

パートナーロボットのための特徴点に基づく画像認識と センサーフュージョン

Image Recognition and Sensor Fusion based on Feature Information for Partner Robots

賈 成達¹ 久保田 直行¹

ChengTa Chia¹, Kubota Naoyuki¹

¹ 首都大学東京大学院 システムデザイン研究科

¹ Department of System Design, Tokyo Metropolitan University

Abstract: This paper discusses a system for real-time object recognition in real-world environment by extracting distinctive invariant features from image. We explain sensor fusion based on measured data from CCD camera and 3-D range sensor to solve the problem of scale variant of image. Finally, we show our experimental results of the proposed method.

1 はじめに

近年、人間と相互作用を行えるパートナーロボットに関する研究開発は盛んに行われている。また、人間とコミュニケーションのための視覚をロボットに持たせるためには、パターン認識は不可欠である。本研究では、まず、人間とのコミュニケーションの観点からロボットの視覚を考える。人間と円滑なコミュニケーションを行うためには、人間の日常生活で存在する物体を全般的に認識できる視覚が必要される。ここでは、特定種類の認識対象に特化された認識手法ではなく、あらゆる物に対し特徴を抽出し、認識を行う手法について議論する。

本稿では、物体から特徴点を抽出する手法を用いて、パターン認識を行う。ここで特徴点というのは、物体の画像上にエッジや濃淡変化が激しいところ。画像上に特徴点の位置から抽出した画像情報を用いて、画像認識を行う。

パートナーロボットが人間と自然なコミュニケーションを行うためには、人間が実世界における行動にあわせる必要があると考えられる。そのためには、物体認識が動的環境下での物体のスケール変化における対処と、リアルタイムの認識スピードを必要とする。本稿では、センサーフュージョンの手法を提案し、カメラ画像と三次元距離センサーの融合によって、従来手法における動的環境下での物体のスケール変化時の認識が困難である問題と、多重解像度分析によって処理時間が膨大になる問題を解決する。その結果、動的環境下でスケール変化する物体のリ

アルタイムに近いロバストな認識が可能になった。

2 センサーフュージョン画像認識システム

開発したシステムでは、一台の CCD カメラと一台の三次元距離センサーが使われている。二つのセンサーが計測した空間をマッピングすることによって、センサーフュージョンを行い、CCD カメラからの画像情報と三次元距離センサーからの三次元距離情報を融合する。次に、CCD カメラの物体画像情報から特徴点を測定し、距離情報を用いて、特徴点位置から抽出する画像情報のスケールを決めて、抽出する。最後に、事前の学習における、同じ手法によって抽出した物体特徴情報とマッチングを行うことより、物体認識を行う。

2.1 センサーフュージョン

ここで使用した CCD カメラは、Logitech 社製のウェブカメラ Logitech QCAM (図 1, 表 1)。使用した三次元距離センサーは MESA 社製の赤外線 TOF 方式三次元距離測定ミニチュアカメラ SR-3000[1] (図 2, 表 2)。測定した距離値を、輝度値を用いてイメージとして再現すると、図 3 のようになる。

この二つセンサーを隣接して設置する時、お互いの 80% の視野をカバーすることができる。次に、両センサーの画像の中に位置が一致する四点を選び、その四点を用いたアフィ変形マッチングを行う (図

4). これによって、CCD カメラの二次元画像に、距離のデータを加えて、三次元の画像データになる。

2.2 特徴点の選定

本稿での特徴点は、CCD カメラ画像において、物体のスケールと視点が変化しても不変の特徴を持つ、再選定可能なポイントである。これらの特徴点は、CCD カメラの画像の輝度値 $f(x,y)$ の微分に基づいて、微分幾何学的アプローチで選定する。グラデーションのような一様な輝度こう配の領域から特徴点を選定するのは望ましくないため、このアプローチでは2階微分までが必要となる。更に、輝度曲面の形状を記述する量であれば、座標回転（ x 軸をどちらの向きに取るか）に関する不変量でなければならない。

不変量の理論によれば、曲面 $f(x,y)$ のこう配 ∇f

とヘッセ行列(Hessian) H を

$$\nabla f = \begin{pmatrix} f_x \\ f_y \end{pmatrix}, \quad H = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} \quad (1)$$

と定義するとき、可能な不変量はラプラス作用素

$$\Delta f = f_{xx} + f_{yy} \quad (2)$$

$$\|\nabla f\|^2 = f_x^2 + f_y^2 \quad (3)$$

$$\det H = f_{xx}f_{yy} - f_{xy}^2 \quad (4)$$

の組み合わせが知られている。この問題に対し、現在最もよく用いられている Harris 作用素[2]は次のように定義される。

$$Harris = \det \hat{C} - k(tr\hat{C})^2 \quad (5)$$

$$\hat{C} = \begin{pmatrix} G_\sigma(f_x^2) & G_\sigma(f_x f_y) \\ G_\sigma(f_x f_y) & G_\sigma(f_y^2) \end{pmatrix} \quad (6)$$

ただし k は定数であり、 G_σ は標準偏差 σ のガウス関数との畳み込む。この手法では1階微分のみしか計算しない。よって処理負荷は大幅に低下させることができる。



図 1: CCD カメラ

表 1: CCD カメラの主な仕様

解像度	320 x 240 pixels
フレームレート	最大 30 fps
色深度	24-bit



図 2: 三次元距離センサー

表 2: 三次元距離センサーの主な仕様

解像度	176 x 144 pixels
視角	47.5 x 39.6 degrees
光学レンズ	f/1.4
フレームレート	Real-time



図 3: 三次元距離センサーの測定値 (左)
同時に CCD カメラの画像 (右)

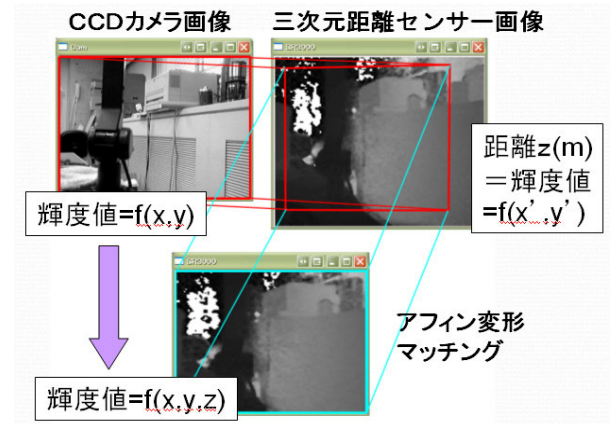


図 4: アフィン変形マッチング

行列 $\hat{C}$ の固有値を λ_1, λ_2 とすると, Harris 作用素

の不変性から $\hat{C}$ が対角化されるため xy 座標系を回転しても値は同じである. したがって Harris 作用素は次のように書ける.

$$Harris = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \quad (7)$$

ただし k は定数である. この Harris 作用素の評価に, Kanade-Lucas-Tomasi 法(KLT)[3]を使う. 特徴点の追跡しやすさの指標として $\min(\lambda_1, \lambda_2)$ を採用している. 図 5 は CCD カメラ画像から特徴点選定の結果.

2.3 特徴点の抽出とマッチング

本稿では, 特徴点がある CCD カメラ画像中のピクセルが対応した三次元距離センサーの距離データに基づいて, 特徴点を中心に抽出する画像情報のスケールを決める. これは CCD カメラから物体までの距離が遠いほど, 画像中におけるスケールが同比率に小さくなるからである. すべての距離の下で特徴点から同じ情報を抽出するために, 実世界における抽出のスケールを SS (Sampling Scale-cm) とし, CCD カメラ画像中に該当する抽出のスケールを DDS (Digital Sampling Scale-pixels) とすると, 次の式になる.

$$DSS = k \frac{SS}{\text{距離}(m)} \quad (8)$$

ただし k は定数である. 特徴点を中心にして, 半径 DDS ピクセル範囲内から抽出した特徴画像は同じスケールのテンプレート画像に転換し, 保存されることによって, 物体認識用の物体特徴が学習される. 認識を行う時, 同じ手法を用いて, 画像中すべての特徴を抽出し, 事前が学習した物体の特徴をマッチングを行うことにより, 物体認識を行う. マッチングは, 特徴のテンプレートを用いて, 次の式のようにテンプレートマッチングの平方差を計算する.

$$J(p, q) = \sum_{(i, j) \in N} |T_p(i, j) - T_q(i, j)|^2 \quad (9)$$

ただし, T_q は事前学習した物体特徴のテンプレート, T_p はそれとマッチングを行う特徴テンプレート, N は物体特徴テンプレートのピクセル行列.

テンプレートマッチングに評価する数値は二つがある. 一つ目は平方差の値. これは小さいほどマッチングが成立する. 二つ目は式(10)のように, 最小の平方差を持つテンプレートは, 平方差をすべてほかのテンプレートと比べる. もし一定程度の差がつか

ない場合, 二つのテンプレートのどちらかのマッチングが成立したのかを曖昧とし, 排除する.

$$J(p_i, q) < k^2 J(p_i, q) \quad (10)$$

$$i = 1, 2, \dots, t-1, t+1, \dots, n$$

ただし, p_i は最小の平方差を持つテンプレート. p_i は最小の平方差を持つテンプレート以外のテンプレート. 特徴を用いた画像認識の結果を, 図 6 に示す.



図 5: 特徴点の選定

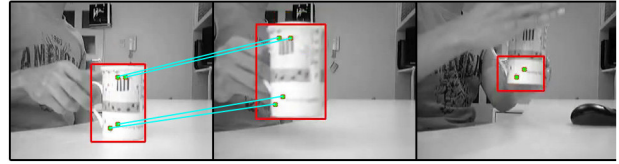


図 6: 特徴を用いた画像認識結果

3 むすび

本稿では, 非特定種類の物体に対し特徴を抽出し, 動的環境下に画像認識を行う手法を提案した. CCD カメラと三次元距離センサーのセンサーフュージョンによって, 従来手法における動的環境下での物体のスケール変化に対処する計算負荷を減らし, リアルタイムに近い画像認識が可能であることを示した. 実験結果 (注 1) では, 特徴ごとの認識率は 82.1%, 複数認識した特徴点の集合による物体の認識率は 91.7% である. 認識スピードは 13 fps.

今後は自動学習の機能を組み込み, 自動的物体の特徴を取り込む. さらに認識における比較的有效な特徴を, 重要度をあげることにより, 無効な特徴を排除する学習方法などを開発する予定である.

(注 1) 式 (6) おけるガウス関数 G_σ の標準偏差 σ を 0.7 に, 式

(8) おける SS を 2cm に, k を 1.5 に, 特徴テンプレートのサイズを 7×7 pixels に, 式 (10) おける k を 0.6 に

謝辞

本研究の一部は相沢直秀さんの助成による。

参考文献

- [1] Oggier T., Lehmann T., Kaufmann R., Schweizer M., Richter M., Metzler P., Lang G., Lustenberger F., and Blanc N. : An all-solid-state optical range camera for 3D-real-time imaging with sub-centimeter depth-resolution(SwissRange), Proc. SPIE, Vol. 5249, pp. 634-545, (2003)
- [2] Harris C. and Stephens M.: A combined corner and edge detector, In Proc. Alvey Vision Conference, (1988)
- [3] Tomasi C. and Knade T.: Detection and tracking of point features, CMU Tech. Rep, CMU-CS-91-132, (1991)