

映像インデキシングのための物体追跡と変化の認識 Object Tracking and Object Change Detection for Video Indexing

津吹 陽介[†]
Yosuke Tsubuku

中村 裕一[‡]
Yuichi Nakamura

大田 友一[†]
Yuichi Ohta

1. はじめに

操作や作業を教示するための教材やマニュアルとして、映像は強力な道具となる。一般企業や教育機関でもこのような映像コンテンツの制作に関心が集まっているが、映像コンテンツの制作には多大な時間と労力を必要とし、その自動化が望まれている。また、このような映像コンテンツを有効に利用するためにはインデキシングが不可欠である。例えば、機械の組み付けの説明では、機械を組み上げるのに必要な部品や重要な道具の説明が与えられるが、学習者が一人で作業をする場合、このようなポイントを状況に応じて簡単に参照できれば効率の良い作業ができる。そのため、本研究では、組み立て作業映像を取得する際に自動的にインデックスを付与することを目的としている。

机上作業を対象とした映像教材や作業マニュアルでは、部品・道具などの物体やそれらの使い方についての説明が重要な情報となるため、物体像やそれに対する説明を認識することが重要となる。さらに、机上作業を対象とする場合には、環境の変化に柔軟に対応できる物体認識が必要とされる。本研究では、そのために、ボクセルカービング、把持物体認識、体積変化の推定等を行なう手法を提案する。

2. 机上作業の認識

組み立て作業や実験を行なう場合、手や物体が机上空間を動き回り、追跡対象となる物体が隠蔽されたり、回転したり、さらに変形、結合、分離など様々な変化が生じる。そのため、物体の位置、テクスチャに関する事前知識を十分に使えない場合が多い。

このような問題に対し、我々のグループでは、可視光カメラ、赤外線カメラ、ステレオカメラの3つの画像センサを用いたシステムを構築し、簡単にロバストな物体追跡を実現した[1]。この手法では、物体に関する事前知識をほとんど使わずに、10 cm角程度の物体を比較的精度良く追跡することができる。しかし、1方向からの観測であるため、隠れに弱いこと、見え方の変化と物体の変化との区別が難しいこと等が課題として残っていた。

本研究では、この問題に対処するため、作業空間中の手や物体を検出するだけでなく、作業空間を占める物体の体積やその変化を捕える。我々はその最も基本的な構成として、天井と側面の壁の2視点にカメラを設置したシステムを構築した。図1にシステムの例を示す。各視点で使用するカメラは可視光カメラ、赤外線カメラの2種類のカメラ対である。赤外線カメラを用いて温かい領域を検出し、可視光カメラから得られる色情報、動き情報と統合することにより、正確に手や把持物体を抽出する。これについて、3.節で説明する。さらに、複数方

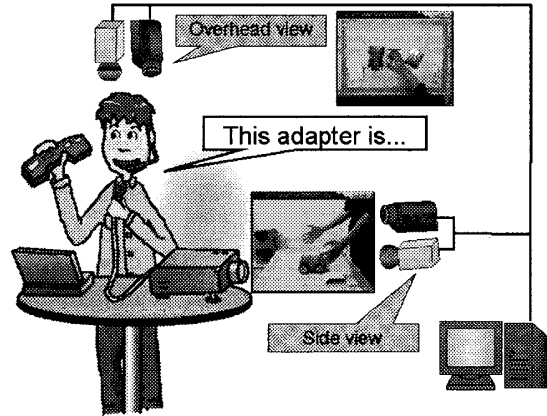


図1: システムの概要

向から撮影された画像を入力としてボクセルカービングを行ない、物体の机上空間に対する位置と体積の計測を行なう。さらに、その結果を用いて物体の変化を推定する。これについては、4.節で説明する。

3. 把持物体検出・追跡

本手法の基本的なアイデアは、ある特定の作業空間中で手と共に移動する領域を検出し、手領域と分離することで把持物体を認識することである。

手領域: 可視光カメラから得られた肌色領域と、赤外線カメラから得られた肌温領域の以下のような論理積を手領域とする。

$$\text{手領域} = \text{肌色領域} \wedge \text{肌温領域}$$

動領域 / 把持物体領域: フレーム間差分とテンプレートマッチングの併用によって動領域を抽出し、それが手領域と近接してれば以下のように把持物体領域と定める。

$$\text{把持物体領域} = \text{動領域} \wedge \neg \text{手領域}$$

4. ボクセルカービングを用いた物体変化の認識

ボクセルカービングを用いて物体の体積とその変化を検出する。我々の実験では、ボクセルを $50 \times 50 \times 50$ (=125000) 個空間上に設置し、1個の大きさは $20 \text{ mm} \times 20 \text{ mm} \times 20 \text{ mm}$ とした。処理の実行速度は、我々の計算機上で約 15 frame/sec である (dual Xeon 3.06GHz)。

このようにして得られた体積を把持物体の体積と定める。体積は物体認識に用いるだけでなく、その変化を認識することにも用いる。しかし、ノイズに弱く、ノイ

[†]筑波大学 大学院 システム情報工学研究科

[‡]京都大学 学術情報メディアセンター

表 1: 検出率

| | 造花 | 車の模型 | 小さな球 |
|-------|-------------|-------------|-------------|
| 正解 | 950 (95.0%) | 920 (92.0%) | 524 (52.4%) |
| 準正解 | 50 (5.0%) | 54 (5.4%) | 297 (29.7%) |
| 誤検出 | 0 (0.0%) | 26 (2.6%) | 179 (17.9%) |
| total | 1000(100%) | 1000(100%) | 1000(100%) |

ズと物体の変化を区別することは難しい。そこでノイズを軽減するため、複数のフレームで観測した値を用いる。現在のところ、連続する5フレームの中で2番目に小さい値を採用している。

また、組み立て作業では、作業中におこりうる物体の変化として、結合、分離、隠蔽、回転等が挙げられる。結合が起こった場合には、体積の大きな変化は起こらず、物体の検出個数が2個から1個へと減少する。隠蔽が起こった場合は、物体の個数、体積ともに減少する。このような変化の定義と時系列に沿った体積の変化の流れを観測し、体積と状況を考慮した状態変化の判定を下す。

5. 実験

5.1 把持物体の検出率

把持物体の追跡精度を調べるために、3種類の物体について検出率を求めた。表1に各物体に対する追跡結果を示す。ここで、“準正解”とは把持物体周辺を検出した場合(検出された領域と物体の領域に重なりがある場合)、“誤検出”とは把持物体以外の領域が検出されたか、検出が起こらなかった場合を表す。準正解を含めて、全ての物体について80%以上の精度で正しく検出できた。

5.2 体積変化の検出

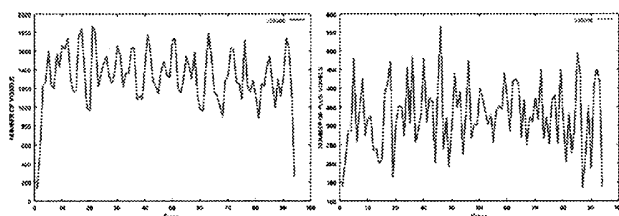


図 2: 背景差分法による体積検出結果 (本文参照)

まず、入力画像が体積検出に及ぼす影響を検証するために、体積が既知の立方体(8000cm³)を机上空間で動かし、背景差分法とフレーム間差分法を比較、体積の観測を行なった。図2は背景差分法によって得られた画像を入力として算出したボクセル数(図左)、ボクセルの変化分(図右)を表したグラフである。物体を大きく動かしたため、結果として大きな誤差が観測された。なお、フレーム間差分法では、ノイズの影響や差分が十分に検出されないこともあり、これらよりも劣る結果が得られた。

この結果から、以後の計測には背景差分法を採用した。図3は上記の立方体をゆっくり動かしながら観測した結

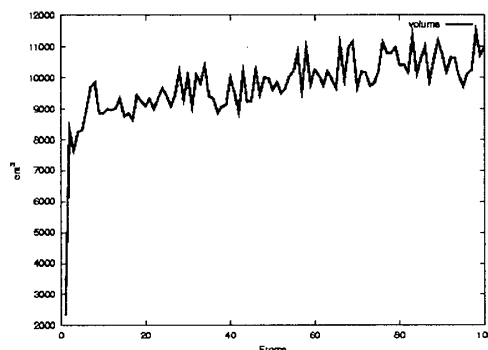


図 3: 体積検出結果

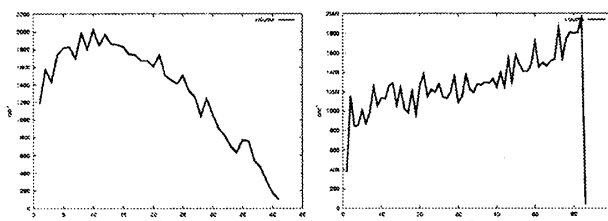


図 4: 風船を用いた体積検出結果 (左: しばませたとき, 右: 割ったとき)

果、図4は、膨らました風船の変化の様子を観測した結果である。グラフより、体積の変化によってそれぞれの現象が捉えられていることが分かる。

6. まとめ

複数の視点からの2種類の画像センサ対を相互補完的に用い、机上作業プレゼンテーションにおける把持物体を検出、追跡する手法を提案した。また、体積の変化を検出し、物体の状態変化を認識する枠組みについて述べた。今後の課題としては精度や安定性の向上があげられる。

参考文献

- [1] M.Itoh, M.Ozeki, Y.Nakamura, and Y.Ohta "Simple and Robust Tracking of Hands and Objects for Video-based Multimedia Production", IEEE Conf. on Multisensor Fusion and Integration for Intelligent Systems, pp.252-257, 2003.
- [2] S.M.Seitz and C.R.Dyer, "Photorealistic Scene Reconstruction by Voxel Coloring", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp.1067-1073, 1998.
- [3] A. Laurentini, "The Visual Hull Concept for Silhouette-Based Image Understanding", Transactions on Pattern Analysis and Machine Intelligence, 16(2), pp.150-162, 1994.