

オプティカルフローと色情報に基づく 掌の検出と追跡によるジェスチャ認識

小渡 悟 †, 星野 聖 ††

† 琉球大学工学部

†† 科学技術振興事業団さきがけ研究21

本稿では、固定単眼カメラで撮影された時系列画像中から、掌を時間差分、空間差分、人物の肌色情報、また抽出領域のオプティカルフローを基に追跡し、その軌跡からジェスチャ認識を行うシステムを提案する。本システムでは、入力画像を対数極座標変換により画像サイズを縮小することで処理の高速化を行っている。また、対象物体の検出と追跡をマーカー等を装着せず、専用のハードウェアを用いずに実時間で安定して行うことができる。評価実験から、移動物体に対して1フレームあたり約40msの実用的な速度で安定して検出することができ、また移動物体の追跡に関しても良好な結果が得られた。

Gesture recognition based on tracking of a human hand with information of the optical flows and colors

ODO Satoru †, HOSHINO Kiyoshi ††

† Faculty of Engineering, University of the Ryukyus

†† Precursory Research for Embryonic Science and Technology,
Japan Science and Technology Corporation

In this paper, we propose a method of gesture recognition in sequential images based on tracking of a human hand with the information of optical flows and colors. Sequential images which are recorded with a single CCD camera are transformed into log polar images in order to reduce the size of images. This system requires no extra peripheral equipments neither such as a digital signal processor board, nor a special optical systems for pre-processing. We carried out experiments to demonstrate the effectiveness of the proposed method. As the experimental results, the tracking and the recognizing were successfully accomplished with high accuracy. The system can track the target with processing time of about 40 milliseconds per a frame with an available personal computer without extra peripheral equipments.

1 はじめに

CCDカメラなどの機械視覚を用いた移動物体の検出と追跡は、侵入者の検知やゲートを通行人や自動車の通行量の計数、自動車のナンバープレートの認識などの自動監視、環境に応じて自らを認識し、予定または教示された行

動を修正しながら目的を果たす自律移動車や自律ロボット、人間の身振り・手振りなどのジェスチャや顔の表情を入力とするユーザーインターフェースなど様々な研究がなされている [1]-[7].

その中でもジェスチャ認識や顔の表情認識は、従来の計算機のキーボードやマウスに加え、更に快適に利用できる人間と計算機のイ

インターフェースとして注目されており、様々な研究がなされている。これらの研究では、認識の容易さからマーカを体や衣服に付けて認識する方法があるが、これはユーザに不自然な使用感を与えてしまい快適なものとはいえず、データグローブ、磁気式モーションキャプチャ [8]、関節角度検出センサといったセンサを体に装着する方式と変わらない。またマーカなど用いない方法では、赤外線カメラなど特殊カメラを用いる方式 [9]、[10] や CCD カメラを複数台用いる方式 [11] など、また計算コストの面から見ても、専用のハードウェアを付加することなしには実時間システムを構築するのは難しい。

本稿では、固定単眼カメラで撮影された時系列画像中において、各フレームに対し対数極座標変換 (Log polar mapping; 以下, LPM) を対象物体を中心に施し、差分法と肌色情報による対象物体の位置決め、対象物体の Optical Flow から移動ベクトルを推定し、その軌跡と掌の形状からジェスチャ認識を行うシステムを提案する。

本システムの特徴は、一つ目は LPM により網膜画像を生成することで、少ない画素数で高解像度、広域の作業範囲を得るだけでなく、画像のデータ量を減らし画像処理の計算時間を減少させている点、二つ目は、網膜での特性と高次視覚野への形態視経路と空間視経路の知見から、動き情報を扱う部分と形状情報を扱う部分の二つに分け、その二つを統合してジェスチャ認識を行う点、三つ目は、LPM を用いていることから、対象物体の回転や拡大縮小にスケーリング不変性があり、また変換中心に対する回転不変性をもっている点である。

以下、2章においてヒトの視覚特性について簡単に解説し、3章において移動物体の検出・追跡、ジェスチャ認識について述べる。4章において本システムによる実験結果を述べ、5章において本論文をまとめる。

2 視覚特性

サルの視野に図 1(a) に示すようなパターンを提示すると、視覚 1 次野の活動度の高くなる

神経細胞の分布は、図 1(b) のようになる [12]。このように霊長類の視覚系は、網膜と大脳視覚領との写像関係を見ると、中心部を重視した不均一構造をしており、注目点近傍で密で視野周辺では粗となるサンプリングを行っている。なお、網膜は単に網膜像を電気信号に変換して脳に送り出しているのではなく、空間的微分や時間的微分に相当する処理など、ある程度の処理を行った後に、第一次視覚野 (V1) に信号を送っている。

網膜からの視覚情報は大きく分けて、網膜から視床枕の神経核を通り、後頭部にある V1 を経て側頭葉に向かう経路と、V1 から頭頂葉に向かう経路の 2 つの流れがある。図 2 に視覚神経系の主要経路を示す。前者は物体の色や形を識別し、物体を認識する形態視の経路と呼ばれ、後者は、物体の運動や視野全体の動きを解析して物体と自己との位置関係や自己の姿勢を把握する空間視の経路と呼ばれている。

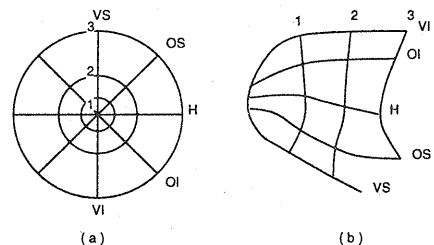


図 1: 網膜座標と視覚 1 次野の座標系

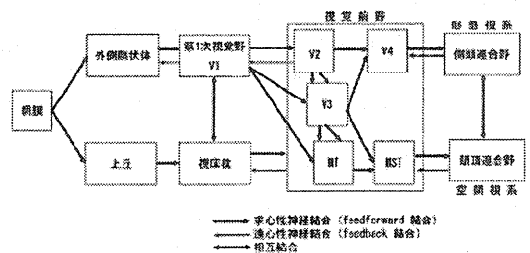


図 2: 視覚神経系の主要ルート

3 ジェスチャ認識システム

3.1 ジェスチャ認識アルゴリズム

固定単眼カメラで撮影された時系列画像の各フレームの画像を、網膜と大脳視覚領との写像関係を近似した網膜画像に変換する。Schwartzはこの網膜 - 皮質間の関係は、LPM によって近似できる [13] としており、本稿においても、このLPMを用いて網膜画像を生成する。次に、網膜画像から時間差分、空間差分、肌色情報を基に輪郭画像を生成し、この輪郭画像をもとに移動物体の位置を決定する。また、移動物体の移動先の推定には、時間差分により生成した画像よりオプティカルフローを計算することによりおこなう。そして、その軌跡を辞書内のデータと比較することでジェスチャ認識を行う。また肌色情報を用いて掌の形状を抽出し形状によるジェスチャ認識をも行う。処理の流れを図3に示す。

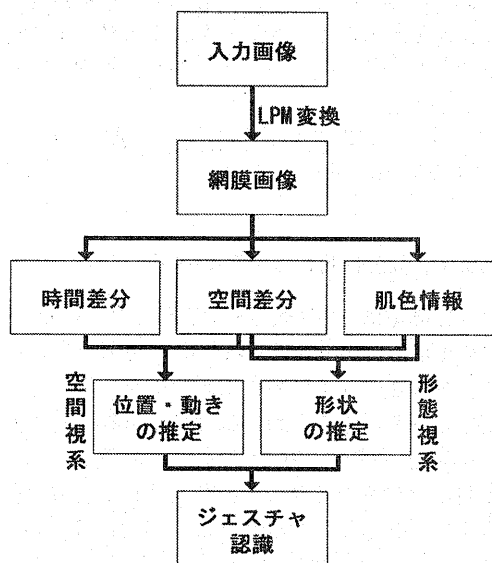


図 3: 処理の流れ図

3.2 網膜画像の生成

本稿はLPMを用いて網膜画像を生成する。この変換を用いることにより、中心視・周辺視

という機能的に分化した視覚を実現し、中心部で高解像度を保ち、周辺部で大まかな状況を抑えるということが実現できる [14]。このようなヒトの視覚系を機械視覚に取り入れると、少ない画素数で高解像度、広域の作業範囲を得ることができ、さらに画像のデータ量を減らすことができるので、計算時間を減少させることもできる。

直交画像での座標 $I(x, y)$ を複素平面 Z とし、そのときの複素平面上の点を $z = x + iy$ と表現する。つぎにLPMの座標 $L(p, q)$ も同様に複素平面 W とし、そのときの複素平面上の点を $w = p + iq$ と表現する。このとき、LPMは式(1)となる。

$$w = \log(z + \alpha) \quad (1)$$

ここで、 α は原点が特異点となるのをさけるための並進量である。

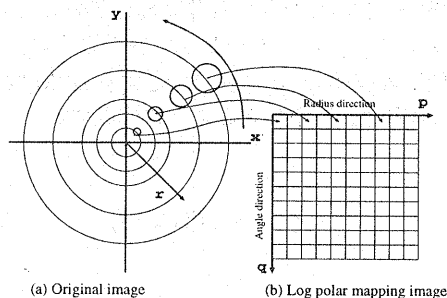


図 4: 対数極座標変換

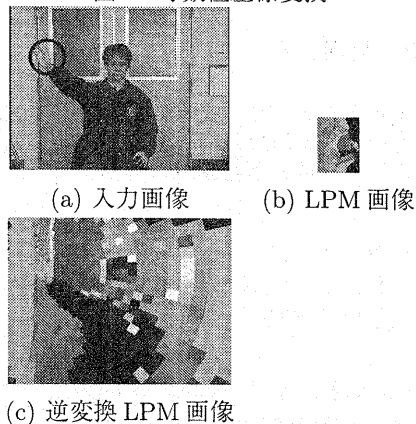


図 5: 対数極座標変換例

LPMにより、原画像 I は角度方向、半径方向の各成分に分解される。図4からも分かるように、半径方向のサンプリングは対数的なものとなるため、周辺部の情報は中心部に比べると大まかに捉えられている。そのため、中心部で高解像度を保ちつつ、周辺部にいくにつれ解像度を下げていくため、空間の状況を大まかに捉えデータ量を大幅に削減し、また処理時間を減少させることができる。LPMの例として、図5(a)の掌の部分(画像中の○印の部分)に着目した場合のLPM後の画像を図5(b)に示す。またその逆変換を図5(c)に示す。

本稿では、通常のCCDからの取り込みを行うため、入力画像からLPMを行う際、各画素がLPM画像のどの画素に対応しているかを計算する必要がある。そこで変換処理の扱いを容易にするために、あらかじめ前処理の段階で各画素の対応点を求め、ルックアップテーブルを作成する[15]。このようにすることで変換処理の扱いが容易になり、計算時間の短縮になる。ルックアップテーブルは、式(2)のように入力画像 I の座標 (x, y) から、対応するLPM画像 L の座標 (p, q) を与えるように作成する。

$$L(p, q) = LUT(I(x, y)) \quad (2)$$

3.3 移動物体の位置推定

3.3.1 時間差分

時間的差分として、式(3)に示すように連続した2フレーム間の差分をとることにより、時刻 t の静止領域と運動領域の分割を行う。

$$L_1(p, q, t) = \begin{cases} 1, & |L(p, q, t) - L(p, q, t-1)| \geq th_t \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

図6に示すように、運動領域 M_t は、生成動領域 G_t 、継続動領域 K_t 、消滅動領域 V_t の和集合となっている。これら G_t, K_t, V_t 領域が、Optical Flowを算出する領域となる。

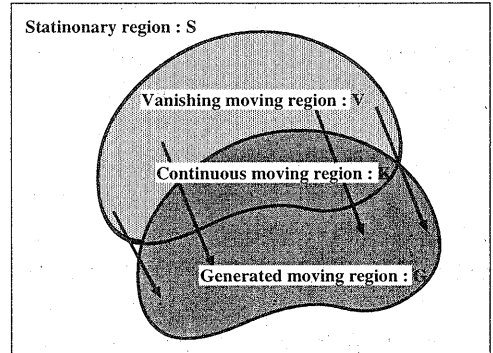


図6: 時間差分画像

3.3.2 空間差分

空間的差分として、時刻 t における画像に対して、式(4)に示す 3×3 のSobelフィルタを施すことで、画像中のエッジを抽出する。式中の $L_2(p, q, t)$, $L_{HS}(p, q, t)$, $L_{VS}(p, q, t)$ は、それぞれ時刻 t における座標 (p, q) の空間的差分値、 p 方向Sobelオペレータによる値、 q 方向Sobelオペレータによる値を表している。

$$L_2(p, q, t) = \begin{cases} 1, & \sqrt{L_{HS}(p, q, t)^2 + L_{VS}(p, q, t)^2} \geq th_s \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

3.3.3 肌色情報による領域選択

入力画像はRGB表色系で表現されているが、このRGB表色系では、各値間の相関が高いため、明るさの変化に大きく影響を受けるといった問題がある。そのため本稿では、明るさの変動に影響をうけず、RGBそれぞれの値が1対1で対応づけられる $L^*u^*v^*$ 表色系(CIE1976表色系)に変換し、その中の u, v 平面を用いる。

図7に、図5(a)の u, v 平面での色分布を示す。図中の矢印部が肌色領域を表している。あらかじめ同じ環境下で撮影した複数の画像の肌色領域の平均色分布を図8に示す。これから肌色情報の平均値 $M(\bar{u}, \bar{v})$ 、分散共分散行列 C を求める。これより、時刻 t における座標 (p, q) の色情報を $L_3(p, q, t)$ とすると、肌色情報による領域の選択は式(5)に示すようになる。

$$L_3(p, q, t) = \begin{cases} 1, & (L_c - M)^T C^{-1} (L_c - M) \geq th_c \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

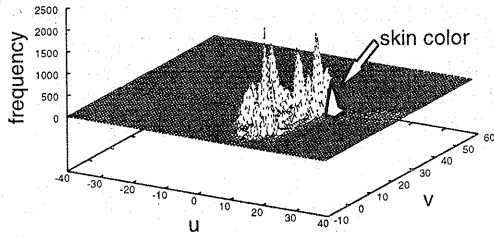


図 7: 図 5(a) の色分布

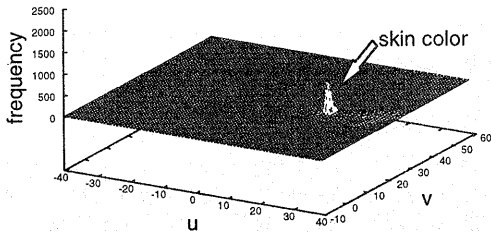


図 8: 肌色領域の平均色分布

3.3.4 輪郭画像

式 (3) から式 (5) より, 式 (6) に示す輪郭画像を得る.

$$L_d(p, q, t) = \begin{cases} 1 & , \sum_{i=1}^3 L_i(p, q, t) = 3 \\ 0 & , \text{otherwise} \end{cases} \quad (6)$$

この得られた輪郭画像 L_d から 1 次モーメントを算出することにより, 時刻 t における入力画像 I の掌の位置 $P(t)$ が算出される.

3.4 移動物体の運動推定

3.4.1 オプティカルフローの算出

画像中の動領域を検出し追跡するため, 式 (3) で得られた連続した 2 フレーム間の差分画像に対して, 一般化勾配法 [17] によりオプティカルフローを算出する.

時刻 t における代表ベクトル $F(t)$ は, 式 (7) に示す一般化勾配法により得られた Optical Flow $F(u(p, q, t), v(p, q, t))$ の平均 Optical Flow (μ_p, μ_q)

を算出することにより得る.

$$F(t) = (\mu_p, \mu_q) \quad (7)$$

$$\mu_p(t) = \frac{1}{N} \sum_i^N u_j(p, q, t)$$

$$\mu_q(t) = \frac{1}{N} \sum_j^N v_j(p, q, t)$$

ただし, 算出されたオプティカルフローの信頼性の評価値を評価式 [18] から求め, 信頼性が小さい点ではオプティカルフローの計算を行わない. またコントラストが小さい点でもオプティカルフローの信頼性がないのでオプティカルフローの計算を行わないとした.

t 時における推定ベクトル $EF(t)$ は, 式 (8) に示すように現在の代表ベクトルと過去 2 フレーム分の代表ベクトルの平均値から推定する.

$$EF(t) = \frac{1}{3} \sum_{i=0}^2 F(t-i) \quad (8)$$

3.5 探索基準点の決定

$t+1$ 時の探索基準点は, 移動物体の位置 $P(t)$ と推定ベクトル $EF(t)$ から式 (9) として表される. また式 (10) から $t+1$ 時の探索基準点を中心点とした探索半径 $r(t+1)$ を決定する. ただし, $R_1 > R_2$ である.

$$gp(t+1) = P(t) + EF(t) \quad (9)$$

$$r(t+1) \begin{cases} R_1 & , \max(\sigma_p^2, \sigma_q^2) > th_R \\ R_2 & , \max(\sigma_p^2, \sigma_q^2) \leq th_R \end{cases} \quad (10)$$

$$\sigma_p^2 = \frac{1}{N} \sum_i^N (u_i(p, q, t) - \mu_p(t))^2$$

$$\sigma_q^2 = \frac{1}{N} \sum_j^N (v_j(p, q, t) - \mu_q(t))^2$$

3.6 ジェスチャの推定

3.6.1 軌跡からのジェスチャの推定

それぞれのジェスチャの開始位置と終了位置は, 予め定めた基準位置と同じ位置とする. そのためジェスチャ推定のための特徴ベクトルは, 開始位置から終了位置までの掌の追跡軌跡の x 座標と y 座標の値を用いる.

ジェスチャの推定は、式 (11) に示す未知の特徴ベクトル $\{S : s_1(x, y), s_2(x, y), \dots, s_i(x, y), \dots, s_N(x, y)\}$ と辞書内の参照される特徴ベクトル $\{T : t_1(x, y), t_2(x, y), \dots, t_j(x, y), \dots, t_M(x, y)\}$ とのコストを DP マッチングにより算出し、その中から最小のものを選択することにより行う。

$$\min_C \{DP^{(c)}(S, T)\} \quad (11)$$

$$C = \{\alpha, \beta, \dots\} \quad :(\text{カテゴリ})$$

s_1, s_2, \dots, s_i と t_1, t_2, \dots, t_j が最もよく合致するときのコスト $f(i, j)$ は式 (12) に示す漸化式により算出でき、これを再帰的に行うことにより最終的な最小コストを算出する。

ここで $d(s_i, t_j)$ は、 s_i と t_j との類似度を表している。

$$\begin{cases} f(i, j) = \min \{ f(i-1, j) + 2, \\ f(i, j-1) + 2, \\ f(i-1, j-1) + d(s_i(x, y), t_j(x, y)) \} \\ f(0, 0) = 0 \\ f(i, 0) = 2 \cdot i \\ f(0, j) = 2 \cdot j \end{cases} \quad (12)$$

ただし、

$$d(s_i, t_j) = \left\{ |s_i(x) - t_j(x)|^2 + |s_i(y) - t_j(y)|^2 \right\}^{\frac{1}{2}}$$

3.6.2 形状からのジェスチャの推定

時刻 t における掌の領域の抽出として、式 (5) による肌色情報による領域選択後、肌色領域のクラスタリングを行い、その中の最大領域を掌とする。図 9 に、入力画像と掌領域を抽出した LPM 画像の例を示す。

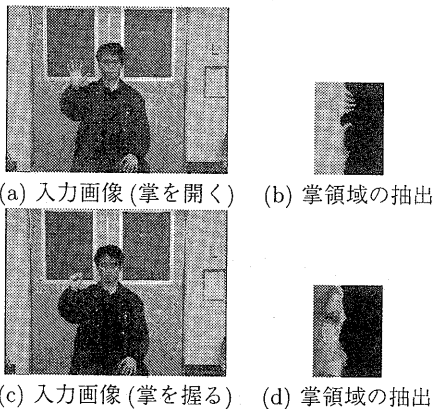


図 9: 掌領域の抽出例

形状からのジェスチャの推定は、式 (13) に示す未知の LPM 画像 S と辞書内の参照 LPM 画像 T とのコストを式 (14) に示す残差逐次検定法により算出し、その中から最小のものを選択することにより行う。またそのときの角度方向シフト量 n より、参照 LPM 画像と相対回転角度を算出する。

$$\min_C \{ER^{(c)}(S, T)\} \quad (13)$$

$$C = \{\alpha, \beta, \dots\} \quad :(\text{カテゴリ})$$

$$ER = \min_N \left\{ \sum_j \sum_i |S(i, j) - T(i, j - n)| \right\} \quad (14)$$

$$N = \{0, 1, \dots\} \quad :(\text{角度方向のシフト量})$$

3.7 評価実験

3.7.1 実験条件

本システムの追跡性能を評価するために評価実験を行った。評価実験にて使用した動画は、通常の照明環境下の室内で撮影した画像サイズ 320 × 240 画素、RGB 各 256 階調、毎秒 30 フレームの AVI 形式の動画ファイルを用いた。

3.7.2 軌跡からのジェスチャ認識実験

実験に用いた基準位置とジェスチャを、図 10 に示す。実験に用いたジェスチャは、動作の開始位置と終了位置はすべて図 10(a) 基準位置になるようにし、ジェスチャとしては、(b) 挙手、(c) バイバイ、(d) 右から左、(e) 左から右の 4 種類である。

実験では、参照ベクトルは、あらかじめ被験者より採取したジェスチャから作製したものを用いた。

評価実験に用いた動画は、図 10(b) から図 10(e) までの動作を連続で行い、総フレーム数 377 フレームとした。連続した軌跡からの入力ベクトルの切り出しは、掌が基準位置近辺にあるとき動作が開始、または終了したと見做し、切り出しを行った。

表 1: 抽出フレーム区間

ジェスチャ	区間 (総フレーム数: 377)	
	開始フレーム番号	終了フレーム番号
挙手	1	65
バイバイ	73	170
右から左	186	260
左から右	285	358

抽出された各ジェスチャのフレーム区間を、表 1 に示す。ただし、掌が基準位置にある場合のフレーム区間は除いてある。また、各ジェスチャにおいて

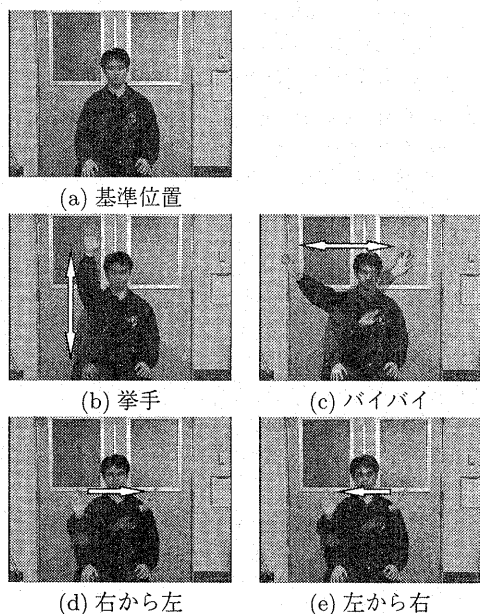


図 10: 実験に用いたジェスチャ

の追跡過程のスナップショットを図 11 に示す。評価実験において、移動物体に対して 1 フレームあたり約 40ms の実用的な速度で安定して検出することができ、また移動物体の追跡、抽出されたフレーム区間に対してジェスチャの推定に関しても良好な結果が得られた。

3.7.3 形状からのジェスチャ認識実験

評価実験に用いたジェスチャは、掌を開いている場合と握っている場合の 2 通りとした。実験では、参照 LPM 画像は、あらかじめ被験者より採取したジェスチャから作製したものをを用いた。

各ジェスチャにおいての実験結果を、スナップショットとして図 12 に示す。ただし、掌を握っている場合は◎印、掌を開いている場合は○印とした。また、掌を開いている場合は、指先が上を向いた状態を参照 LPM 画像としたので、図 12(b), (c), (d) 中の△印は指先の方角を表している。評価実験において、ジェスチャの推定に関して良好な結果が得られた。

4 まとめ

本稿では、固定単眼カメラで撮影された時系列画像中において、各フレームに対し LPM を掌を中心に施し、差分法と肌色情報にて対象物体の位置を、

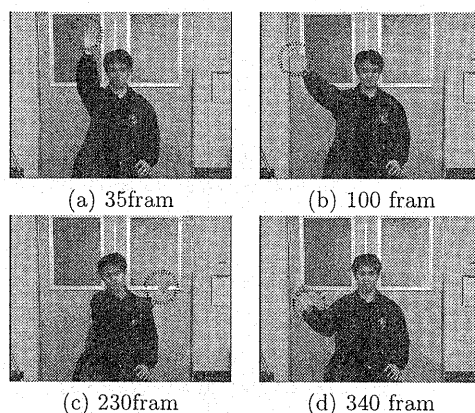


図 11: 掌の追跡過程のスナップショット

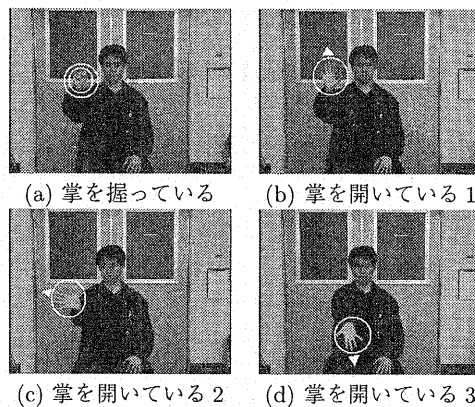


図 12: 各ジェスチャのスナップショット

対象物体の Optical Flow から移動ベクトルを推定し、その軌跡からのジェスチャ認識と形状からのジェスチャ認識を行うシステムを提案した。本システムの特徴は、一つ目は LPM により網膜画像を生成することで、少ない画素数で高解像度、広域の作業範囲を得るだけでなく、画像のデータ量を減らし画像処理の計算時間を減少させている点、二つ目は、網膜での特性と高次視覚野への形態視経路と空間視経路の知見から、動き情報を扱う部分と形状情報を扱う部分の二つに分け、その 2 つを統合してジェスチャ認識を行う点、三つ目は、LPM を用いていることから、対象物体の回転や拡大縮小にスケーリング不変性があり、また変換中心に対する回転不変性をもっている点である。

評価実験から、移動物体に対して 1 フレームあたり約 40ms の実用的な速度で安定して検出することができ、また移動物体の追跡、抽出されたフ

レーム区間に対してジェスチャの推定に関しても良好な結果が得られた。

謝辞

本研究の一部は、通信・放送機構 ギガビットネットワーク利活用研究開発制度の援助を受けて実施された。

参考文献

- [1] 谷口慶治, 水沢利栄, 飯田徹, 山田裕久, “スキヤの動きの画像処理的解析,” 映像情報メディア学会誌, vol.43, no.12, pp.1370-1374, 1989.
- [2] 渡辺嘉二郎, 小林一行, 小山明美, “複数の移動体の軌跡追跡,” 計測自動制御学会論文集, vol.27, no.4, pp.394-401, 1991.
- [3] 遠藤利生, 鳥生隆, 吉田真澄, “時空間に関する滑らかさ制約を用いたオプティカルフローの算出,” 電子情報通信学会論文誌, vol.J74-D-II, no.12, pp.1678-1685, 1991.
- [4] 稲川淳, 前嶋司, “動画像を用いた滑らかさ条件による柔物体の追跡,” 電子情報通信学会論文誌, vol.J77-D-II, no.2, pp.362-369, 1994.
- [5] 村瀬洋, V.V.Vinod, “局所色情報を用いた高速物体探索-アクティブ探索-,” 電子情報通信学会論文誌, vol.J81-D-II, no.9, pp.2035-2042, 1998.
- [6] 興梠正克, 村岡洋一, “グローバルなアフィン動きパラメータの実時間推定手法,” 電子情報通信学会論文誌, vol.J82-D-II, no.7, pp.1161-1170, 1999.
- [7] 足立佳久, 中西知, 久野義徳, 島田伸敬, 白井良明, “顔の視覚情報を用いた知的車椅子,” 日本ロボット学会誌, vol.17, no.3, pp.423-431, 1999.
- [8] 福井一夫, “モーションキャプチャ,” 映像情報メディア学会誌, vol.51, no.8, pp.1120-1123, 1997.
- [9] 三原功雄, 山内康晋, 土井美和子, “モーションプロセッサを用いたビジョン型インターフェースの構築,” インタラクティブシステムとソフトウェア VII(WISS'99), 近代科学社, pp.33-42, 1999.
- [10] 小林貴訓, 佐藤洋一, 小池英樹, “Enhanced Desk のための赤外線画像を用いた実時間指先認識インタフェース,” インタラクティブシステムとソフトウェア VII(WISS'99), 近代科学社, pp.49-54, 1999.
- [11] 松本吉央, 小笠原司, Alexander Zelinsky, “リアルタイム視線検出・動作認識システムの開発,” 電子情報通信学会技術研究報告 PRMU, vol.99, PRMU150-164, pp.9-14, 1999.
- [12] R.B.H. Tootell, et al., “Deoxyglucose analysis of retinotopic organization in primate striate cortex,” Science, vol.218, pp.902-904, 1982.
- [13] E.L. Schwartz, “Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding,” Vision Research, vol.20, no.8, pp.645-668, 1980.
- [14] 青西亭, 福島邦彦, “網膜と皮質の不均一性を考慮した注視点移動モデル,” 電子情報通信学会論文誌, vol.J78-D-II, no.9, pp.1363-1371, 1995.
- [15] 大城尚紀, 丸典明, 西川敦, 宮崎文夫, “網膜と皮質間の写像関係に基づく両眼追跡,” システム制御情報学会論文誌, vol.10, no.6, pp.287-296, 1998.
- [16] 亀田能成, 美濃導彦, 池田克夫, “差分画像を利用した人体の動作認識,” 電子情報通信学会技術研究報告 PRU, vol.165, PRU95-98, pp.115-120, 1995.
- [17] M.V.Srinivasan, “Generalized gradient schemes for the measurement of two-dimensional image motion,” Biological Cybernetics, vol.63, pp.421-431, 1990.
- [18] 岡田隆三, 白井良明, 三浦純, 久野義徳 “オプティカルフローと距離情報に基づく移動体追跡,” 電子情報通信学会論文誌, vol.J80-D-II, no.6, pp.1530-1538, 1997.