

濃淡値の時系列変化を利用した画像からの手振りの検出

入江 耕太*¹ 梅田 和昇*²

Detection of Waving Hands from Images Using Time Series of Intensity Values

Kota Irie*¹ and Kazunori Umeda*²

This paper proposes a method of detecting waving hands from images as a key technology 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. Additionally, the method is applied to an omnidirectional vision images and an infrared camera, so as to show the wide applicability of the method.

Key Words: Image Processing, Gesture Recognition, FFT, Low-Resolution Image, Man-Machine Interface, Omnidirectional Vision, Infrared Camera

1. 序 論

現在、我々の生活環境の情報化、インテリジェント化が進んでいる。その一例として、身近な機器である家電製品のネットワーク化が実現し、利便性の向上が図られている。一方、多機能化することで、操作が複雑化するという問題が生じる。特に生活に密着する機器においては、ユーザーが直感的に操作可能なインタフェースが有効であると考えられる。そのためには、人間の自然な行動を利用することが重要である。

我々の日常的なコミュニケーション手段を考えた場合、意識的あるいは無意識的に身振り、手振りなどのジェスチャを頻繁に用いている [1]。そこで、直感的でかつ非接触での操作を可能にするマン・マシン・インタフェースの一つとして、ジェスチャを用いることが考えられる。これまでに、ジェスチャ認識を用いたマン・マシン・インタフェースの研究が数多く報告されている。しかし、これらの多くは、動作者の位置がある程度限定されているため、生活空間での実用性に欠ける。背景差分を適用することで動作者の場所を特定せずにジェスチャ認識を行う手法も提案されている [2] が、人物が複数存在するなど移動領域が複数の場合、適用が困難である。つまり、ジェスチャ認識を用いたマン・マシン・インタフェースを実環境で実現するには、広い空間における動作者の位置の特定や、動作者の操作す

る意思を事前に認識することが必要である。我々が普段注視を促すジェスチャとして、手を左右に振る手振り動作が挙げられる。本研究は、動作者の位置を特定することなどを目的とした、画像から手振りを検出する手法を提案する。

すでに提案されている画像からのジェスチャ認識 [3]~[6] や、手振りの検出手法 [7] 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 [8]、振動成分を評価する [9] [10] ことで手振りの検出を実現している。これらの手法では、肌色抽出が不可欠なため、照明条件や肌の個人差の影響を受けやすい。

本手法では、低解像度化した濃淡画像の各画素に対して時間軸方向の FFT を行い、その中で手振り領域の検出を行う [11]。カラー情報を用いていないため、照明条件や肌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う画像処理は不必要であり、高速な計算が可能な FFT [12] を適用することで、極めて簡潔な処理を実現している。以上のことから、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。

なお、時系列の濃淡変化に対して FFT を適用し画像中から繰り返し動作を認識するというアイデア自体は、文献 [11] とほぼ同時期に、浜田らによっても独立に、映像の索引付けを目的として提案されている [13] [14]。ただし、上記のアイデア以外は本研究の手法とは異なり、また用途・インプリメンテーションも違う。

応用例として本手法を全方位視覚センサと赤外線カメラに対して適用する。全方位視覚センサに適用することで、動作者の操作可能領域が拡大し、より実用的なシステムが構築できると

原稿受付 2002年8月20日

*¹中央大学大学院理工学研究科

*²中央大学理工学部

*¹Graduate School of Science and Engineering, Chuo University

*²Faculty of Science and Engineering, Chuo University

考えられる．一方，赤外線カメラの適用においては，照明の影響を受けず，照明のない部屋や照明条件の悪い部屋などで有効である．

以下，2章で解像度と濃淡変化との関係，3章で濃淡値の時系列に対するFFTの適用手法，4章で手振りの認識手法について述べ，5章で手振りの検出実験，6章で全方位視覚センサ，赤外線カメラへの適用結果を報告する．

2. 画像の低解像度化

濃淡画像中で手振りを行った場合，手領域の画素において，手の濃淡値と背景の濃淡値とで変化が生じる．この時系列の濃淡変化を取得するための前処理として，画像の低解像度化を行う．これにより，計算量の軽減やノイズに対するロバスト性の向上や，滑らかな濃淡変化パターンが得られるといった効果が期待できる．Fig. 1に手振りによる濃淡値の周期変化を示す．

カメラの画角を θ [rad]，取得画像の水平画素数を a [pixel]，計測距離を L [m]，手振り幅の距離を L_H [m]，手振りの動作平面と画像平面とのなす角度を α [rad] とすると，取得画像における手の重心移動量 H [pixel] は，式(1)で表される．Fig. 2に模式図を示す．

$$H = \frac{aL_H \cos \alpha}{2 \tan \frac{\theta}{2}} \cdot \frac{1}{L} \quad (1)$$

低解像度化した画像において，手領域と背景との間で濃淡が変化するためには，手振りの幅が低解像度画像において1画素以上の幅を持たなければならない．標準化画素数を P_{lim} とす

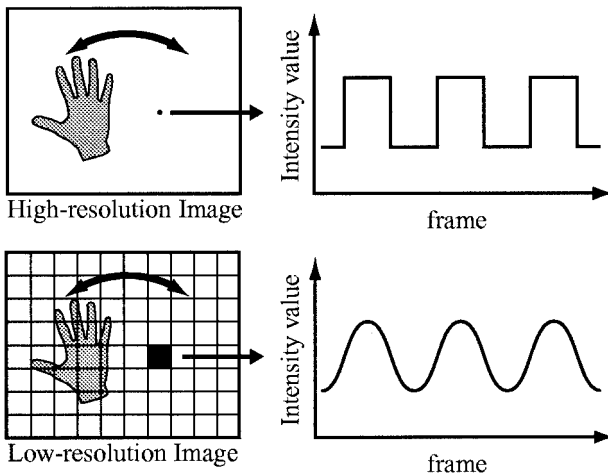


Fig. 1 Cyclic change of intensity values by waving hands

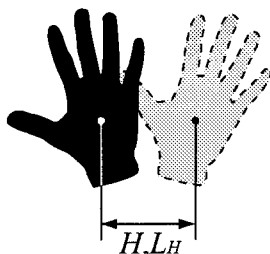


Fig. 2 Width of waving hands

ると， $H > P_{lim}$ が満たされる必要がある．

3. 濃淡値の時系列に対するFFT

2章に基づき，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. 3に模式図を示す．

Fig. 4(a)に取得画像，Fig. 4(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. 5(a)に示すように時間変化に伴って濃淡値 $I(i, j, t)$ が手のときと背景のときで変化する．この濃淡変化は，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．各画素における時間軸方向の濃淡値に対してFFTを行い，スペクトルが

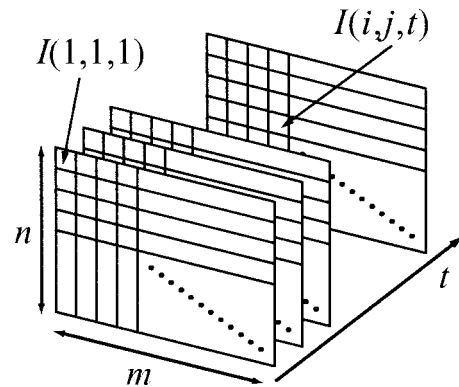
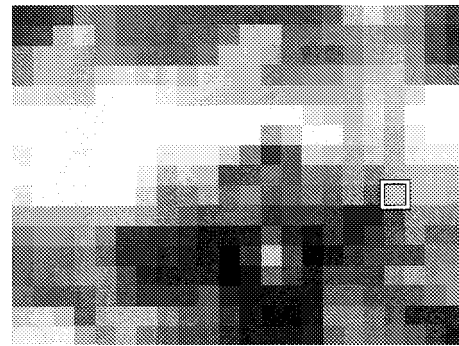


Fig. 3 Time series of low-resolution images



(a) Original image



(b) Low-resolution image

Fig. 4 Conversion of image to low-resolution

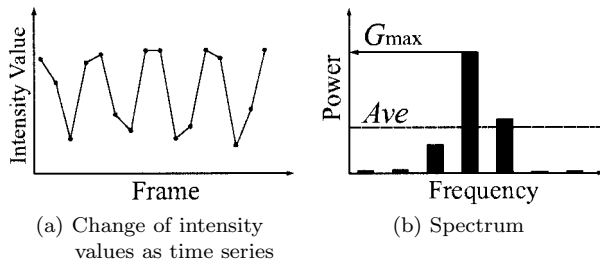


Fig. 5 Application of FFT to time series of intensity values

ら手振りを評価する．得られるスペクトルの例を Fig. 5 (b) 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 u フレームにおける濃淡値の最大値と最小値の差が I_{dif} 以上ある画素のみに対して FFT を行う．

$$\max(I_{t-u+1}, \dots, I_t) - \min(I_{t-u+1}, \dots, I_t) \geq I_{dif} \quad (2)$$

4. 手振りの認識手法

4.1 クラスタの形成

低解像度化した画像の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ーの最大値 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 (3)，(4) で与えられる．なお， N はサンプリング数， W は DFT (Discrete Fourier Transform) における回転子を表す．

$$G_{\max} = \max_n \left(\left| \sum_{k=1}^{N-1} I_k W_N^{nk} \right| \right) \quad (3)$$

$$Ave = \frac{2}{N} \sum_{n=1}^{N/2} G_n \quad (4)$$

4.2 判別分析による識別関数の決定

識別には G_{\max} ， Ave 空間における線形判別法 [15] を用いる．特徴ベクトル $\mathbf{x} = [G_{\max}, Ave]^T$ ，手振りのクラス w_1 ，手振り以外のクラス w_2 とし，それぞれのパターン平均を \mathbf{m}_1 ， \mathbf{m}_2 とする．さらに，クラス内変動行列を S_W とすると，線形識別関数 $g(\mathbf{x})$ は式 (5) で与えられる．

$$g(\mathbf{x}) = \mathbf{A}^T \mathbf{x} + a_0 \quad (5)$$

ただし，

$$\mathbf{A} = S_W^{-1} (\mathbf{m}_1 - \mathbf{m}_2) \quad (6)$$

である．また， a_0 の決定には各クラスの標準偏差で内分する手法 [16] を適用する．

$$a_0 = \frac{\tilde{m}_1 \tilde{\sigma}_2 + \tilde{m}_2 \tilde{\sigma}_1}{\tilde{m}_1 + \tilde{m}_2} \quad (7)$$

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．

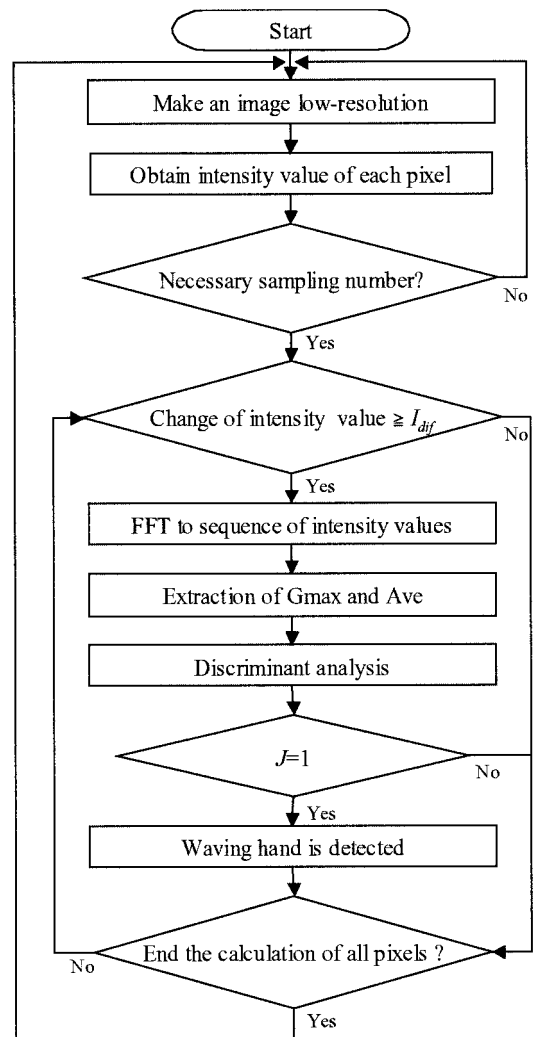


Fig. 6 Flow chart of detecting waving hands

4.3 手振りの認識

より確実な認識を行うために，フレームごとでの $g(\mathbf{x})$ の値を認識条件に加味する．識別用の変数を D_t とし，連続 c フレームの間 $g(\mathbf{x}) < 0$ ($\mathbf{x} \in w_1$) を満たした場合 ($J = 1$) に手振りとして認識させる．これを式 (8)，(9) で与える． t はフレーム数である．

$$\begin{aligned} \text{if } g(\mathbf{x})_t < 0 \text{ then } D_t &= 1 \\ \text{else } D_t &= 0 \end{aligned} \quad (8)$$

$$J = \prod_{k=t}^{t+c-1} D_k \quad (9)$$

このとき，手振りを始めてから認識するまでに要する時間 T は，サンプリング周期を τ とすると式 (10) で与えられる．

$$T = \tau(N + c) \quad (10)$$

2章，3章，4章で述べた手振り認識処理の流れを Fig. 6 に示す．

Table 1 Frequency of waving hand

Subject	A	B	C	D	E	Average
Frequency[Hz]	3.7	2.7	2.9	3.3	3.6	3.2

5. 手振り検出実験

提案手法の有効性を検証するために実験を行った。カメラにはSONY製のEVI-G20、画像入力ボードには、Leutron製のPicPort Colorを用いた。また、画像の低解像度化には画像処理ソフト(MVTec製HALCON)を用い、FFT計算、認識処理等はPC(Pentium 4 1.4 [GHz])で行った。

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(2)における前フレームとの濃淡値の差 I_{dif} は5とし、フレーム数 u は4とした。また、式(9)におけるフレーム数 c は検出の応答性と安定性を考慮し、4とした。

5.1 手振りの速さ

本研究で扱う手振り動作は、普段我々が注視を促す場合に行う動作として想定している。そこで、5人の被験者に対して注視を促すことを想定して手振り動作を行ってもらい、手振りの振動数を計測した。手振りを30往復するのに要する時間から振動数を算出した。結果をTable 1に示す。この結果から、検出の対象とする手振りの振動数は2.5 [Hz] から4.0 [Hz]の間とする。

5.2 解像度の決定

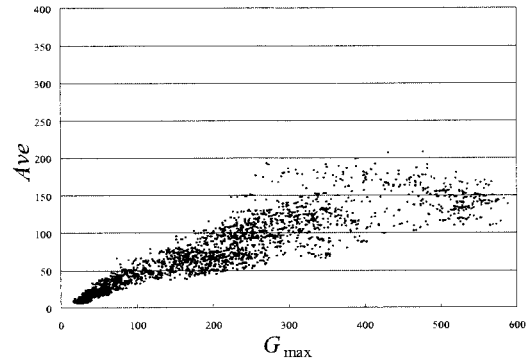
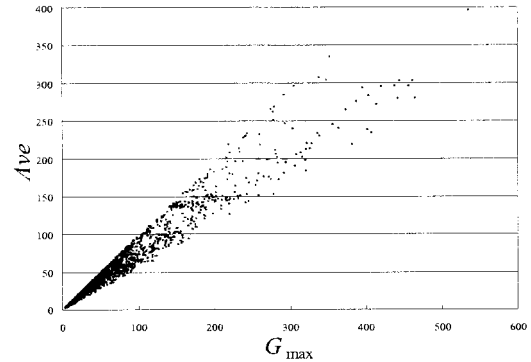
式(1)に基づき、解像度を決定した。各パラメータの値は、 $a = 640$ [pixel]、 $\theta = \text{約}\pi/4$ [rad]、 $L_H = 0.3$ [m]である。また、手振りはカメラに向かって行うこととし、 α は0 [rad]とした。最大の計測距離 $L = 8$ [m]とした場合、 $H = 29$ [pixel]となる。このことから $P_{im} = 25$ [pixel]とし、 25×25 [pixel]を1画素として低解像度化する。以上のことから、処理対象画像の解像度は 25×19 [pixel]である。一連の処理に要したサンプリング周期は約80 [ms]であった。すなわち、手振りを開始してから認識されるまでに要する時間 T は式(10)より約1.6 [s]となる。また、5.1節で定めた振動数に対応するスペクトルの周波数は3~5となる。

5.3 線形判別関数の決定

実験により手振りのクラス w_1 と手振り以外の動作のクラス w_2 を形成した。手振りのクラス形成に関しては、あらかじめ手振りの位置に対応する画素を定め、その範囲内で手振りを行い、 G_{\max} と Ave を取得した。また、手振り以外の動作は、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 w_1, w_2 の G_{\max} - Ave の分布をそれぞれFig. 7(a)とFig. 7(b)に示す。データ数はそれぞれ約2,500である。式(6)におけるクラス内変動行列 S_W 、パターン平均 m_i は以下の通りである。

$$S_W = \begin{bmatrix} S_{G_{\max}, G_{\max}} & S_{G_{\max}, Ave} \\ S_{G_{\max}, Ave} & S_{Ave, Ave} \end{bmatrix} \\ = \begin{bmatrix} 1296.3 & 4616.3 \\ 4616.3 & 2310.2 \end{bmatrix}$$

(a) Distribution of waving hands: ω_1 (b) Distribution of other motions: ω_2 Fig. 7 Distribution of G_{\max} - Ave

$$m_1 = \begin{bmatrix} \overline{G_{\max}} \\ \overline{Ave} \end{bmatrix} = \begin{bmatrix} 214.6 \\ 75.6 \end{bmatrix} \\ m_2 = \begin{bmatrix} \overline{G_{\max}} \\ \overline{Ave} \end{bmatrix} = \begin{bmatrix} 61.5 \\ 48.6 \end{bmatrix}$$

よって、式(6)の変換行列 A は、

$$A = \begin{bmatrix} -0.02657 \\ 0.04141 \end{bmatrix}$$

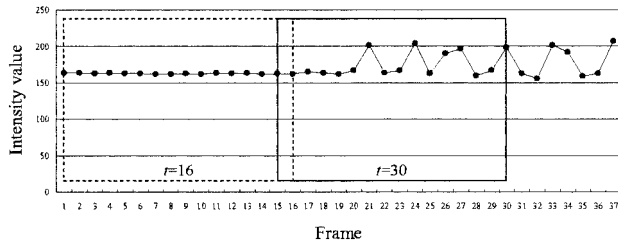
となり、式(7)により $a_0 = -0.143$ となる。以上のことから、線形識別関数 $g(x)$ は次式で与えられる。

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

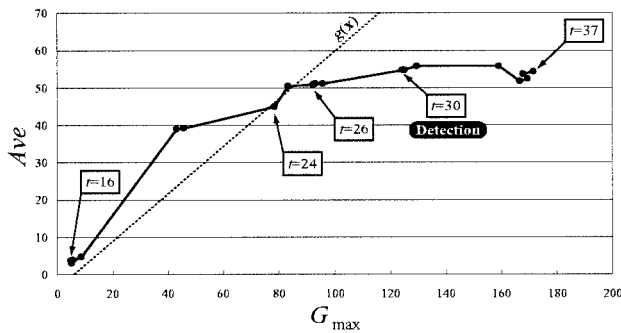
ここで、算出した $g(x)$ について評価するために、誤判別率を定義する。クラス i のサンプル数を n_i 、誤って判別された数を n_{wi} 、誤判別率を p_i とすると、

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

と定義できる。 w_1, w_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。

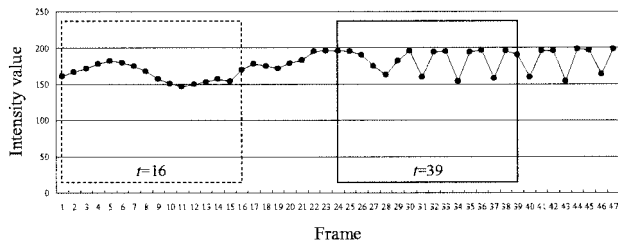


(a) Change of intensity values

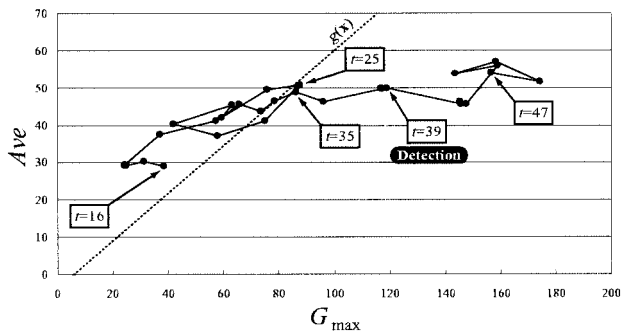


(b) Transition of features

Fig. 8 Transition of features for waving hands (1)



(a) Change of intensity values



(b) Transition of features

Fig. 9 Transition of features for waving hands (2)

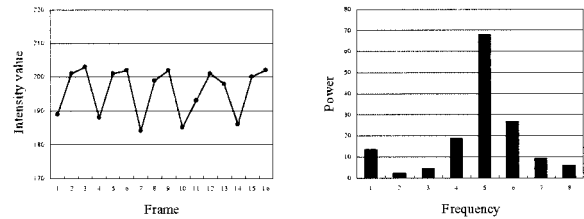
5.4 手振りの検出

5.4.1 セグメンテーション

提案手法では、4章で示したように、式(5)の線形識別関数が定められた連続フレームの間負であった場合に手振りであると認識される。すなわち、手振り動作のセグメンテーション(切り出し)は自動的に行われることになる。これを検証するための実験を行った。結果を Fig. 8, Fig. 9 に示す。Fig. 8 は静止状態から手振りを始めた場合、Fig. 9 はランダムな動作から手振りを始めた場合である。手振り動作が始まると特徴ベクトル

Table 2 Recognition rate of waving hands

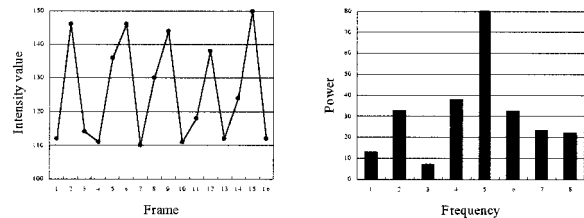
Distance	60 -190 lx	250-315 lx
4m	99%	96%
5m	100%	96%
6m	100%	97%
7m	96%	92%
8m	91%	83%



(a) Change of intensity values

(b) Spectrum

Fig. 10 Distance: 4 [m] (Bright room)



(a) Change of intensity values

(b) Spectrum

Fig. 11 Distance: 4 [m] (Dark room)

が遷移し、Fig. 8 で手振り開始後 0.8 [s] ($t = 30$), Fig. 9 で手振り開始後 0.9 [s] ($t = 39$) で手振りであると認識する。なお、Fig. 9 (b) で手振り動作でない $t = 25, 26$ で線形識別関数が負になっているが、ここでは連続 4 フレームで負となった場合に手振りであるとしているため、手振りであるとは認識されない。以上、手振り動作のセグメンテーションが適切に行われた。また、式(10)で与えられる時間 T より短い時間で認識しうることが示された。

5.4.2 距離、照明による認識率

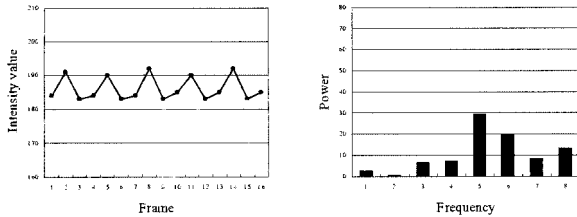
被験者 5 人を対象に、計測距離と室内照明を変化させて実験を行った。実験環境は蛍光灯のみの照明とし、手領域付近の鉛直面照度は約 60 ~ 190 [lx] (室内の照明を約 1/3 にした場合) と約 250 ~ 315 [lx] (室内の照明をすべてつけた場合) で行った。認識率の算出に当たり、手振り方法に関して次のことを条件とした。

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

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

計測距離が 4 [m] から 8 [m] 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。Fig. 10 ~ 13 に計測距離 4 [m] と 8 [m] における時系列の濃淡変化とそのスペクトルの例を示す。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小さくなる。照明を明るくした場合は、壁面が十分に照らされ、その

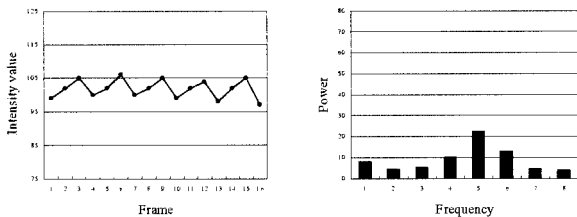
結果、手と背景で濃淡の差が生じにくい結果となった。なお、壁面周辺の鉛直面照度は室内の照明を約1/3にした場合で約30~50 [lx]、室内の照明をすべてつけた場合で約230~300 [lx]であった。その他の認識失敗原因として、手振りの幅が小さい場合が挙げられる。



(a) Change of intensity values

(b) Spectrum

Fig. 12 Distance: 8 [m] (Bright room)



(a) Change of intensity values

(b) Spectrum

Fig. 13 Distance: 8 [m] (Dark room)

計測距離 4 [m] と 8 [m] における手振りの検出例を Fig. 14, 15 にそれぞれ示す。図中において、矩形で示している部分が手振り検出位置である。

5.5 誤認識に関する実験

誤認識の要因として、人間の手振り以外の動作と蛍光灯などの周期性を持つ発光体が考えられる。よって、以上の二つの場合について誤認識に関する実験を行った。

5.5.1 手振り以外の動作

手振り以外の動作として、以下の動作を定義する。

- (1) 腕を回す動作
- (2) 膝の屈伸動作
- (3) 歩行

以上の動作を組み合わせながら10分間連続して行い、誤認識した回数を数えた。なお、動作範囲はカメラからの距離が2 [m] から 8 [m] の間とした。

以上の条件で実験を行った結果、誤認識することはなかった。

5.5.2 周期性を持つ発光体

実環境における周期的な発光体として、以下の機器を用いた。

- (1) 蛍光灯 (100 [Hz])
- (2) TV モニタ (60 [Hz])
- (3) PC モニタ (75 [Hz])

Fig. 16 に実験に用いた機器を示し、図中の矩形で示した画素の濃淡変化とスペクトルを Fig. 17~19 に示す。

蛍光灯においては、濃淡変化がほぼ5以下になっていたため、式(2)のしきい値処理でFFT処理の対象外となった。



(a) Bright room



(b) Dark room

Fig. 14 Example of detecting waving hands (Distance: 4 [m])



(a) Bright room



(b) Dark room

Fig. 15 Example of detecting waving hands (Distance: 8 [m])

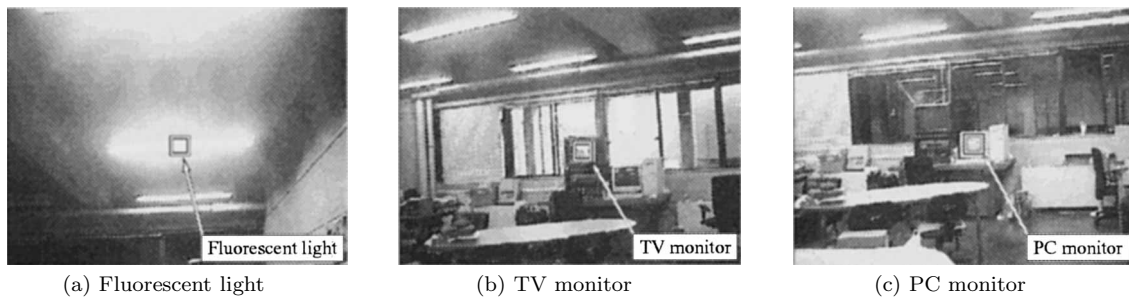


Fig. 16 Examples of luminous bodies with periodicity

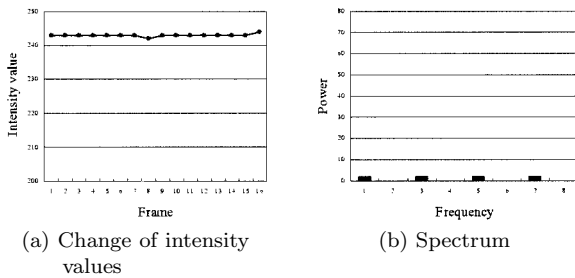


Fig. 17 Fluorescent light

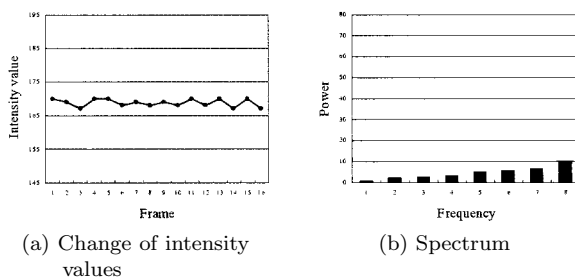


Fig. 18 TV monitor

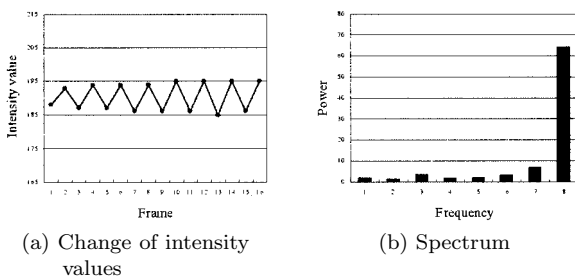


Fig. 19 PC monitor

TV モニタにおいては、濃淡変化が0から7の間であった。式(2)のしきい値処理でFFT処理の対象となる場合があるが、濃淡変化の周期性が見られなかったため、誤認識することはなかった。また、この結果からTVの走査周期による影響は少ないと考えられるが、TVに映される映像に周期的な変化がある場合には誤認識する可能性がある。

PC モニタにおいては、濃淡変化が1~10の間であり、周期性が見られた。しかし、PCモニタのピーク周波数と手振りのピーク周波数とで違いがあるため識別は可能であり、誤認識することはなかった。

以上三つの機器における実際のリフレッシュレートは本シス

テムのサンプリング周期の約5~8倍である。システムのサンプリング周期によってはエリアシング誤差の影響による誤認識が生じる可能性があると考えられるが、本システムではほぼ問題なく動作することが確認できた。

6. 他のカメラへの適用

本手法の特徴として、時系列の濃淡変化のみに着目していることがあげられる。つまり、手振り動作が画像中で濃淡変化として取得できれば適用可能である。このことから、本手法の汎用性を示すために、全方位視覚センサと赤外線カメラへの適用を行った。

6.1 全方位視覚センサへの適用

これまでに述べた手振り検出手法を全方位視覚センサ(Vstone VS-C41N)から得られる全方位画像に対して適用した。一般的に双曲面鏡を用いた全方位視覚センサにおいて、輝度のダイナミックレンジの大きなカメラが必要とされることや、カメラの解像度の低さが大きな問題となる[17]。前者については、室内における使用であることと、時系列の変化値変化を利用しているという手法の性質から、大きな問題とはならない。後者についても、低解像度化を行うといった手法の性質上、問題とならない。実験に用いた入力画像は、 400×400 [pixel]とした。また、手振りはカメラに向かって行うこととし、 α は0 [rad]とした。全方位視覚センサは、画像中心部と周辺部で空間分解能が異なるが、本実験では中心投影など特別な変換は行わず、得られた全方位画像を用いた。さらに、画像全体に対して重み等を持たせず、 40×40 [pixel]に低解像度化した。これは全方位画像の周辺部の空間分解能と手振りの幅を基に決定した。PCにPentium4 2.2 [GHz]を使用し、サンプリング周期は約90 [ms]であった。全方位視覚センサを高さ約1.5 [m]、室内の中央に設置し、鉛直面照度250~315 [lx]の明るい室内とした。Fig. 20に全方位視覚センサによる手振り検出例を示す。全方位視覚センサを中心に半径約3 [m]の範囲で良好な検出結果が得られた。

6.2 赤外線カメラへの適用

次に、赤外線カメラ(日本アビオニクス IR-30、レンズ焦点距離25 [mm])から得られる画像に対して適用した。赤外線カメラは、温度情報を濃淡情報として得られるため、本手法の適用が容易に行える。さらに、照明条件の影響がないため、暗い部屋、明るい部屋問わず安定した認識が可能となる。式(1)に基づき、解像度の決定を行った。各パラメータの値は、 $a = 640$ [pixel]、 $\theta =$ 約0.45 [rad]、 $\alpha = 0$ [rad]、 $L_H = 0.3$ [m]である。また、

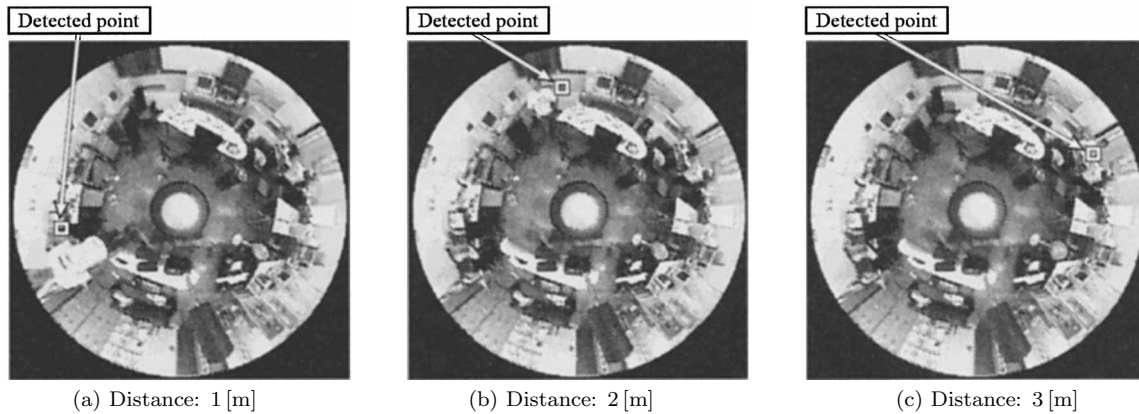


Fig. 20 Example of detecting waving hands using an omnidirectional vision images

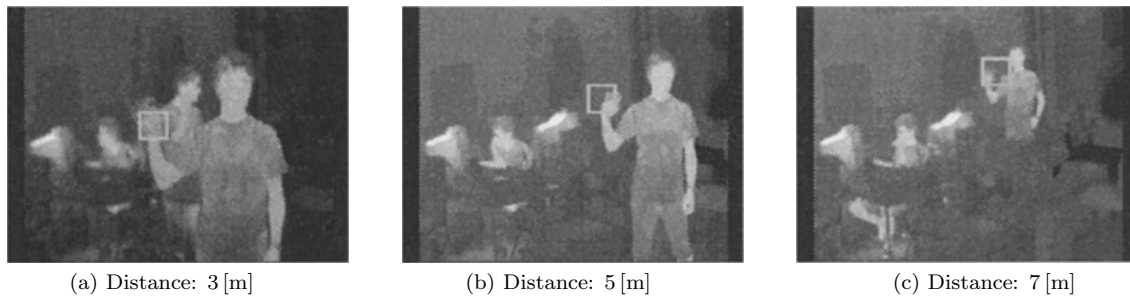


Fig. 21 Example of detecting waving hands using an infrared camera

最大の計測距離 $L = 8$ [m] とした場合, $H = 52$ [pixel] となる. このことから $P_{lim} = 50$ [pixel] とし, 50×50 [pixel] を 1 画素として低解像度化する. 以上のことから, 処理対象画像の解像度は 13×10 [pixel] とした. PC に Pentium4 2.2 [GHz] を使用し, サンプル周期は約 60 [ms] であった. また, 人間の皮膚温はある一定の範囲にあると考えられる. 被験者 3 人の掌の濃淡値を計測した結果, 130 ~ 160 程度であった. よって, 本実験ではこの濃淡値の範囲を最大, 最小のしきい値とし, 前 16 フレーム以内において, しきい値を満たす画素のみを処理対象とした. これにより, 計算コストの削減が可能となる. Fig. 21 に赤外線カメラによる手振り検出例を示す. 図中において, 矩形で示している部分が手振り検出位置である.

7. 結 論

マン・マシン・インタフェースとしての利用を目的とした, 画像からの手振りの検出手法を提案した. 色情報を用いていないため, 照明条件に対してロバストであり, かつ高速な処理を実現した. また, 低解像度の濃淡変化に対して FFT を行うことで, 計算コストを低減させた. 実験では距離に対するロバスト性も確認でき, 安定した認識結果が得られた. また, 全方位視覚センサと赤外線カメラに対して適用し, 本手法の汎用性を確認し, 実験により有効性を示した. 全方位視覚センサを用いることで計測範囲が拡大し, 操作位置を拘束されにくいジェスチャ認識システムが実現すると考えられる. 一方, 赤外線カメラへの適用においては, 照明条件に左右されないため, 照明環境に対してロバストであることを確認した. さらに人間の体温を利用す

ることで, より安定した手振り検出が可能となった.

今後の予定として, インテリジェントルームなどにおけるジェスチャ認識を利用したシステムへの応用が挙げられる. 手振りを認識した後, 動作者の手周辺部を拡大し, より複雑なジェスチャを認識するといった具体例が考えられる. さらに, 手振り位置の色情報を利用することで, 個人差や照明条件にロバストな肌色抽出が行えると考えられる. これまでに肌色情報を用いたジェスチャ認識手法が数多く提案されているが, それらに適用することで, 認識率の向上が期待される.

参 考 文 献

- [1] 黒川隆夫: ノンバーバルインタフェース. オーム社, 1994.
- [2] 西村拓一, 十河卓司, 小木のぶ, 岡隆一, 石黒浩: “動き変化に基づく view-based aspect model による動作認識”, 電子情報通信学会論文誌, vol. J84-D-II, pp. 2212-2223, 2001.
- [3] J. Sherrah and S. Gong: “VIGOUR: A System for Tracking and Recognition of Multiple People and their Activities,” Proc. of the International Conference on Pattern Recognition, 2000.
- [4] P. Hong, M. Turk and T.S. Huang: “Gesture Modeling and Recognition Using Finite State Machines,” IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2000.
- [5] H. Wu, T. Shioyama and H. Kobayashi: “Spotting Recognition of Head Gestures from Color Image Series,” Proc. of the International Conference on Pattern Recognition, pp. 83-85, 1998.
- [6] 山岸崇志, 中西裕一, 梅田和昇: “インテリジェントルーム実現のためのジェスチャ認識”, 画像の認識・理解シンポジウム (MIRU'2000), I, pp. 289-294, 2000.
- [7] 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”, 日本機械学会ロボティクス・メカトロニクス講演会'01 講演論文集, 2A1-C8, 2001.
- [8] 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”, 電子情報通信学会論文誌, vol.J78-D-II, no.4, pp.608-617, 1995.
- [9] R. Cutler and L. Davis: “View-based Detection and Analysis of Periodic Motion,” Proc. of the International Conference on Pattern Recognition, 1998.
- [10] R. Cutler and L.S. Davis: “Robust Real-Time Periodic Motion Detection, Analysis, and Applications,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22, no.8, pp.781-796, 2000.
- [11] 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回日本ロボット学会学術講演会講演論文集, pp.747-748, 2001.
- [12] W. Press, S. Teukolsky, W. Vetterling and B. Flannery: Numerical Recipes in C. Cambridge University Press, 1988.
- [13] 浜田玲子, 佐藤真一, 坂井修一, 田中英彦: “料理映像における繰り返し動作のスポッティング手法”, 電子情報通信学会技術研究報告, パターン認識・メディア理解研究会 PRMU2001, 2001.
- [14] R. Hamada, S. Satoh, S. Sakai and H. Tanaka: “Detection of Important Segments in Cooking Videos,” Proc. IEEE Workshops on Content-Based Access of Image and Video Libraries at Kauai, Hawaii, USA, pp.118-123, 2001.
- [15] R.O. Duda and P.E. Hart: Pattern Classification and Scene Analysis. John Wiley & Sons, 1973.
- [16] K. Fukunaga: Introduction to Statistical Pattern Recognition (2nd ed.). Academic Press, 1990.
- [17] 石黒浩: “全方位視覚の歴史と展望”, 日本ロボット学会誌, vol.20 no.4, pp.373-376, 2002.



入江耕太 (Kota Irie)

1977年10月11日生。2001年中央大学理工学部精密機械工学科卒業。2003年同大学大学院理工学研究科博士前期課程修了。現在、株式会社日立製作所に勤務。(日本ロボット学会正会員)



梅田和昇 (Kazunori Umeda)

1967年2月6日生。1989年東京大学工学部精密機械工学科卒業, 1994年同大学大学院精密機械工学専攻博士課程修了, 博士(工学)。同年中央大学理工学部精密機械工学科専任講師, 1998年より同助教授。ロボットビジョン, センサフュージョン, マン・マシン・インタフェースの研究に従事。精密工学会, 日本機械学会, 電子情報通信学会, 情報処理学会, 日本時計学会, IEEEの会員。(日本ロボット学会正会員)