

画像からの手振り検出

－ 濃淡値の時系列に対する FFT の利用 －

○入江 耕太（中央大学） 梅田 和昇（中央大学）

Detection of Waving Hands from Images - Application of FFT to time series of intensity values -

*Kota IRIE, Kazunori UMEDA

Abstract — This paper proposes a method of detecting waving hands from images as a method for man-machine interface. The method applies FFT to time series of intensity values, and realizes robustness for lighting condition and high speed. Experiments show that this method is robust for change of distance.
Key Words: Image Processing, Gesture Recognition, FFT, Man-Machine Interface, Pattern Recognition

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、色情報が不要で照明条件に対してロバストであり、かつ高速な手振り検出手法を提案する。低解像度画像の濃淡変化のみに着目し、濃淡変化を定量化するためにFFTを用いる。FFTは高速な処理であり、かつシステムの小型化、ハードウェア化が期待でき、汎用性のあるインタフェースが実現できると考えられる。

2 濃淡値の時系列に対するFFT

取得した濃淡画像を $m \times n$ の適当な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。

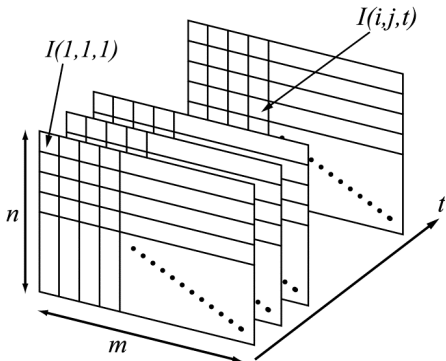


Fig.1 $m \times n$ 画素の時系列の濃淡値

Fig.2 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.3 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.4に示す。



(a)元画像 (b)低解像度化画像

Fig.2 画像の低解像度化

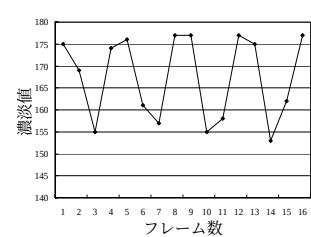


Fig.3 濃淡変化の例

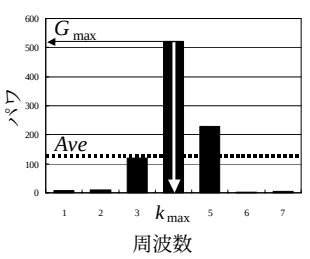


Fig.4 スペクトルと特徴量

3 手振りの認識手法

まず、手振り領域に対応する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。認識時には、各画素で得られた特徴量とのマハラノビス距離を求め、その値が小さい画素がある場合に手振りであると認識する。

特徴量にはパワの最大値 G_{max} と、 G_{max} を平均値 Ave で正規化した値 R を用いる。 R および Ave は(1)式で与えられる。

$$R = \frac{G_{\max}}{Ave}, \quad Ave = \frac{2}{N} \sum_{i=1}^{N/2} G_i \quad (1)$$

N : サンプル数

手振りのクラスタの平均ベクトルを $\mathbf{M}=[M_{G_{\max}} \ M_R]^T$ とし、共分散行列を $\mathbf{V}$ とする。計測された特徴量ベクトル $\mathbf{X}$ とのマハラノビス距離 d_M は次式で与えられる。

$$d_M^2 = (\mathbf{X} - \mathbf{M})^T \mathbf{V}^{-1} (\mathbf{X} - \mathbf{M}) \quad (2)$$

2 章, 3 章で述べた手振り認識処理の流れを Fig.5 に示す。

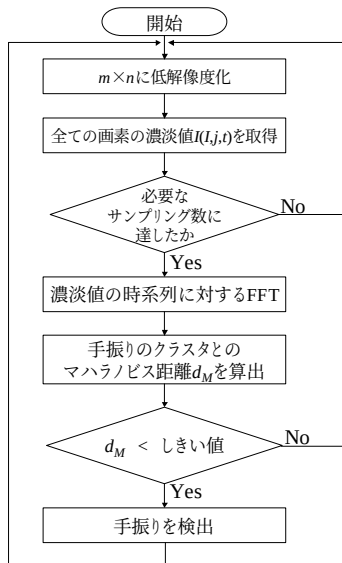


Fig.5 処理の流れ

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。入力画像は, 640×480 画素を 22×17 画素に低解像度化したものを用いた。サンプリング数 N は 16 とし, サンプリング周期は約 80[ms]であった。

4.1 クラスタの形成

実験により手振りのクラスタを形成した。手振りの速さ約 2.5Hz で特徴量の計測を行った。本実験では, 手振りの 91% で $k_{\max}=4$ に集約したため, $k_{\max}=4$ のみでクラスタを形成した。計測距離 2[m] から 7[m] における手振りに対応する画素の $G_{\max}$ - R の分布を Fig.6 に示す。データ数は約 2600 である。Fig.6 において, 分布の塊が二つ形成されている。 $G_{\max}$, R の値が小さい方の塊は, 距離が 6[m], 7[m] の時である。これは, 画像の解像度を低くしたことにより, 手振

りの振幅が 1 画素以内になり, 濃淡値の変化が微小となったためである。よって別々のクラスタを形成し, それぞれの平均ベクトル $\mathbf{M}$, 共分散行列 $\mathbf{V}$ は以下の通りである。

$$\mathbf{M}_1 = \begin{bmatrix} G_{\max} \\ R \end{bmatrix} = \begin{bmatrix} 80.2 \\ 3.42 \end{bmatrix} \quad \mathbf{V}_1 = \begin{bmatrix} 69.9 & 3.61 \\ 3.61 & 0.363 \end{bmatrix}$$

$$\mathbf{M}_2 = \begin{bmatrix} G_{\max} \\ R \end{bmatrix} = \begin{bmatrix} 16.6 \\ 2.54 \end{bmatrix} \quad \mathbf{V}_2 = \begin{bmatrix} 443 & 11.8 \\ 11.8 & 0.938 \end{bmatrix}$$

次に, 2 人の人物が椅子に座り, 2 人の人物が左右に歩行している室内の風景に対し, 約 1 分間計測した時の, 全画素に対する $G_{\max}$ - R の分布を Fig.7 に示す。データ数は約 4500 である。 $G_{\max}$ と R はともに小さい値に集中し, 手振りの場合との区別がほぼ可能である。Fig.6 と Fig.7 を重ねて見た場合, 形成したクラスタ内に手振り以外のデータが約 400 個含まれている。フレーム毎で見た場合, 手振りの動作 (Fig.6) では $G_{\max}$, R は連続的に大きい値をとり, 手振り以外の動作 (Fig.7) では一時的であるため, 両者の区別は可能であった。

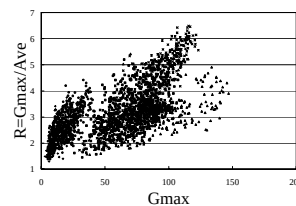


Fig.6 手振りの $G_{\max}$ - R の分布

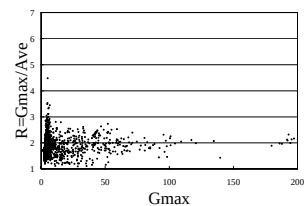


Fig.7 $G_{\max}$ - R の分布

4.2 手振りの検出

被験者 10 人を対象に手振りの検出実験を行った。計測距離が 2[m] から 7[m] で, 実時間での検出が行えた。検出に失敗した例として, 極端に手振りが速い, または遅い場合が挙げられる。また, 手振り以外の動作や蛍光灯のちらつきを手振りとして誤認識することはなかった。

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, 実験による定量的な認識率等の検討が挙げられる。

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.