

画像からの手振りの検出

— 低解像度画像の時系列に対する FFT の利用 —

Detection of Waving Hands from Images

- Application of FFT to Time Series of Low-Resolution Images -

○入江 耕太 (中央大学) 梅田 和昇 (中央大学)

Kota IRIE, Kazunori UMEMA

Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Abstract: This paper proposes a method of detecting waving hands from images as a method for man-machine interface. FFT is applied to time series of intensity images. The images are converted to low-resolution ones, and FFT is applied to each pixel of the low-resolution images. The proposed method is robust to lighting condition and individual difference of skin color, because it doesn't use color information at all. Experiments show the stability and robustness of the proposed method.

Key Words: Image Processing, Gesture Recognition, FFT, Low-Resolution Image, Man-Machine Interface

1 序論

人間と機械との間の自然なマン・マシン・インタフェースを実現するには、操作者の存在および操作する意思を認識することが重要である。その一手法として、操作者にとって自然な動作である手振りをを用いることが考えられる。既に提案している画像からの手振りの検出手法¹⁾では、カラー画像から肌色抽出を行うことで手領域を抽出し、この重心座標に対して FFT を行い²⁾、振動成分を評価することで手振りの検出を実現している。この手法では、肌色抽出が不可欠なため、照明条件や肌色の個人差の影響を受けやすい。

本論文では、低解像化した濃淡画像のすべての画素に対して FFT を行い、その中で手振り領域を検出する手法を提案する³⁾。カラー情報を用いていないため、照明条件や肌色の個人差に対してロバストである。さらに、あらかじめ手領域を抽出するといった画像処理は不必要であり、高速な計算が可能な FFT を適応することで、極めて簡潔な処理を実現している。

文献3)では識別手法に最短距離法を用いているが、本論文では線形判別法を用い、ノイズに対するロバスト性を向上させる。

2. 濃淡値の時系列に対する FFT

取得した濃淡画像を $m \times n$ の画素数に低解像度化し、全画素の濃淡値 $I(i, j, t)$ を時系列で取得する。ただし、 $i=1, 2, \dots, m$, $j=1, 2, \dots, n$, t はフレーム数である。Fig.1 に模式図を示す。

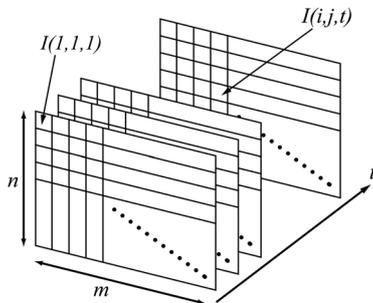


Fig.1 $m \times n$ 画素の時系列の濃淡値

Fig.2(b)に低解像度化した画像の例を示す。図中の矩形で示す手振り領域に対応する画素では、Fig.2(c)に示す様に時間変化に伴って濃淡値 $I(i, j, t)$ が手の時と背景の時で変化する。この濃淡変化は、ある一定の周期を持つため、定量化する手法として FFT を用いる。各画素における濃淡値に対して FFT を行い、スペクトルから手振りを評価する。得られるスペク

トルの例を Fig.2 (d)に示す。なお、計算量の軽減や蛍光灯などのちらつきによる誤認識を防ぐために、前のフレームとの濃淡値の差が I_{dif} 以上ある画素のみに対して FFT を行う。

$$|I(i, j, t+1) - I(i, j, t)| \geq I_{dif} \quad (1)$$

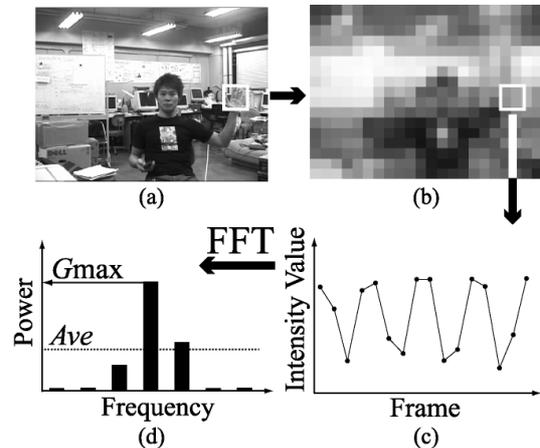


Fig.2 濃淡値の時系列に対する FFT

3. 手振りの認識手法

3.1. クラスタの形成

全画素の濃淡値に対して FFT を行い、スペクトルから特徴空間を定め、手振りのクラスタを形成する。特徴量にはスペクトルにおけるパワの最大値 G_{max} と、スペクトルの平均値 Ave を用いる。 G_{max} , Ave はそれぞれ式(2), 式(3)で与えられる。なお、 N はサンプリング数、 W は DFT における回転子、 $\max(f)$ は f の最大値を表す。

$$G_{max} = \max \left(\frac{1}{N} \sum_{k=1}^{N-1} I_k W_{nk} \right) \quad (2)$$

$$Ave = \frac{2}{N} \sum_{i=1}^{N/2} G_i \quad (3)$$

3.2. 判別分析による識別

認識には G_{max} , Ave 空間における線形判別法を用いる。特徴ベクトル $\mathbf{x} = [G_{max}, Ave]^T$ 、手振りのクラス ω_1 、手振り以外のクラス ω_2 とし、それぞれのパターン平均を \mathbf{m}_1 , \mathbf{m}_2 とすれば、変動行列 \mathbf{S}_1 , \mathbf{S}_2 は式(4)で定義される。

$$\mathbf{S}_i = \sum_{\mathbf{x} \in \mathcal{X}_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \quad (4)$$

次に、2クラスの全特徴ベクトルを用いて、クラス内変動

行列 \mathbf{S}_W とクラス間変動行列 \mathbf{S}_B は式(5)と式(6)で定義される。 n_i は ω_i のパターン数、 \mathbf{m} は全パターンの平均を表す。

$$\mathbf{S}_W \equiv \mathbf{S}_1 + \mathbf{S}_2 = \sum_{i=1,2} \sum_{\mathbf{x} \in \mathcal{X}_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^t \quad (5)$$

$$\mathbf{S}_B \equiv \sum_{i=1,2} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^t \quad (6)$$

ここで d 次元特徴空間から 1 次元空間の変換を表す行列を \mathbf{A} とすると、式(7)の関係が導かれる。

$$\mathbf{g}(\mathbf{x}) = \mathbf{A}^t \mathbf{x} + a_0, \quad \mathbf{A} = \mathbf{S}_W^{-1} (\mathbf{m}_1 - \mathbf{m}_2) \quad (7)$$

$$a_0 = \frac{\tilde{m}_1 \tilde{\sigma}_2 + \tilde{m}_2 \tilde{\sigma}_1}{\tilde{m}_1 + \tilde{m}_2} \quad (8)$$

ただし、 \tilde{m}_i , $\tilde{\sigma}_i$ は、決定境界線の法線に G_{\max} , Ave 空間の点を投影したときの平均値と標準偏差である。

より確実な認識を行うために、フレーム毎での $\mathbf{g}(\mathbf{x})$ の値を認識条件に加味する。識別用の変数を D_i とし、連続 c フレームの間 $\mathbf{g}(\mathbf{x}) < 0$ ($x \in \omega_1$) を満たした場合 ($J=1$) に手振りと認識させた。これを式(9), (10)で与える。 t はフレーム数である。

$$\text{if } \mathbf{g}(\mathbf{x})_t > 0 \text{ then } D_t = 0 \text{ else } D_t = 1 \quad (9)$$

$$J = \prod_{k=t}^{t+c} D_k \quad (10)$$

4. 手振り検出実験

画像の入力、および低解像度化には画像処理ソフト (MVTec HALCON) を用い、FFT 計算、認識処理等は PC (Pentium 4 1.4GHz) で行った。また、手振り位置を注視させるために Pan-Tilt 機能を搭載したカメラ (SONY EVI-G20) を使用した。

サンプリング数 N は計算コストと要求される周波数分解能から 16 とした。式(1)における前フレームとの濃淡値の差 I_t は、蛍光灯のちらつきなどの影響を避けるために 5 とした。また、式(10)における変数 c は検出の応答性と安定性を考慮し、5 とした。サンプリング周期は約 80 [ms] であった。

4.1. 線形判別関数の決定

実験により手振りのクラス ω_1 と手振り以外の動作のクラス ω_2 を形成した。ここで、手振り以外の動作とは、室内を前後左右に歩行する動作や、室内における無造作な動作とした。計測距離 3 [m] から 8 [m] における ω_1 , ω_2 の G_{\max} -Ave の分布をそれぞれ Fig.3 と Fig.4 に示す。データ数はそれぞれ約 2500 である。式(5)におけるクラス内変動行列 \mathbf{S}_W , パターン平均 \mathbf{m}_i , は以下の通りである。

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} -0.02657 \\ 0.04141 \end{bmatrix}^t \begin{bmatrix} G_{\max} \\ Ave \end{bmatrix} - 0.143$$

ここで、算出した $\mathbf{g}(\mathbf{x})$ について評価するために、誤判別率を定義する。クラス i のサンプル数を n_i , 誤って判別された数を n_{w_i} , 誤判別率を p_i とすると、

$$p_i = \frac{n_{w_i}}{n_i} \quad (11)$$

と定義できる。 ω_1 , ω_2 の値をもとにそれぞれ算出すると、 $p_1 = 4.20\%$ (手振りの画素を手振りでない画素と認識する確率), $p_2 = 2.28\%$ (手振りでない画素を手振りの画素と認識する確率) である。

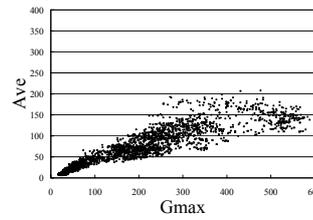


Fig.3 手振りの分布 ω_1

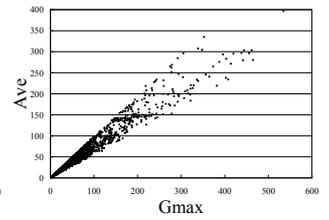


Fig.4 手振り以外の分布 ω_2

4.2. 手振りの検出

4.2.1. 距離、照明による認識率

被験者 5 人を対象に、計測距離と室内照明を変化させて実験を行った。実験環境は蛍光灯のみの照明とし、手領域付近の鉛直面照度は約 60~190 [lux] (条件 1) と約 250~315 [lux] (条件 2) で行った。認識率の算出にあたり、手振り方法に関して次のことを条件とした。

- (1) 画像中の任意の場所で手振りを約 2 秒間行う。
- (2) 手振りを止めてから約 2 秒おいて次の手振りを再開する。

上記の動作を 20 回試し、そのうち約 2 [s] 以内で手振りを検出した回数から認識率を算出した。実験結果を Table 1 に示す。

Table 1 認識率

Distance	60-190 lux	250-315 lux
4m	99%	96%
5m	100%	96%
6m	100%	97%
7m	96%	92%
8m	91%	83%

全てので高い認識率が得られた。照明条件で認識率に違いが生じた理由として、背景の照度と手の照度の関係が挙げられる。本実験では、照明を暗くした場合、背景となる壁面の濃淡値が極めて小さくなり、照明を明るくした場合は、壁面が十分に照らされる。その結果、明るくすると、手と背景で濃淡の差が生じにくい結果となった。なお、壁面周辺の鉛直面照度は条件 1 で約 30~50 [lux], 条件 2 で約 230~300 [lux] であった。その他の認識失敗原因として、手振りの幅が小さい場合が挙げられる。

4.2.2. 誤認識に関する考察

誤認識に関する評価を行った。実験は 1 人の人間が手振り以外の動作を行ったり、無造作に室内を歩行したりすることを条件とした。また、歩行範囲としてカメラからの距離が 1m から 8m の間で行い、計測時間は 10 分間である。以上の条件で実験を行った結果、誤認識することは無かった。

5. 結論

色情報を用いず、照明条件に対してロバスト、かつ高速な手振り検出手法を提案した。低解像度の濃淡変化に対して FFT を行うことで、低い計算コストで手振りの検出手法を構築した。実験では距離に対するロバスト性も確認できた。今後の課題として、手法の汎用性から、様々なマン・マシン・インタフェースとしての応用が期待される。

参考文献

- 1) 入江耕太, 梅田和昇, 中西裕一, 画像処理を用いた手振り検出のインテリジェントルームへの適用, 日本機械学会ロボメック講演会'01 講演論文集, 2A1-C8, 2001.
- 2) 安富敏, 森英雄, 清弘智昭, 歩行のリズムに基づく歩行者検出の一手法, 信学論, Vol.J78-D-II, No.4, pp.608-617, 1995.
- 3) 入江耕太, 梅田和昇, 画像からの手振りの検出 -濃淡値の時系列に対する FFT の利用-, 第 19 会日本ロボット学会学術講演会講演論文集, 2L22, 2001.