

**科学研究費助成事業 研究成果報告書**

平成 29 年 6 月 16 日現在

機関番号：32689

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26330209

研究課題名(和文) 動画像処理による非剛体の挙動解析の研究

研究課題名(英文) Research on Analyzing Non-rigid Objects' Behaviors by Temporal Image Processing

研究代表者

大谷 淳(OHYA, Jun)

早稲田大学・理工学術院・教授

研究者番号：90329152

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：近年、動画像処理の研究が様々な応用への適用を目指して活発に行われている。本研究では、(1)動的動画像中の動物体の抽出・追跡、および(2)動画像処理に基づく手術の状況解析、に焦点をあてた。(1)については、剛体が移動するシーンを動いているRGBDカメラにより動画像で撮像し、特徴点のクラスタリングに基づき各剛体を抽出・追跡する方法を提案した。(2)については、手術者の手の動作を認識する方法として、Bag-of-Wordsに基づく特徴と学習に基づく方法、早期認識する方法として、サブ動作を導入し、SVMや確率計算に基づく方法を提案した。実験により提案手法の有効性を検証した。

研究成果の概要(英文)：In recent years, towards the application to various applications, temporal image processing has been actively researched. This research focuses on the following two topics: (1) Extracting and tracking moving objects in dynamic, temporal images, and (2) Analyzing surgical situations based on temporal image processing. For (1), the following method is proposed: a moving RGBD camera acquires temporal images that contain scenes in which rigid objects moves, and then each rigid object is extracted and tracked based on clustering the feature points. For (2), the following two methods are proposed: for recognizing surgeons' hand actions, a method based on Bag-of-words and learning; for early recognition, a method, for which "sub-action" is introduced, based on SVM and probability computations. Experimental results show the validity of the above-mentioned proposed methods.

研究分野：画像情報処理

キーワード：動画像処理 動的動画像処理 認識 早期認識 剛体 手術 確率

## 1. 研究開始当初の背景

近年、動画処理の研究がさかんに行われるようになってきており、ロボット、医療、介護、セキュリティ、交通、市場調査、各種作業の自動化、ヒューマンインタフェースなど様々な分野に応用されることに対する期待と重要性が高まっている。動画処理の具体的な検討課題は多様である。そこで、本研究では、(1) 移動するカメラにより獲得される動画から、撮像されたシーン中を移動する物体を抽出・追跡する技術、および(2) 動的な物体(人物を含む)の不定形な(一つの認識カテゴリーに分類されるものの、変動が大きい)挙動を認識・解析する技術、の2つのテーマに焦点を絞り、研究を行う。なお、(2)では、外科的手術のシーンを動画として撮像し、医者の手による手術具の操作等を自動的に認識し、自律的に医者の支援を行うことが可能な看護師ロボットの実現に資する動画処理技術を研究する。以下に、(1)と(2)それぞれの課題について、研究開始当初の背景を述べる。

### (1) 動的動画中の動物体の抽出・追跡

自動車の車載カメラにより獲得される動画を利用して、歩行者の飛び出しを事前に検出する研究が最近見られるようになってきている。また、移動ロボットの研究の近年の進展を考慮に入れると、今後は移動ロボットと人間の共存を前提とした環境やサービスが実現される可能性が高く、ロボットに搭載したカメラからの動画に対する処理(ロボットと人間との衝突回避等)による自律的な知的制御技術が必要である。一方、視覚障害者が屋内外を歩行する場合、身体に装着されたカメラにより獲得される動画を解析することにより、歩行可能な方向の提示や障害物回避行動の支援の実現が期待される。また、子供等の社会的弱者に対する誘拐や傷害事件を防ぐためには、身体に装着したカメラからの動画を解析し、周囲の人物の不審な行動を事前に検知することが有効と考えられる。

動画から移動物体を識別する研究は過去に非常に多く行われており、特に静的なカメラや静的なステレオカメラを対象とした研究は数多く見られるものの、移動するカメラにより獲得された動画には適用できない。これに対して、移動するカメラにより獲得される動画を対象とした研究としては、M. Agrawal らの手法があるが、背景からの移動物体の分離のみを対象としており、物体毎の領域抽出は行えない。物体ごとの移動物体領域の抽出を永橋らの手法は行ったが、処理のための重要な情報を手動で与える必要があった。近年、Choiらは、動的カメラからの動画を解析し、道路の歩行者やオフィスの人物の追跡を行う方法を開発したが、追跡処理が人物に特化され、他の物体には適用できなかった。

これらの問題を解決するため、研究代表者らは TMR (Temporal Modified RANSAC (RANDOM Sample Consensus))に基づく方法を提案し、検討を進めている。本手法は、3次元空間中で運動している複数の剛体を、移動する RGB-D カメラ(マイクロソフト社製 Kinect<sup>®</sup>等のように、カラー画像と距離情報を動画として獲得可能。)により撮像することにより得られる動画から検出される特徴点(TMRにより、各剛体と静止背景に分類されている)と Super-voxel(類似した色をもつ、隣接する画素を連結することにより得られる領域を時間軸方向に連結した物:以後 SV)とを利用して、グラフを構築する。グラフのノードは SV、リンクは隣接する SV の境界で、ノードのラベル(各剛体と背景の識別子)と特徴点の整合性等により定義されるエネルギーを最小化することにより、動画における各フレームにおいて各剛体および背景に対応する領域を得る。本研究では、研究代表者らのこの方法を発展させる。

### (2) 動画処理に基づく手術の状況解析

近年の医療現場における看護師の不足は深刻化しており、手術を行う際の障害になりかねない状況である。即ち、手術時には、実際に手術を行う医者に看護師が次に必要な手術具を迅速に手渡す等の支援を行うからこそ、効率的かつ効果的な手術が実現できる。そこで、RSN (Robotic Scrub Nurse)と呼ばれるロボット看護師の研究がいくつか見られるようになってきている。例えば、GestoNurse と呼ばれるシステムは、医者の手形状と手術具を予め対応付けておき、手術中に医者が提示する手形状を GestoNurse が画像処理により自動的に認識し、その手術具をマニピュレータで医者に手渡す。しかし、このようなシステムに対して、現場の医者は否定的であり、有効でないと考えている。有能な看護師の多くは、術野を一瞥しただけで、その時点での手術の進行状況が理解でき、次に医者に手渡すべき手術具が判断できる。医者が真に望むのは、前述のような有能な看護師を代行可能なロボットの実現である。

手術の種類により術野に違いがあるが、開腹手術の場合、術野に存在するのは大別して、医者の手、手術具、患者の身体内部の臓器等である。従って、医者の手動きなどの動作は、手術の状況を理解するためには、重要な情報源と言える。そのため、医者の手や手術具に各種のセンサを装着し、検出される手や手術具の動きから手術の状況を推定する研究が見られる。しかし、このようなセンサ装着は医者の動作の妨げになりうる。また、手術具の臓器への作用は扱われておらず、状況理解の精度に課題を残している。

研究代表者らは、センサ装着の不要な動画処理により、手術の状況を認識する方法を研究している。具体的には、傷の縫合手術を対象に、同じ手術具を使うため、手術具では

区別できない縫合（傷を縫い合わせる）と結紮（手術用糸を結ぶ）の2つの状況を、医者の手の動きの動画像処理に基づき識別する方法を提案し、検討している。提案手法では、色彩情報を用いて抽出された手の領域において多数の特徴点を検出し、SVM(Support Vector Machine)により、前述の2つの状況を識別する。本研究計画の着想は、この研究の遂行により導かれたと言える。

## 2. 研究の目的

課題(1)と(2)の目的を述べる。

### (1) 動的動画像中の動物体の抽出・追跡

前述の研究代表者らの方法の問題点は、取り扱える物体が剛体に限定されていること、および隣接するフレーム間の動きが小さい場合の検出精度の低下である。人物像のような関節構造をもつ非剛体を扱おうとすると、フレーム間の小さな動きを扱える必要があるため、まずこの問題の解決を目指す。これに基づき、現状の方法を非剛体物体へ拡張する方法を明らかにする。

### (2) 動画像処理に基づく手術の状況解析

前述の研究代表者らの方法は、縫合あるいは結紮のみの動画像を処理対象とする場合は高精度での認識が可能だが、種々の状況が連続する場合は扱えない。また、より難易度の高い手術を対象とするための改良法が見出せていない。そこで、より難易度の高い手術として開腹手術を対象に、医者が手にもつ手術具の、患者の臓器への作用の進行度(例えば、切開すべき距離のうち、どの程度切開したのか)を含めて、手術の状況を理解可能な、新たな方法の確立を目指す。

## 3. 研究の方法

課題(1)と(2)の方法を述べる。なお、(1)については、剛体を対象とした手法の改良法の検討を行ったので、その方法を述べる。(2)については、手術時の手術者の手の動作の認識に加えて、動作が終了する前に認識を行う早期認識の検討を行ったので、その方法についても述べる。

### (1) 動的動画像中の動物体の抽出・追跡

提案手法では、まず初めに時系列画像から、スーパーピクセルを生成する。スーパーピクセルの生成において、画素の色情報に加えて、距離情報も用いるように改良を加える。次に、各スーパーピクセルに含まれる特徴点のフローに対して、RANSACを適用して誤対応のフローやスーパーピクセル内部の別の物体のフローを除去して、同時に回転・並進行列の推定を行う。

スーパーピクセルによっては、特徴点の数が少なかったり、RANSACの結果、移動情

報の推定に失敗したりするケースがあるため、グラフカットのラベルは、移動情報の推定に成功したスーパーピクセル毎に定義する。そして、各ラベルに対応した回転・並進行列を各フローの起点に対して適用し、終点との誤差を計算してスーパーピクセル毎に集計した結果をデータ項として用いる。各ノードを接続するエッジは、スーパーピクセルの中点を元にドロネー分割を行い、ピクセル間のエッジの本数に基づき定義する。作成したグラフをカットして、クラスタの統合を行う。統合した結果に基づき、回転・並進行列を再推定して、グラフ構築及びグラフカットをエネルギーが小さくなる限り繰り返して、エネルギーが最小になるラベルを決定し、その結果を最終的なクラスタリング結果として得る。

提案手法では、従来法の「初期クラスタの生成処理における問題」を、スーパーピクセルを初期クラスタの単位として利用することで、解決する。スーパーピクセルは必ずしも物体とピクセルの対応は1対1ではなく、過分割な傾向があり、一つのスーパーピクセルが異なる物体に跨ることはない。そのため、スーパーピクセルを初期クラスタとして用いた場合は、DeLongらの従来法においてランダムに選択した特徴点3つを初期クラスタとして利用する場合と異なり、初期の回転・並進行列の推定時に使用する特徴点が、異なる二つの物体にまたがることのない。また仮に跨ることがあっても、姿勢推定時にRANSACを適用して、異なる物体に属する特徴点は除去される。

さらにこのRANSAC処理を通じて、精度の悪いフローも除去されるため、DeLongらの方法の「フロー生成処理の誤差の悪影響」の問題も同時に解決している。また、このRANSACの閾値の設定は、TMRのようにRANSACを繰り返し適用して複数のモデルを抽出する場合に比較すると、移動情報を推定するモデルは一つで良いため、閾値の決定は容易である。

### (2) 動画像処理に基づく手術の状況解析

手術時の手術者の手の動作の認識法は以下のように学習に基づいている。学習処理ではまず、学習データである各動作のビデオを構成するフレームをsectionに分割する。Sectionごとにランダムに選択された点についての3D-SIFT記述子を計算し、これに対してクラスタリングを行うことによりword listを作成する。各学習データの各sectionにおける各wordの頻度を計算してヒストグラムを作成し、全sectionのヒストグラムを連結したSwoBS (spliced bag-of-words of sections)特徴を求める。各学習データのSwoBS特徴を用いてニューラルネットの学習を行う。認識処理では、認識対象の手術の連続ビデオに対してsliding windowを移動させ、sliding windowの各場所において

sliding window 内のフレームから SWS 特徴を計算し、前述のニューラルネットに入力して、動作の認識結果を得る。連続ビデオ中で認識結果の妥当性を評価して、最終的な認識結果を得る。

手術動作の早期認識法は以下の通りである。動作全体の情報が使える認識よりは技術的難易度が高い早期認識を実現するため、本方法では、動作は複数の subaction (サブ動作) から構成されていることを前提とする。サブ動作は各動作に固有のものであり、早期認識の対象である。提案手法は動作開始点検出と早期認識から構成されている。

動作開始点検出処理では、手術動作以外の動作の混入に対処するため、one-class Support Vector Machine を利用して開始候補点を検出する。早期認識処理を各候補点から開始し、各時刻においてテストビデオがサブ動作モデルに対応する確率を計算し、確率が閾値を超えた場合、サブ動作が認識されたと判断する。全ての候補点から得られたサブ動作の認識結果を解析することにより、最終的な動作の認識結果を得る。

#### 4. 研究成果

##### (1) 動的動画中の動物体の抽出・追跡

剛体が移動するシーンを、移動する Kinect により撮像して得られる RGBD 動画を対象に、提案手法の有効性を検討した。まず、提案手法を用いて、原画像の特徴点に対してクラスタリングを行った。提案手法を用いた場合、特徴点は物体毎に正しく分類されていることがわかる。

提案手法の有効性を示すために、時系列画像の複数のシーケンスに対して、提案手法を適用した。このシーケンスは、二つの移動物体と、静止している一つ、又は複数の静止物体を移動する Kinect を用いて撮像した。局所的にラベリングが誤っている特徴点も存在するが、殆どの特徴点は正しくクラスタリングされることがわかった。

##### (2) 動画処理に基づく手術の状況解析

傷の縫合手術を対象に連続ビデオを用いて認識実験を行った。傷の縫合手術を構成する縫合と結紮の手術動作を認識対象とした実験の結果、90%以上の認識率を達成した。手術時動作のような、固定パターンの無い手の動作の認識率としては良好な結果と言える。

早期認識については、傷の縫合手術を対象に連続ビデオを用いた認識実験を行い、動作の進捗度(動作全体の時間に対する、動作開始時点からの経過時間の割合)が70%の時点において、90%以上の認識率を達成した。サブ動作という新たな概念の導入が有効であると考えられる。

#### 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4件)

Ye Li, Jun Ohya, Toshio Chiba, Xu Rong, and Hiromasa Yamashita, "Subaction Based Early Recognition of Surgeons' Hand Actions from Continuous Surgery Videos", IIEEJ Transactions on Image Electronics and Visual Computing, Vol. 4, No. 2, pp. 124-135, (2016.12), 査読有。

Naotomo Tatematsu, Jun Ohya and Larry S. Davis, "Supervoxel-Based Detection and Segmentation of Moving Objects from Dynamic RGBD Images", IIEEJ Transactions on Image Electronics and Visual Computing, Vol. 4, No. 1, pp. 32-42, (2016.6), 査読有。

Ye Li, Jun Ohya, et al., "Early Recognition of Surgeons' Hand Actions from Continuous Surgery Videos", Proceedings of Interactive Medical Image Computing (IMIC), affiliated with MICCAI2015, pp.78-85, (2015.10), 査読有。

Ye LI, Jun OHYA, Toshio CHIBA, Rong XU, Hiromasa YAMASHITA, "Recognizing Hand Actions during Suture Surgeries from Video Sequences Based on Sliding Windows, SBoWS Feature and Neural Network", IIEEJ Transactions on Image Electronics and Visual Computing, Vol. 3, No. 1, pp. 34-43, (2015.6), 査読有。

[学会発表](計5件)

坪田実和, 李 イエ, 大谷 淳, "縫合手術を構成する6つの段階の動画中の手術者の手の動きに基づき予測する方法", 信学技報, vol. 116, no. 464, IE2016-112, pp. 257-262, 2017年2月, 北海道大学(北海道・札幌市)

三浦朋也, 李イエ, 大谷 淳, "縫合手術を構成する5つの段階の動画からの識別法の検討" 画像電子学会研究会, 15-03-016, pp.135-141, (2016.3), 九州工業大学(福岡県・飯塚市)

Ye Li, Jun Ohya, Toshio Chiba, Rong Xu and Hiromasa Yamashita, "Study of Early Recognition and Progress Level Calculation of Surgeon's Action during Suture Surgery", 2015年度画像電子学会第43回年次大会, R2-3, pp.1-2, (2015.6), 姫路市市民会館(兵庫県・姫路市)。

Ye Li, Jun Ohya, Toshio Chiba, Rong Xu, and Hiromasa Yamashita, "Fundamental Study of Early Recognition of Surgeon's Action during Suture Surgery from Video Sequences", 2015年電子情報通信学会総

合大会, 情報・システム講演論文集 2, p. 152, (2015.3), 立命館大学びわこ・くさつキャンパス(滋賀県・草津市).

Ye Li, Jun Ohya, Toshio Chiba, Rong Xu, and Hiromasa Yamashita, "Recognizing Hand Actions during Suture Surgeries by using sh-SIFT algorithm", 第 32 回 日本ロボット学会 学術講演会, 3H1-06, pp.1-4, (2014.9), 九州産業大学(福岡県・福岡市).

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

## 6. 研究組織

### (1) 研究代表者

大谷 淳 (JUN OHYA)

早稲田大学・理工学術院・創造理工学部・総合機械工学科・教授

研究者番号: 90329152

### (2) 研究分担者

無し

### (3) 連携研究者

無し

### (4) 研究協力者

無し