様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ５ 月 ２６ 日現在

機関番号：１７１０２
研究種目：挑戦的萌芽研究
研究期間：2013～2014
課題番号：２５５４００７０

研究課題名（和文）工学的想像 ― 大量事例を用いた動画像推定

研究課題名（英文）Engineering Imagination -- Inferring past frames using massive video instances

研究代表者
内田　誠一（Seiichi, Uchida）

九州大学・システム情報科学研究科（研究院・教授

研究者番号：７０３１５１２５

交付決定額（研究期間全体）：（直接経費）　　2,900,000円

研究成果の概要（和文）：本研究では、我々人間が持つ想像能力の一側面を工学的に実現した。ここで言う「想像」は次の例によって説明される。「静止画像Xには、壊れた花瓶の脇にボールが転がっている様子が写っている。このとき、その画像Xに至る過程の動画像を生成せよ」。この想像能力の工学的実現のために、本研究では「大量の事例（動画像）を適切に参照しながら、推定結果としての動画像を最適合成する機構」を実装した。この実装においては，最適事例の選択法と事例を用いた推定画像の生成法について，それぞれ立案，実装，検証を行った。ペットボトル落下の様子を写した1000枚の動画を準備・利用し，手法の妥当性を検証した.

研究成果の概要（英文）：The aim of this research is to mimic the imagination skill of human-being by computer. Human can imagine the process that an object falls down and bounds at the floor just by watching a final image where the object lies on the floor. One assumption for this imagination skill is that we refer to similar instances that we have already seen before. Based on the assumption, we prepare a large instance data set and realize a method to estimate the past frame images by referring instances appropriate for the individual frame images.

研究分野：画像情報学

キーワード：パターン認識　映像解析　映像合成　時系列処理　大規模データ

１．研究開始当初の背景

百万オーダー以上の大規模画像データを手元の計算機システムで利用可能になってきた。それを受け、これまで夢とされてき様々な研究が萌芽している。「膨大な画像集合があれば、ある画像について、それに似た別の画像もある。」―この単純な原理が言えるようになることで、複雑化一辺倒だった画像処理、特に画像認識について、むしろ原点回帰とでも言える研究動向が見られる。申請者らも、1 クラスあたりのラベル付きデータが世界最大である画像データベース（80 万手書き数字データベース。1 クラスあたり 10 万弱の目視ラベル付きデータ）を利用し、画像空間における真の分布の姿を解明し、さらには認識率向上のための一般的方策を研究してきた。（挑戦的萌芽研究 H23-24 採択課題。）

２．研究の目的

本研究では、大規模な動画像データを用いることで、我々人間が持つ想像能力の一側面を工学的に実現する。本研究で言う「想像」は、次の例によって説明される。「静止画像 X には、壊れた花瓶の脇にボールが転がっている様子が写っている。このとき、その画像 X に至る過程の動画像を生成せよ」。この例題に対し、我々ならば、「ボールが花瓶に向かって飛んできて、花瓶に当たり、そして花瓶が割れ、ボールはその近くに転がり、いずれ止まる」という様子を写した動画像を生成、すなわち「想像」するだろう。

この想像能力の工学的実現には、次の２点が手掛かりになる。第一に、我々は膨大な量の視覚的経験を事例として持っており、それらを参照して画像Ｘに至る動画像を推定している点である。すなわち、想像のためには大量の事例（動画像）を扱う必要がある。上の例で言えば、我々は様々な物体が壊れる様子を見たことがあるから、こうした想像ができるのである。第二に、我々は想像のために、画像Ｘに合わせて動画像を新たに合成している点である。これは、大量の事例を持っていても、それは画像 X と完全に同じ状況ではないために、別途新たな画像を作りだす必要があるためである。上の例で言えば、画像 X 内の花瓶の壊れる様子を、過去に見た別の花瓶に関する動画像を用いて合成する必要がある。以上により本研究では、「大量の事例（動画像）を適切に参照しながら、推定結果としての動画像を最適合成する機構」の実現を目的とする。

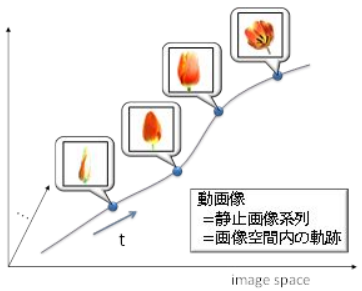
上述の通り、過去の推定（もしくはその逆方向としての未来の予測）を動画像として行うことは、全く新しい研究課題である。もちろん、カルマンフィルタを用いた動き予測など比較的単純かつ短時間の推定・予測については、既に様々な研究がなされている。しかしそれらはあくまで数理的なモデルに依拠した推定方式であり、モデルに当てはまらないようなより複雑かつ長期間の推論には全く適さない。例えば上図にあるような、「現在の花の状態に至る過去の様子」などは、従来のモデルベースでは複雑すぎて扱うことは事実上不可能である。これに対し当課題では、大量の事例を用い、それらを参照することで、現状態に至る様子を動画像として合成するという試みである。事例が大量であることは極めて重要である。少数では参考するに足る適切な事例が見つからないためである。まさに大量データを扱える昨今だからこそ、挑戦すべき課題と言える。

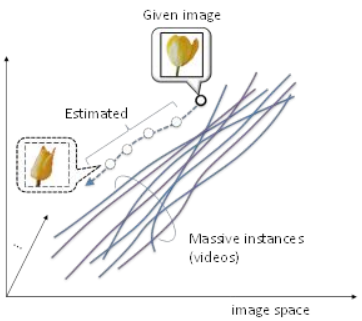
ここで強調するべきは、この動画像推定問題は決して単純ではない、という点である。すなわち、事例を十分に準備して最近傍事例をそのまま推定結果しても、解決にはならない。この理由は、入力される静止画像と完全一致する事例はまず存在しないためである。従って、事例動画像に写された時間変化のパターンを抽出し、さらに事例と入力静止画像間の空間的変化を抽出し、それらを同時に入力静止画像に適用することで、新たな動画像を合成する必要がある。換言すれば、画像に潜む空間的変化と時間的変化を分析し、それらを用いて画像を最適合成する必要がある。このように、大量事例を用いた動画像推定という新しい問題の解決原理は、「事例の選定」ならびに「選択された事例と入力を組合せて、推定画像を合成」するという２つの新しい着想に基づく。

３．研究の方法

(1) 本研究の基本的な方法論：

右図に示すように、静止画像が１点として表現される画像空間において、動画像は軌跡として表現される。従って、大量の動画像事例は、それらは大量の軌跡を成す。さらに、それら事例が似たような対象を写した動画像であれば、それら軌跡は互いに類似しており、画像空間の中において「一定の方向に流れる束」のように存在することになる。





このように大量の事例が与えられた状況において、１枚の静止画像が与えられたとする。すなわち、画像空間において１点を定める。この時、我々が解くべき推定問題は、「この

点すなわち入力静止画を起点として、事例を参照しながら、新たな軌跡を合成する問題」と考えることができる。軌跡を伸ばす方向によって、過去の動画像の推定、もしくは未来の動画像の予測を行うことになる。

既に述べたように、この問題を解くためには、「どのように参照する事例を選ぶか」そして「参照した事例をどのように用いて、軌跡を合成するか」の２点を検討する必要がある。前者については様々な方式が考えられる。例えば、選択の基準や、参照する事例が単一か複数か、等がある。後者については、前述のように、事例をそのまま利用できない点が根底にある。入力静止画像が事例上に含まれることは事実上有り得ないため、「事例に沿った」軌跡の合成法を開発する必要がある。

(2)大量事例からの最適事例の選択：
推定時に参照する事例を如何に選択するかは、推定結果を左右する重要な点である。基本的方針は類似事例の参照である。この方針に基づき以下の検討を行う。

①事例選択時のメトリック：単純には画像間のユークリッド距離や正規化相関だが、事例と入力間のズレに弱い。注目物体ではない領域（背景など）の影響もある。このため、ズレに強い Chamfer distance や、2 画像間の弾性マッチング距離(例えば SIFTflow)を用いる。

②単一事例 or 複数事例：入力静止画像との類似性によって推定に用いる事例を唯一つに確定する方法が最も単純な選択法であろう。一方、推定が進むにつれて（すなわち過去の状況を合成するにつれ）類似事例も変化し得ることを鑑み、複数の事例を動的に選択する方法も考え得る。

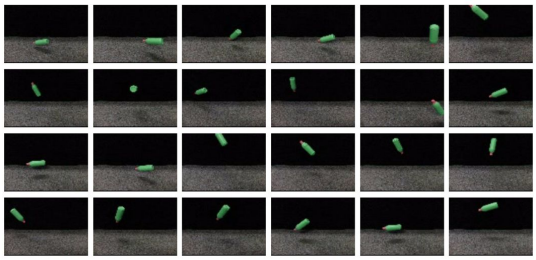
③最適化プロセス：次項の合成と事例選択を独立・逐次的に推定するか（すなわち局所的最適な推定）、あるいはあらゆる選択と合成の組合せを考えて大局的に最適化するか。

(3)選択した事例からの推定結果の合成：
選択した事例を用いて、推定結果すなわちフレーム画像を合成する必要がある。再三述べているように、単純な事例参照、すなわち事例中のフレームをそのまま用いるのは不適切である。入力静止画像に矛盾しないような画像を、「時間的変化」「空間的変化」を分析しながら、事例を用いて新たに合成する必要がある。これについても様々な検討課題がある。

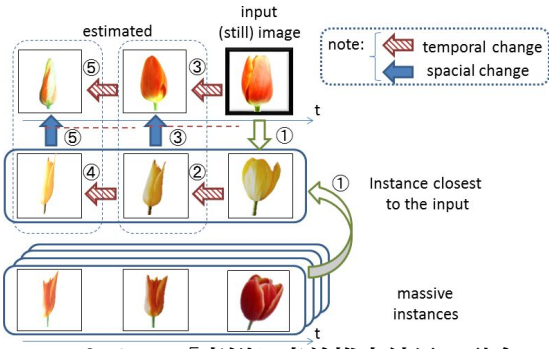
①事例と直前推定結果の融合：時刻 t の推定結果画像の合成には、上記により選択された事例および t-1 の推定結果の両方を考慮する必要がある。これら２つをどのようにブレンドするべきかを検討する。

②直前推定結果との連続性：時刻 t と t-1 の推定結果画像は、比較的類似している必要がある。（＝推定結果が成す軌跡は連続的である必要がある。）合成時に、この類似性をどのように担保するべきかを検討する。

４．研究成果
(1)H25 年度
①データ収集：
研究開始段階においては，まず，推定のための事例となる大量の動画像を準備した．どのような動画像を収集するかは極めて重要である．これに対し，本研究では，推定問題を極力単純化するため、単一物体が落下し，床に当たって跳ね返って止まるまでの状況を写した動画像 1000 枚を撮影した。これら事例を用いて、落下中のナップショット１枚から、その前の動画像の推定実験を行った。事例数の変化による推定結果の影響など、多様な評価実験を行った．下図は，その中からランダムに選んだ 24 枚の動画像における，フレーム画像例である．



②最適事例選択：
H25 年度は，推定結果の合成法に力点を置くべき，最近傍となった単一事例を用いた．
③選択した事例からの推定結果の合成：
入力静止画像に矛盾しないような画像を、事例の「時間的変化」「空間的変化」を分析しながら推定、すなわち合成する。以下はその手順を図化したものである。

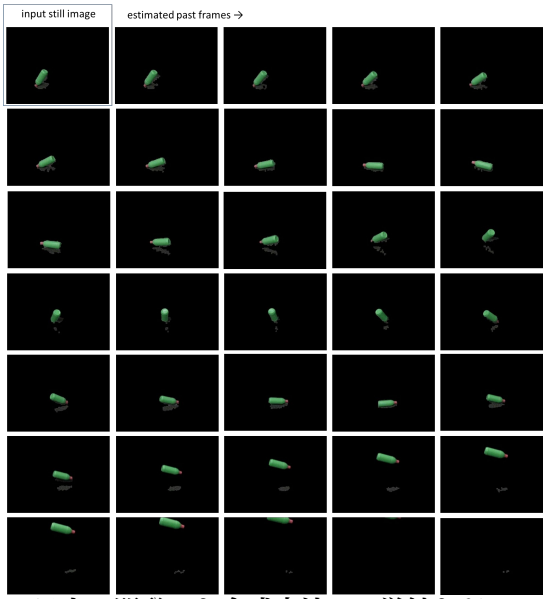


このために，「事例と直前推定結果の融合」「直前推定結果との連続性の担保」の実現法を開発した。この方法として、異なる５種類の方法を実装し、それらの相互比較を行いながら、極力スムーズな推定結果を得るための方法論を模索した。その結果、落下対象の輪郭を用いる方法や、素直にフレーム間差分を用いる方法が安定的であり、画像間非線形マッチングを用いる方法が不安定であることを示した。

(2)H26 年度
①最適事例選択：
H25 年度で用いた事例が単一であったのに対し，H26 年度では，複数事例を常に使い続け

るようなアプローチを検討した．具体的な選択手法については，推定結果合成と密接に関連するので，次項②において述べる．
②選択した事例からの推定結果の合成：
複数事例を利用しながら推定を行う方法論として，具体的には，大量事例参照→複数事例の選択→フレーム合成→大量事例参照→...を繰り返すアプローチを採用した．その結果，H25 年度の単一事例に比べ，大量事例を活用するほうがより安定的に合成結果が得られることも実験的に検証された．その際，どのように事例を参照するかが，方法論検討の段階で重要となってくる．複数の参照法を検討した中，合成されたフレーム画像について，k 近傍を求め，その距離（類似度）を重みとして次の時刻のフレーム画像を合成するという，原点回帰にも近い，比較的単純な方法が最も安定することが分かった．

下図はその実行例である．左上が入力された静止画であり，そこから右そして下の方向に順次推定されたフレーム画像系列を示している．本例では極めて安定して，過去の状況が推定できていることがわかる．



以上で開発した合成方法は，単純なだけに拡張性も高く，例えば様々な距離尺度の利用も可能である．

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計１件）
①Volkmar Frinken, Ryosuke Kakisako and Seiichi Uchida, A Novel HMM Decoding Algorithm Permitting Long-Term Dependencies and its Application to Handwritten Word Recognition, Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition, 査読有, pp. 128 - 133, 2014. DOI: 10.1109/ICFHR.2014.29

〔学会発表〕（計 5 件）
岩切裕太郎，フォン ヤオカイ，内田誠一，大規模事例に基づく動画像推定のための対象の動き表現，電子情報通信学会パターン認識･メディア理解研究会, 2013 年 12 月 13 日, 三重大学

内田誠一，柿迫良輔，深澤大我，フリンケン フォルクマー，フォン ヤオカイ，弾性マッチング二題 ～ 最適化法を変えて広がる応用 ～，電子情報通信学会パターン認識・メディア理解研究会，2013 年 12 月 13 日，三重大学

深澤大我，藤崎顕彰，フォン ヤオカイ，内田誠一，K-近傍弾性マッチングを用いたオンライン文字認識，電子情報通信学会パターン認識・メディア理解研究会，2014 年 2 月 13 日，福岡大学

上村 将之，岩切 裕太郎，フォンヤオカイ，内田 誠一，大規模事例に基づく動画像推定 -時間的および空間的変化の抽出法の検討-，画像の認識・理解シンポジウム，2015 年 7 月 30 日, 岡山コンベンションセンター，岡山市

深澤大我，フォン ヤオカイ，内田誠一，K-近傍弾性マッチングに関する諸検討，電子情報通信学会パターン認識・メディア理解研究会，2015 年 2 月 20 日，東北大学

〔その他〕
ホームページ等
human.ait.kyushu-u.ac.jp/~uchida

６．研究組織
(1)研究代表者
内田　誠一（UCHIDA SEIICHI）
九州大学・大学院・システム情報科学研究院・教授
研究者番号：70315125