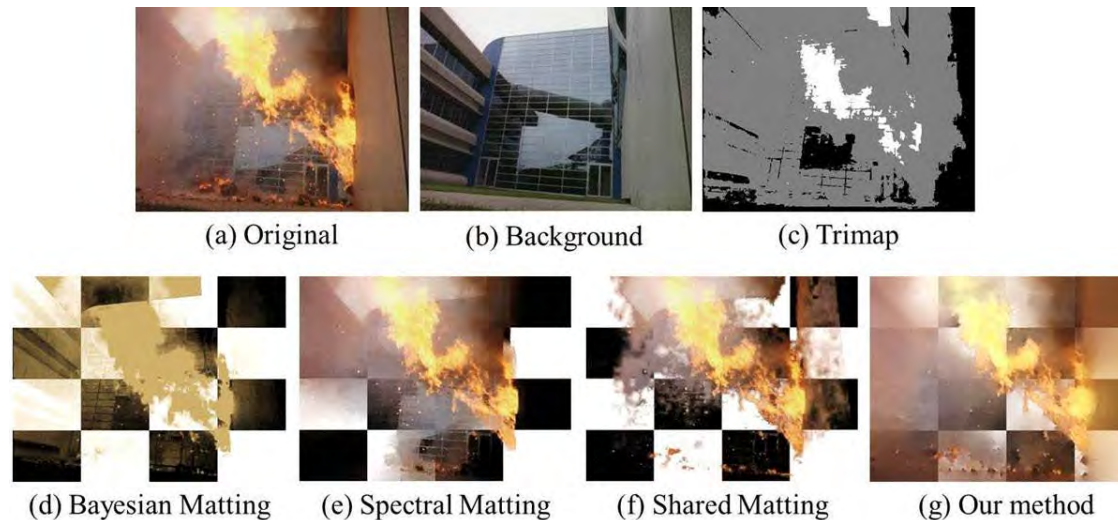
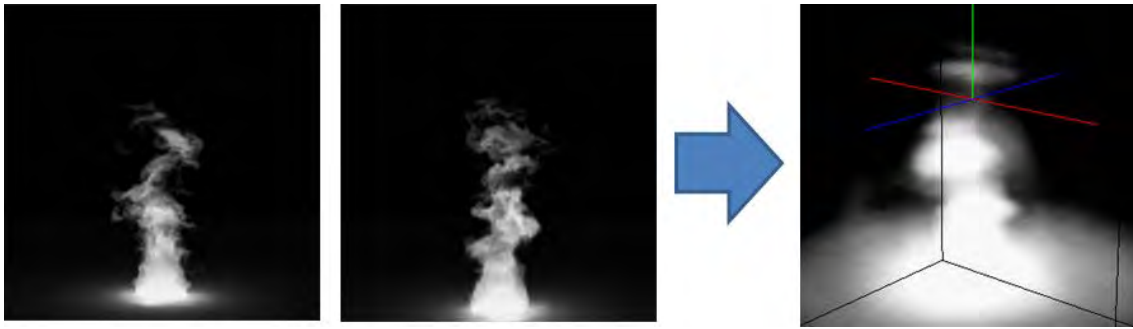


図. 既存手法との比較。オリジナルフレーム(a)と背景画像(b)との差分からトライマップ(c)を作成して既存手法への入力とした。Bayesian Matting で作成した合成画像(d)。Spectral Matting で作成した合成画像(e)。Shared Matting で作成した合成画像(f)。本手法で作成した合成画像(g)。

3. 今後の展開

本年度までの3年半、2次元動画データベースの解析と合成の研究に従事したと共に、世界中で映像データベースに基づく映像合成が盛んに研究されている事実も踏まえ、動画データベースの研究を続けることが未来の映像表現に貢献できると確信しつつ、同時に2次元動画の限界も悟るに至った。申請者の考える2次元動画の限界は大きく2つ。1つは上述した「後で視点を変えられない」という3次元的表现への限界である。もう1つは、2次元動画を解析・学習させても、得られるコンピュータの映像理解力に限界があるという事である。我々人間も両眼立体視や焦点調節を通じ、周囲の環境を3次元的に学習して、その結果、現在の視覚能力を得ているのである。近年、我々の研究環境もコンピュータに人間と同様、3次元環境を学習させるのに都合の良い世界になってきた。最も顕著な事は3次元撮像技術の発展である。近年ハリウッドで製作される映画の多くが3次元映画で、全ての映像が2台のカメラでステレオ撮影された3次元映像であることが普通である。Microsoft Kinect に代表される3次元スキャナも安価に普及している。この情勢を見れば、今後、数十年の間に3次元動画コンテンツが普及していくのは明らかであり、現在の2次元画像・動画の世界は大きく変わっていくと思われる。

今後の大きな予定の1つは、上で述べた2次元流体動画の応用研究を、3次元流体動画に拡張していくことである。3次元流体を測定する技術はいくつか提案されているが、特別な機材を必要とするなど、扱いやすいものがない。したがって、安価な測定技術自体を作るという研究が可能である。一方、さまざまな流体を測定 & 収集して3次元動画データベースを作り、検索、加工、動画合成のためのオーサリングツールを作る等の研究を進めることも可能である。現在、2つのカメラで撮影した動画像から煙の3次元化を行うところまで行っている(下図)。



4. 自己評価

「①ビデオデータベースを用いて流体画像をアニメーションさせる手法」は「研究のねらい」に書いた内容を、ほぼそのまま実現しようとした研究である。ユーザは絵と簡単なアノテーション、という入力でコンピュータにアイデアを与える。後は流体動画データをあらかじめ分析したコンピュータが自動的にアニメーションを制作してくれる。ユーザが結果を修正したければ、アノテーションの修正及びパラメータ調整をして、コンピュータに再計算を依頼できる。研究結果は雑誌、及び、2つの学会発表、2つの表彰という形でまとめる事が出来た。プレゼンをする時(フレンドリーな会場だと)「おー」と歓声上がる事もあったので、良い研究が出来たのではないかと考えている。今後は手法として誰しも簡単に使えるツールという形で公開できるように、手法を簡略化したいと考えている。

「②動画鑑賞のためのスケッチインタフェース」は新しい映像閲覧のためのユーザインタフェースの研究である。こちら、ユーザはスケッチというシンプルなアノテーションを与えるだけで、予め動画を解析したコンピュータがユーザの見た動画の即座に検索してくる事ができる。物体の回転等を考慮した検索が可能なのだが、3次元形状を分析する必要が無く、2次元的分析のみで済むという辺りがウリである。この研究も国際学会での発表と国内特許をとる形で成果にまとめた。現在より大規模なデータで実験を行ったものの論文を執筆中である。

「③入力音楽をあたかもバンド演奏しているかのように見える動画の生成技術」「④アマチュア演技者の動作を格好良く見せるように動画を編集する手法」も今までにない動画の編集技術の提案を行い、国際会議での発表を行う事が出来た。特に④の研究では学会会場の発表等で、聴衆から笑い声を聞く事ができ、本来の研究の目的を達成できたと考えている。「⑤勾配画像処理に基づく動画からの流体抽出」は上記のような研究を進めるために、リッチな動画データベースを構築するのを助ける手法であり、本研究のテーマの補助的な研究であるが、今まで解けなかった問題を解く事ができた。こちらは学会発表を行い、ソフトウェアを公開している。

5. 研究総括の見解

動画 DB を参照しながら静止画を動画として合成する研究である。

大量の映像データを統計的に分析して知識を抽出し、ユーザが直感的な映像製作を可能にする環境を作ることがテーマである。このテーマに沿って、「ビデオデータベースを用いて流体画像をアニメーションさせる手法」、「動画鑑賞のためのスケッチインタフェース」、「入力音楽をあた

かもバンド演奏しているかのように見える動画の生成技術」、「アマチュア演技者の動作を格好良く見せるように動画を編集する手法」、「勾配画像処理に基づく動画からの流体抽出」等の基本的な技術の研究開発を行った。それぞれ、「目で見て分かる」成果をあげている。

今後、このさきがけ研究を通じて得られた、2次元動画の解析・学習の限界とコンピュータに人間と同様に3次元環境を学習させるべきとの知見に基づき、2次元流体動画の応用研究を3次元流体動画に拡張していくと同時に、これらの成果を統合して、どのような方向に向かうのかを明確にしながら研究を進めてほしい。

6. 主な研究成果リスト

(1) 論文(原著論文)発表

1. Makoto Okabe, Ken Anjyo, Rikio Onai, "Creating Fluid Animation from a Single Image using Video Database", Computer Graphics Forum, vol. 30, no. 7, pp. 1973-1982.
2. Makoto Okabe, Yuta Kawate, Ken Anjyo, Rikio Onai, "Video Retrieval based on User-Specified Appearance and Application to Animation Synthesis", In Proc. of MMM 2013.
3. Makoto Okabe, Ken Anjyo, Rikio Onai, "Extracting Fluid from a Video for Efficient Post-Production", In Proc. of DigiPro 2012.
4. Kenta Mizui, Makoto Okabe, Rikio Onai, "Choreographing Amateur Performers Based on Motion Transfer between Videos", In Proc. of MMM 2013.
5. Tomohiro Yamamoto, Makoto Okabe, Rikio Onai, "Synthesis of a Video of Performers Appearing to Play User-specified Band Music", ACM SIGGRAPH 2012 poster.

(2) 特許出願

研究期間累積件数: 3 件

1.

発明者: 平野 廣美, 三條 正裕, 川手 裕太, 岡部 誠, 尾内 理紀夫
発明の名称: 動画検索装置、動画検索方法、記録媒体、ならびに、プログラム
出願人: 楽天株式会社
出願日: 2012/1/20

2.

発明者: 平野 廣美, 岡部 誠
発明の名称: 画像処理装置、画像処理方法、及びプログラム
出願人: 楽天株式会社
出願日: 2012/8/24

3

発明者: 尾内 理紀夫, 岡部 誠, 栗原 竜矢, 三條 正裕, 平野 廣美
発明の名称: 動画処理装置、動画処理方法、ならびに、情報記録媒体
出願人: 楽天株式会社
出願日: 2012/9/7



(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

●学会発表

1. Makoto Okabe, Ken Anjyo, Rikio Onai, "Creating Fluid Animation from a Single Image using Video Database", Pacific Graphics 2011.
2. Yuta Kawate, Makoto Okabe, Rikio Onai, "Video Retrieval based on User-Specified Deformation", ACM SIGGRAPH 2012 poster.
3. Tomohiro Yamamoto, Makoto Okabe, Rikio Onai, "Synthesis of a Video of Performers Appearing to Play User-specified Band Music", ACM SIGGRAPH 2012 poster.
4. Kenta Mizui, Makoto Okabe, Rikio Onai, "Choreographing Amateur Performers using Video Examples", To appear at SIGGRAPH Asia 2011 poster, pp. 8:1--8:2.
5. Tomohiro Yamamoto, Makoto Okabe, Rikio Onai, "Synthesis of a Video of a Performer Appearing to Play User-specified Music", ACM SIGGRAPH 2011 poster.
6. 岡部 誠, 安生 健一, 尾内 理紀夫, "勾配画像処理に基づく動画中の流体部分抽出", ビジュアルコンピューティング '12, 2012年6月.
7. 水井 健太, 岡部 誠, 尾内 理紀夫, "動画間の動作転送に基づくアマチュア動画編集システム", ビジュアルコンピューティング '12, 2012年6月.
8. 山本 知大, 岡部 誠, 尾内 理紀夫, "バンド演奏における疑似演奏動画の合成手法", ビジュアルコンピューティング '12, 2012年6月.
9. 川手 裕太, 岡部 誠, 尾内 理紀夫, 平野廣美, 三條正裕, "静止画内オブジェクトへの指示による動画検索手法の提案", DEIM 2012, 2012年3月.
10. 鈴木 喜也, 岡部 誠, 尾内 理紀夫, "無音動画に対する効果音貼付けシステムの試作", DEIM 2012, 2012年3月.
11. 岡部 誠, 安生 健一, 尾内 理紀夫, "ビデオデータベースを用いた流体画像に基づくアニメーション生成", ビジュアルコンピューティング '11, 2011年6月.
12. 山本 知大, 岡部 誠, 尾内 理紀夫, "楽曲を演奏しているように見える動画の生成手法", ビジュアルコンピューティング '11, 2011年6月.

●受賞

1. 岡部 誠, 安生 健一, 尾内 理紀夫, "ビデオデータベースを用いた流体画像に基づくアニメーション生成", 2011年6月.(VC賞)
2. 岡部 誠, 安生 健一, 尾内 理紀夫, "ビデオデータベースを用いた流体画像に基づくアニメーション生成", 2012年6月.(グラフィクスとCAD研究会優秀研究発表賞)
3. 鈴木 喜也, 岡部 誠, 尾内 理紀夫, "無音動画に対する効果音貼付けシステムの試作", DEIM 2012, 2012年3月.(学生奨励賞)
4. Tomohiro Yamamoto, Makoto Okabe, Rikio Onai, "Synthesis of a Video of a Performer Appearing to Play User-specified Music", ACM SIGGRAPH 2011 poster. (Student Research Competition Semi-Finalist)
5. 湊 匡平, 岡部 誠, 尾内 理紀夫, "俳句に適合する画像を合成するシステムの開発", 第3回楽天研究開発シンポジウム, 2010年12月.(優秀ポスター賞)

●著作物

1. Computer Graphics Gems JP 2012 -コンピュータグラフィックス技術の最前線(共著)