

H-008

## 動的カメラ画像からの動物体追跡のためのロボットプラットフォームの検討 Study of Constructing a Robot Platform for Tracking Moving Objects from Mobile Video Sequence

羅 丹† 謝 英第† 大谷 淳†  
Ra Tan Yingdi Xie Jun Ohya

### 1. はじめに

コンピュータビジョンの分野では、動画画像から移動物体を抽出し、追跡・位置同定を実現することはセンサーを用いたシステムにおいて最も重要な技術の一つである。

先行研究では、視覚装置が固定となるケースは一般的である。また、物体の移動情報獲得と物体追跡は別の問題として取り上げられている。自立的な空間認識機能の実現には、追跡対象物体を自動認識し、それに基づいて追跡を行うことに関する処理が含まれている必要がある。

本研究では、移動領域分割・移動追跡・位置同定ということを目指し、能動的にコントロールをすることが可能とする視覚移動ロボットのプラットフォームについて検討した。

### 2. システム構成

研究プロジェクトは人間行動認識を最終の目的としている。人物の「移動分析」、「位置同定」、「行動理解」という一連の技術は、各種防犯システムや介護ロボット、あるいは店舗での顧客行動分析などのマーケティングへの応用といった幅広い分野で、注目されている。本研究はコンピュータビジョンの視点から人間と共存し、協調作業ができるロボットシステムを開発することを目指している。

#### 2.1 ハードウェア構成

実験用のモデルは図1の(a)のステレオカメラ BumbleBee と図1の(b)のトラッキング型ロボットを使用した。ステレオカメラをロボットの先方に搭載した。また、操作インターフェースとして、タッチパネル付の ThinkPad Tablet X60 を用いた。

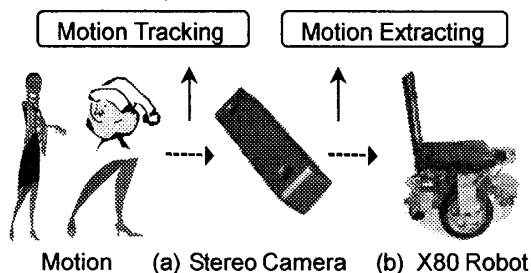


図1 ハードウェアのアーキテクチャ

#### 2.2 ソフトウェア構成

ロボティクスの研究の究極の目標は、動的環境に適応するために、センサー情報から自身の内部構造を組織化する自律的な知能ロボットを実現することである。真の意味でのロボットの知能を実現させるためには、ロボット自身の感覚や行動による環境との密な総合作用を通して、種々の行動を学習し、それらを状況に応じて統合することが必要である。最終の目的を実現するために、まず動

的ステレオカメラで撮影したステレオ画像を用いて移動領域分割を行う。そこから得たシルエットに対して特徴を抽出する。また、人間行動データベースを参照し行動に対する認識結果を計算する<sup>[3]</sup>。さらに、移動領域分割のステップで抽出した背景情報からシーンの理解を行い、ロボットを能動的に行動させるために行動判別結果からロボットの制御情報を計算する。<sup>[1]</sup>

本研究では上記の初期検証として、移動領域分割、移動データの収集、移動認識するための移動追跡のプロセスに沿い、ハードウェアに実装し、実験を行った。検証計画の全体構成は図2のようになる。

- (1) 移動領域の分割・抽出
- (2) 移動追跡
- (3) 能動視覚による行動の獲得

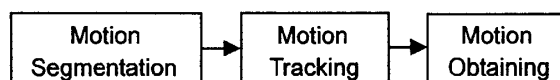


図2 計算処理のアーキテクチャ

### 3. プラットフォームの構築

本研究は 2.2 で述べた目的を達成するために、基礎的な行動を実現でき、簡易に操作できる実験モデルのロボットプラットフォームを構築した。さらに、日常の環境に動的プラットフォームで獲得したデータを用いて移動領域分割・追跡等の処理モデルについて述べる。

#### 3.1 ハードウェア構成

今回構築したハードウェアのモデルは高度な視覚センサーによる認識・追跡機能、超音波センサーによる障害検知機能、また音響モジュールを持ち、さらに、遠隔及び PC による操作が可能となる。X60 ロボットは高機能、軽量化、省エネかつ長時間 (3700mAh 3 時間) 稼働可能、また延長機能の追加を付けやすい特徴となる。

##### (1) 視覚センサー

フレーム毎に認識精度を高めるために、移動対象の奥行きを参照とした 3 次元空間情報が必要である。高速処理と高解像度で高品質な 3 次元情報取得アプリケーションとして、Bumblebee 両眼デジタルカメラが、人物追跡、障害物検出など、動的な認識に適していると考えられる。最大のフレームレート 30FPS を使う。カメラはアーム上に付着したので、上下方向も回転でき、追跡ものの高さにより視覚方向を決めることができる。図3の赤い枠に示す。

##### (2) 移動プラットフォームの配置。

- 機械部 (図3 A) : 12DV のサーボモータ 2 個と後方のキャスター 1 個で駆動する 2 輪方式で、最大走行速度は 1m/s である。重量は 3.5kg で、可付加重量は 10kg となる。
- センサー (図3 B) : 視覚センサー、音響、超音波センサー。

† 早稲田大学大学院国際情報通信研究科, GITS

- アーキテクチャ (図3 C) : Low-level 信号指令はボード上で管理する。High-level 信号指令はリモートからPCを介して処理を行う。
- 制御コントローラ : Multimedia Controller(PMB5010), Sensing-Motion Controller(PMS5005), VC++言語で開発した“WiRobot SDK ActiveX Module”という APIs でセンサー情報をアクセスし、コマンド指令とシステム設定を行うことができる。図4に示す。

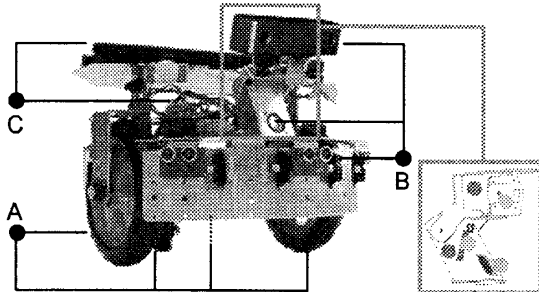


図3 ハードウェアの構成図

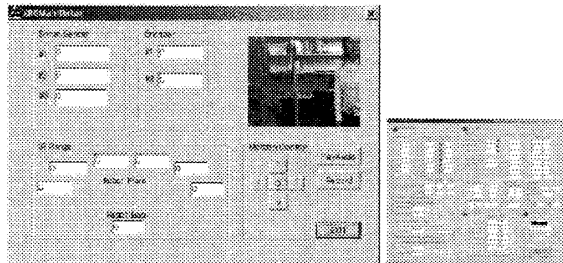


図4 コントロールインターフェースの一部

### 3.2 ソフトウェア構成

3.1 で述べたハードウェアの構成に実装する計算モデルについて説明する。

#### (1) ステレオカメラによるデータの取得

本研究では、Digiclops Stereo Vision SDK Demo II を使用する。データは 30FPS のフレームレートで、曲面キャリブレーションを行わず通常の照明環境で撮影した。

#### (2) 移動領域の自動抽出

入力画像は 2 フレームから左右両眼画像の計 4 枚を用い、計算モデルは同研究室の博士の参考文献の[1]と[2]に基づき、検証を行う予定である。

#### (3) 移動追跡

Particle Filter と Mean Shift の統合手法<sup>[4]</sup>で追跡を行う。Mean Shift 法で求められた観測ベクトルは観測モデルを用い、トラッキングを続ける。

- 探索用画像 : ステップ(2)で抽出した画像を用いる。
- 初期設定 : 求めた画像を追跡対象画像として格納し、探索用画像の中心座標を計算する。
- 色確率分布 : 色ヒストグラムを作成し、H(色相)を 0-255 の階調値に変換する。x, y 座標と階調値を軸に 3 次元座標に色確率分布を作成する。
- 類似した色分布の探索 : 探索用画像の中心座標より 0 次モーメントを求め、画像  $p(x,y)$  の 1 次モーメント計算する。探索画像の平均位置を計算し、上昇勾配のピークが収束した位置を中心位置とする。

上記のトラッキングプロセスで実験を行った。その結果を 4 章で示す。

## 4. 実験結果と考察

3.2 章で説明した計算モデルに基づき、移動追跡の実験を行った。

移動領域を選択し、追跡サンプルとして、5 秒間隔にて 2 枚のフレームのトラッキング結果を示している。追跡移動領域のモデル画像には、目標が小さい時また色情報が不十分の場合、追跡結果が落ちることがわかった。図 5 足の例に示す。目標が大きいつい追跡結果になった。図 5 の人と人形の例に示す。動領域の自動抽出手法を検討する際、目標の大きさ色情報を考慮する必要があるとわかった。実験では 3.1 章で述べたロボットプラットフォームを用い、実空間で目標物体を追跡し、能動的な活動の可能性を検討した。

今後、移動ロボットに基づく移動追跡に対する安定かつ高速の手法を目指し、3 次元情報を利用した能動視覚によるロボット行動の獲得も検討したいと思っている。

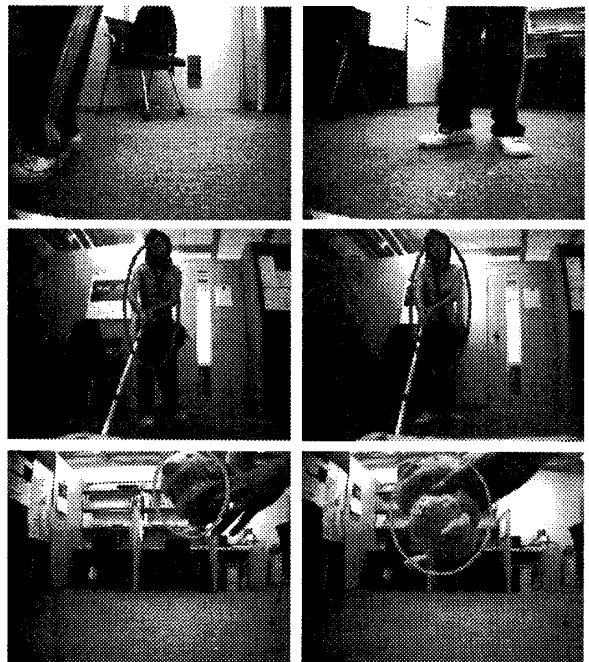


図5 トラッキング処理結果

### 参考文献

- [1] Yingdi Xie, Jun Ohya, "Motion Segmentation Based on Active Stereo Camera", Technical Committee on Image Engineering (IE), Vol.105, No.689, pp1-6, Mar, 2006
- [2] Yingdi Xie, Jun Ohya, "Efficient multiple independent motion segmentation from an active platform by utilizing modified RANSAC", International Conference on Graphics and Visualization in Engineering (GVE 2007), 562-057
- [3] Acep Irawan, Xie Yingdi, Jun Ohya, "Implementation of a Video Sequence Based System for Recognizing an Unknown Person's Gesture by a Tensor Decomposition Based Approach", Proceedings of the 2007 IEICE General Conference, D, P239, (2007.3)
- [4] Maggio E., Cavallaro, A., "Hybrid Particle Filter and Mean Shift tracker with adaptive transition model", Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05).