

赤外線カメラで撮影した画像による空中での手書き文字入力システム

6 J-4

園田 智也 村岡 洋一

早稲田大学 理工学研究所

1. はじめに

本稿では、夜間などの照明が十分でない環境におけるジェスチャ認識の手法として、赤外線ライトと赤外線カメラを用いて撮影した画像を処理することで、利用者が空中で描く文字を計算機に入力するシステムについて述べる。

画像処理によるジェスチャ認識は、人と計算機が対話するための要素技術として、広く研究されてきた [1, 2]。しかし、従来手法は、固定カメラでの撮影を前提とし、背景や照明条件が一定である必要や、利用者が特定のマーカー等を身に付ける制約があった。

一方、装着型ビデオカメラと装着型ディスプレイで構成されるウェアラブルコンピュータ [3, 4] では、カメラは固定されておらず、背景や照明の条件も、使用する場面で毎回異なる。また、利用者は特定のマーカー等を身に付けられない方が望ましい。そこで、従来のシステム [4] では、利用者の手の領域を求めため、肌の色情報を利用する手法が提案された。

これに対して、本研究が対象とする撮影環境は、照明量が十分でない状況であり、色情報を用いた手法では、手の動きを認識することが困難である。

そこで、本研究では、画像フレーム間の輝度差分を用いて手の動きを認識する手法を提案し、赤外線ライトと赤外線カメラを用いて得られる映像を処理することで、昼夜問わずに利用できるシステムを実現する。

本システムによって、空中での手書き文字の認識率を計測したところ、約 78% であった。

2. ウェアラブルコンピュータ環境

図 1 に、本研究におけるウェアラブルコンピュータ環境を示す。利用者は装着型の赤外線ライト、赤外線カメラおよびディスプレイを身に付ける。また、画像処理用の計算機を携帯する。

赤外線カメラでは、赤外線ライトの照射によって映し出される光景を撮影し、その映像は計算機に送られる。計算機では、その映像に注釈などの付加情報を加え、ディスプレイに出力する。本システムでは、付加情報として、利用者が入力した文字を表示している。

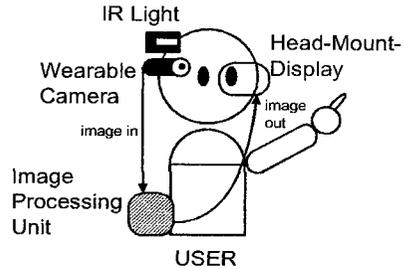


図 1: ウェアラブルコンピュータ環境

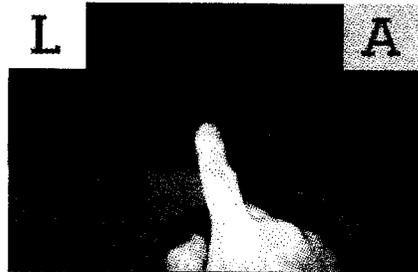


図 2: 赤外線映像

3. 撮影環境

赤外線ライトは赤外線カメラの近辺に設置し、相対位置は固定する。また、文字入力の際には、背景には移動物体が写っていないことを前提とし、カメラは、ほとんど揺らさないこととする。この制約は、利用者にとっては、地面や壁を向いて利用することで容易に実現でき、システム側では精度の高い認識を行うことができる利点が得られる。撮影対象となる利用者の手とカメラとの距離は約 30cm~50cm 程度とする。

4. システムの仕様

図 2 は、本システムのディスプレイ上の表示画像を示す。画像中の右上の文字は、利用者が入力できる文字の種類を 'A' (英字) や '0' (数字) 等で示し、左上の文字は入力された文字を示す。利用者は、人差し指や親指を立て、画像フレーム内に 1 つの文字を書くことで、計算機に入力する。利用者はディスプレイの中央に表示される仮想的なボタンを押す動作により入力開始できる。文字・記号を書き終えた際、手を画像フレーム外に出すことで、システムは認識を終了する [5]。

4.1 画像処理

本研究で利用する赤外線画像は、以下の特徴を持つ。

1. 利用者の手は、画像フレーム中のすべての物体よりも手前にあり、赤外線ライトの照射によって、背景よりも比較的明るく表示される。
2. 赤外線ライトにより、手の影が背景に映し出される。

以上の特徴を考慮した上で、画像処理に [5] で用いた手法を適用し、手の動きを求める。

まず、利用者の手の領域を求める。入力開始の際、利用者は画像フレーム中央のボタンを押す動作を行い、手は画像フレーム下方の中央領域にあると想定できる。そこで、特徴 1. より、その領域でフレーム全体の輝度値の平均値より高い画素を手の領域とみなす。

次に、隣接画像フレーム間の各画素における輝度値の差分により、その重心位置を求め、手の動きを特定する。ここで、上述の特徴 2. により、利用者自身の手の影が背景物体に写され、その動きが処理に影響を及ぼすことを考慮する必要がある。

そこで、本研究では、特徴 1. で求めた手の領域内の各画素の輝度の平均値 ave を求め、画像フレーム中の画素が ave に近い輝度を持たない場合、処理の対象としない。時刻 $t, t-1$ における各画像フレームの画素の輝度値の差分 $|I(t, x, y) - I(t-1, x, y)|$ を重みとしたとき、その重心の位置 $G(gx, gy)$ を次式で求める (ただし、 $I(t, x, y) < ave - \epsilon$ または $I(t, x, y) > ave + \epsilon$ のとき $I(t, x, y) = 0$)。

$$X = \sum_i \sum_j i * |I(t, i, j) - I(t-1, i, j)|$$

$$Y = \sum_i \sum_j j * |I(t, i, j) - I(t-1, i, j)|$$

$$W = \sum_i \sum_j |I(t, i, j) - I(t-1, i, j)|$$

これより、求めた $G(gx, gy) = (X/W, Y/W)$ の軌跡のパターンと、予め記録している文字の軌跡パターンを、DP マッチングで照合し、文字を特定する。

利用者が手を画像フレームの外に出し、重心位置 G の移動が無い状態が、一定時間 T の間続いた場合、システムは認識を終了する。

5. 実験

赤外線映像を用いて、空中での手書き文字認識を行う実験を行い、照明が蛍光灯の場合と比較した。

5.1 実験環境

本実験では、ウェアラブルコンピュータ環境を想定し、デジタルビデオカメラ (SONY DCR PC-100:IR 波長感度 780nm) を利用者が顔周辺に持ち、映像をデ

スクトップ型の計算機 (CPU AMD:Athlon 500MHz, RAM:128M, OS:LINUX, Graphic Board:BT878) で処理し、計算機のディスプレイに表示した。画像は、256 階調画像 (サイズ: 320*240) で処理した。また、 $T = 0.5sec$, $\epsilon = 5$ とした。この実験における画像表示のスループットは、平均 6.7 fps、遅延時間は、平均 8.8 フレーム (約 293 msec) であった。

5.2 照明条件による比較実験

一筆書きで表記した英数字 [5] 36 文字を、被験者 1 人が各 10 回ずつ合計 360 回入力する実験を、赤外線画像、および通常の画像 (蛍光灯照明) のもとで各々行い比較した。表 1 に結果を示す。

赤外線映像を用いた結果は、蛍光灯を用いた結果よりも、全体で約 10% 低下していた。これは、赤外線ライトの照射範囲に起因し、画像フレーム周辺では、手が比較的薄暗く表示されることが原因であると考えられる。このため、文字の設計および認識方法を再考することで、より精度向上を計られると考えられる。

表 1: 空中での手書き文字認識率

	赤外線	蛍光灯
英字 (赤外線)	71.0 %	88.1 %
数字 (赤外線)	81.1 %	89.0 %
英数字総合 (赤外線)	78.3 %	88.3 %

6. まとめと今後の課題

本稿では、赤外線映像を用いた空中での手書き文字を認識するシステムを述べた。また、その精度を、蛍光灯を照明として使用した場合と比較実験した。今後は、文字の書き方の設計、および認識処理部分の設計の研究を進め、精度向上を計る予定である。

参考文献

- [1] T. Darrell and A.Pentland: *Space-Time Gestures*, Proceedings IJCAI'93 Looking at People Workshop, Aug. 1993.
- [2] 高橋 勝彦, 関 進, 小島 浩, 岡 隆一: ジェスチャ動画像のスポッティング認識, 電子情報通信学会論文誌, D-II, vol.J77-D-II, No8, pp.1552-1561, Aug. 1994.
- [3] Tony Jebara, Cyrus Eyster, Josh Weaver, Thad Starner and Alex Pentland. "Stochastic: Augmenting the Billiards Experience with Probabilistic Vision and Wearable Computers", Proceedings of the International Symposium on Wearable Computers, October 1997.
- [4] 蔵田 武志, 興沼 正克, 栗田 多喜夫, 村岡 洋一, 坂上 勝彦: 気の利いたウェアラブルビジョンシステムの実現に向けて, 信学技報, PRMU-99-270, pp.65-71, March 2000.
- [5] 園田 智也, 村岡 洋一: 空中での手書き文字入力システム - ウェアラブルコンピュータ環境での入力インターフェース -, 情報処理学会 インタラクシオン 2001 論文集, pp.3-10, 2001.