

[論文]

(2007年12月28日 Accepted)

MRRF 法を用いた複雑背景下における人物の姿勢認識大場 統¹, タン ジュークイ², 金 亨燮², 石川 聖二²

1) 九州工業大学大学院・工学研究科, 2) 九州工業大学・工学部

要約: 本研究では, 画像から人物領域を抽出し, その動作・姿勢を認識する手法を提案する. 抽出にはテクスチャと輝度の情報を用いた背景差分を用い, 動作・姿勢の認識には2次元画像から3次元特徴を判断できる手法を用いる. 多くの手法では実験環境が屋内であることや領域抽出の際の閾値設定が手動であることや, 認識の際, 特定の特徴量の抽出やモデル化が必要であるなどの条件が必要なことが多いが, 本提案手法ではそれらの条件を必要としないため, 抽出から認識までの流れをほぼ自動で行うことができる. 提案法を用いた屋外実験により, その有効性が確かめられた.

キーワード: 背景差分, 固有空間, 認識

Recognizing Human Postures in a**Complex Background Employing MRRF Method**Osamu OOBA¹, Joo Kooi TAN², HyounGSeop KIM², Seiji ISHIKAWA²

1) Graduate School of Engineering, Kyushu Institute of Technology

2) Faculty of Engineering, Kyushu Institute of Technology

Abstract: In this paper, we propose a technique for extracting a human region from images and for recognizing human postures. We employ a background subtraction technique using texture information and gray scale for extracting a human region, and a technique for recognizing a 3-D object from a 2-D image. Many extraction techniques must be performed in the outdoors and need thresholds tuned by manual operations. Moreover many recognition techniques need extraction of specific characteristics or modeling. Since the proposed technique does not necessitate manual operations, it can perform the recognition from extraction automatically. The proposed technique obtained satisfactory results by the experiment performed in the outdoor environments.

Keywords: Background subtraction, Eigenspace, Recognition

Osamu OOBA

Sensuichou 1-1, Tobata, Kitakyushu, 804-8550, Japan

Phone: +81-93-691-7159, Fax: +81-93-691-8380, E-mail: nakano@med.uoeh-u.ac.jp

1. はじめに

近年、防犯カメラの設置台数の増加とともにその監視業務の軽減が強く求められている。それにともない、カメラを使用した人物の抽出や動作認識研究の、セキュリティ分野への応用への期待も増加している。

画像から人物領域を抽出する方法として、人物など動物体を含まない背景画像と動物体を含む画像を比べることにより、動物体領域のみを抽出する背景差分法という手法がある。この手法は基本的に各画素における濃度変化しか見ておらず、色が以ていた動物体は抽出が困難なことや、照明変化が引き起こす濃度変化による誤抽出などの問題点を含んでいる。また、抽出の判断に用いられる濃度変化の閾値設定は使用者の主観に任される部分が多い。これらの問題を解決するため、照明変化に対応した手法[3]や閾値設定が主観によらない手法[4]が提案されているが、いずれも各画素のみにおける情報しか用いない為、異なるテクスチャであっても色が以ていた物体は抽出できないなどの問題が残っている。これに対し本提案手法では、Radial Reach Filter 法[1]というテクスチャ情報に基づく背景差分法を用いる。この手法は光にロバストで、かつ閾値の自動設定が可能であるが、屋内のようなテクスチャの少ない環境では非常に多くのノイズが乗るといった問題がある。そこで本研究では、この手法と従来の背景差分法を複合的に用いることにより、そのノイズを低減する手法を提案する。

人物の認識についても様々な手法[5]が提案されているが、その多くが特定の特徴量を抽出するかモデル化するなどの処理を必要とする。これに対し本提案手法で用いる固有空間法[2]は2次元画像から3次元物体の認識を行う手法であり、3次元特徴の抽出が不要であり、一枚の画像が固有空間上の一点に投影されるので、その距離によって画像の相関を単純に比較することができる。その上、あらゆる方向からの情報を学習することにより、方向によらない姿勢認識が可能となる。本研究ではこの二つの手法を用いて、人物領域の抽出およびその姿勢の認識を行う手法を提案する。本論文では、まず対象の抽出法について説明し、次に固有空間法について述べ、最後に屋外環境における人物の姿勢認識実験の結果を示す。

2. 人物領域抽出

本節では、領域抽出法として、Radial Reach Filter 法(以下 RRF)と従来の輝度値を用いた背景差分法(以下 BS)からなる、Mixed Radial Reach Filter 法(以下 MRRF)について説明する。

2.1 Radial Reach Filter 法

RRF は以下のステップで求めることができる。

1. 背景画像内の任意の着目画素の位置をベクトル $\mathbf{p}=(x,y)$ と定義し、8近傍方向を表すベクトルを $\mathbf{d}_k (k=0,1,\dots,7)$, $\mathbf{d}_0=(1,0)^T, \mathbf{d}_1=(1,1)^T, \dots,$

$\mathbf{d}_7=(1,-1)^T$ と定義すると、閾値 T_c 以上の明度差

をもつ最近傍画素 (以後ペア) までの距離 r_k (以後リーチ) は次式で表される。

$$r_k = \min \{ r \mid |f(\mathbf{p} + r\mathbf{d}_k) - f(\mathbf{p})| \geq T_c \} \quad (1)$$

ここで f は背景画像、また $k=0,1,\dots,7$ である。背景画像 f の着目画素 \mathbf{p} と各方向のペアとの明度差を比較し、その大小関係を2値化した値 $b_k(\mathbf{p})$ を以下のように定義する(図1参照)。

$$b_k(\mathbf{p}) = \begin{cases} 1 \dots \text{if } f(\mathbf{p} + \mathbf{d}_k r_k) \geq f(\mathbf{p}) \\ 0 \dots \text{otherwise} \end{cases} \quad (2)$$

2. 対象画像に対しても同じペアの画素でその大小関係 $b'_k(\mathbf{p})$ を調べる。

$$b'_k(\mathbf{p}) = \begin{cases} 1 \dots \text{if } g(\mathbf{p} + \mathbf{d}_k r_k) \geq f(\mathbf{p}) \\ 0 \dots \text{otherwise} \end{cases} \quad (3)$$

ここで、 g は背景画像である。

3. 両画像で求められた各画素におけるペアの大小関係の一致数 $E(\mathbf{p})$ を以下の式のように調べる。

$$E(\mathbf{p}) = \sum_{k=0}^7 \{ b_k(\mathbf{p}) \cdot b'_k(\mathbf{p}) + (1 - b_k(\mathbf{p})) \cdot (1 - b'_k(\mathbf{p})) \} \quad (4)$$

$E(\mathbf{p})$ と類似度を表す閾値 T_b により着目画素が背景であるか対象であるかを示す結果画像 $C(\mathbf{p})$ は次式で表される。

$$C(\mathbf{p}) = \begin{cases} \text{foreground} \dots E(\mathbf{p}) < T_b \\ \text{background} \dots \text{otherwise} \end{cases} \quad (5)$$

4. 明度差を表す閾値 T_c は2枚の背景画像をそれぞれ、 $f^1(p), f^2(p)$ とすると、以下の式で求めることができる。

$$\Delta f(p) = f^1(p) - f^2(p) \quad (6)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{k=1}^N (\Delta f(p) - \overline{\Delta f(p)})^2} \quad (7)$$

$$T_c = 2\sigma \quad (8)$$

ここで、 N は画素数、 $\overline{\Delta f(p)}$ は $\Delta f(p)$ の平均値である。類似度を表す閾値 T_b は一致数 E の確率密度関数 $p_E(E)$ を用いて以下の式で表される。

$$p_u = \binom{8}{u} q^u (1-q)^{8-u} \quad (9)$$

$$p_E(E) = \sum_{u=0}^8 p_u \delta(E-u) \quad (10)$$

$$\frac{p_1(T_B)}{p_2(T_B)} = \frac{c_2}{c_1} \quad (11)$$

ここで、 $\delta()$ はクロネッカーのデルタ、 q は一つのペアの符号が一致する頻度の期待値、 p_1 は背景同士の確率密度、 p_2 は背景と対象画像による確率密度、 c_1, c_2 は誤認識するコストである。コストに偏りがないとすると、 T_B は以下の式で表される。

$$p_1(T_B) = p_2(T_B) \quad (12)$$

2.2 Mixed Radial Reach Filter 法

2.1 で説明した RRF は任意の着目画素とその周辺画素との明度差の関係をj用いて背景であるかどうかを判断している。この考え方では、逆に周辺画素との明度差に乏しい環境つまり、テクスチャの少ない屋内や、壁などの背景下ではその効果が十分発揮できず、ノイズがのる事が考えられる。ここで、本論文では、テクスチャの少ない画素においては今まで用いられていたグレイスケールを用いた背景差分を用い、それ以外の画素においては RRF を用いるという、二つの手法を複合した MRRF 法を提案する。以下にその手法を説明する。

任意の着目画素 $p=(x, y)$ から、8方向のペアまでのリーチ長の合計画素数 $N_{RL}(p)$ は式(13)のように定義され

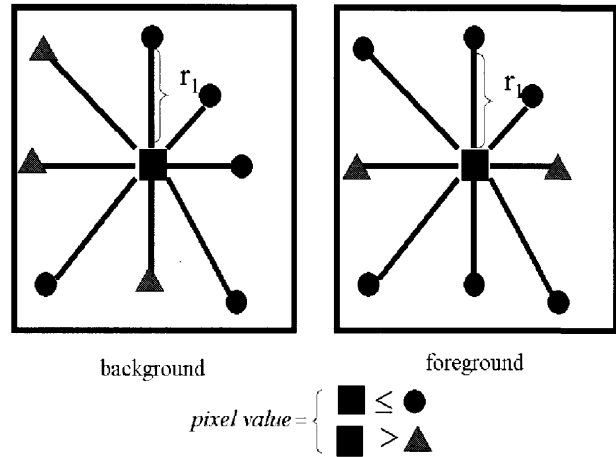


図1 Radial Reach Filter

る。

$$N_{RL}(p) = \sum_{k=0}^7 r_k(p) \quad (13)$$

2枚の背景画像から得られた $N_{RL}(p)$ の値をそれぞれ $N_{RL}^1(p)$ と $N_{RL}^2(p)$ とすれば、その差 $DN_{RL}(p)$ は次式で表される。

$$DN_{RL}(p) = N_{RL}^1(p) - N_{RL}^2(p) \quad (14)$$

理想的には $DN_{RL}(p)=0$ であるが、実際はカメラによるノイズの影響で差が生じる。ここで2枚の背景画像におけるリーチ長を $r^1(p)$ 、 $r^2(p)$ とするとその差 $\Delta r(p)$ 、分散 σ_{RL} は次式で定義される。

$$\Delta r(p) = |r^1(p) - r^2(p)| \quad (15)$$

$$\sigma_{RL} = \sqrt{\frac{1}{N} \sum_p (\Delta r(p) - \overline{\Delta r})^2} \quad (16)$$

ここで、 N は画素数であり、 $\overline{\Delta r}$ はリーチ長誤差の平均であり、 $\overline{\Delta r}=0$ とみなせる。

実際の環境においてはカメラによるノイズ以外にも照明の影響などをうけることになる。リーチ長の誤差 $\Delta r(p)$ はテクスチャの少ない画素ほどこの影響を受け、その値が大きくなる傾向にある。逆に、強いテクスチャが存在する画素では $\Delta r(p)$ の変化が少なく、照明が変化しても分散 σ_{RL} の範囲内に収まり続けることが考えられる。よって、対象画像 $g(p)$ のリーチ長の合計画素数を $N_{RL}g(p)$ とすると、背景画像 $N_{RL}^1(p)$ との誤差

ΔN_{RL} と分散 σ_{RL} を用いることにより、リーチ長誤差 $\Delta r(p)$ の変化の大小は次式によって推測される。

$$\Delta N_{RL}(p) = |N_{RL}^{-1}(p) - N_{RL}g(p)| \quad (17)$$

$$\begin{cases} \Delta N_{RL}(p) < \sigma_{RL} \dots \text{変化小} \\ \text{otherwise} \dots \text{変化大} \end{cases} \quad (18)$$

式(18)により、誤差の出やすい、即ちテクスチャの少ない画素を判別することができるが、この式には本当にテクスチャの大小関係が変わった画素も含まれている(図 2(c))。ここで、テクスチャが少ない画素であるか、本当にテクスチャが変化した画素であるかを判別するために、リーチ長の相対誤差を用いる。

2枚の背景画像を用いたリーチ長の相対誤差 N_{RR} は次式によって求められる。

$$N_{RR}(p) = \frac{|N_{RL}^{-1}(p) - N_{RL}^{-2}(p)|}{N_{RL}^{-1}(p)} \quad (19)$$

また、標準偏差 σ_{RR} は次式によって求められる。

$$r_{RR}(p) = \frac{|r^1(p) - r^2(p)|}{r^1(p)} = \frac{\Delta r(p)}{r^1(p)} \quad (20)$$

$$\sigma_{RR} = \sqrt{\frac{1}{N} \sum_p (r_{RR}(p) - \overline{r_{RR}})^2} \quad (21)$$

ここで、 $r_{RR}(p)$ は各画素のリーチ長の相対誤差、 $\overline{r_{RR}}$ はその平均であり、 $\overline{r_{RR}} = 0$ とみなせる。理想的には $N_{RR}(p) = 0$ であるが、 $\Delta N_{RL}(p)$ と同様にカメラによるノイズの影響を受ける。テクスチャがない画素においては、 $r^1(p)$ の値が大きいため、照明の変化による $\Delta r(p)$ の影響を受けにくいと考えられる。その為、 $r_{RR}(p)$ の値が σ_{RR} の値よりも大きい場合、テクスチャの大小の変化が生じたと考えられる(図 2(d))。

上記で述べた σ_{RL} と σ_{RR} により、背景画像と対象画像において、テクスチャの変化が生じたことを判別することができ(図 2(e))、MRRF は以下の式で画素ごとに RRF と BS を使い分ける。

$$\begin{cases} \{\Delta N_{RL}(p) < \sigma_{RL}\} \cup \\ \{\{\Delta N_{RL}(p) \geq \sigma_{RL}\} \cap \{N_{RR}(p) \geq \sigma_{RR}\}\} \Rightarrow RRF \\ \{\{\Delta N_{RL}(p) \geq \sigma_{RL}\} \cap \{N_{RR}(p) < \sigma_{RR}\}\} \Rightarrow BS \end{cases} \quad (22)$$

2.3 人物領域画像

MRRF 法では、RRF 法に比べ多くのノイズが低減されるが、多少のノイズは残る。認識においてはこのノイズによる影響をさらに低減させるために差分画像のエッジ画像を固有空間への投影に用いる。

3. 固有空間法

ラスタスキャンした画像を $\mathbf{x}_p = [x_1, x_2, x_3, \dots, x_N]^T$ と定義する。ここで $p(p=1, 2, \dots, P)$ はサンプル姿勢番号、 N は画素数であり、このベクトルは大きさ 1 に正規化されている。P 枚の画像を用いて行列 X を以下のように定義する。

$$\mathbf{X} \equiv \{\mathbf{x}_1 - \mathbf{c}, \mathbf{x}_2 - \mathbf{c}, \dots, \mathbf{x}_p - \mathbf{c}\} \quad (23)$$

ここで \mathbf{c} は全画像の平均ベクトルであり、次式で表される。

$$\mathbf{c} = \frac{1}{P} \sum_{p=1}^P \mathbf{x}_p \quad (24)$$

共分散行列 Q を以下のように定義する。

$$Q = \mathbf{X}\mathbf{X}^T \quad (25)$$

式(25)について、次式のような固有方程式を解く。

$$Q\mathbf{e} = \lambda\mathbf{e} \quad (26)$$

ここで、 $\mathbf{e}(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N)$ は固有ベクトル、 $\lambda(\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N)$ は固有値を表している。累積寄与率 K は次式で求められる。

$$K = \frac{\sum_{i=1}^k \lambda_i}{\sum_{j=1}^N \lambda_j} \quad (27)$$

累積寄与率 K は 1 に近づくほど画像を完全に固有空間に投影することを表す。しかし、すべての固有ベクトルを使用すると、冗長な情報まで使用することになるので、通常は k 個のベクトルを使用する。 k 個の固有値に対応する固有ベクトルを用いると、それぞれの画像ベクトル \mathbf{x}_p は以下の式で固有空間上の点 \mathbf{g}_p に投影される。

$$\mathbf{g}_p = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k)^T (\mathbf{x}_p - \mathbf{c}) \quad (28)$$

ある未知画像を固有空間に投影した点を \mathbf{z} とすれば、以下の式で計算された姿勢 p^* が未知画像の姿勢となる。

$$p^* = \min \| \mathbf{z} - \mathbf{g}_p \| . \quad (29)$$

4. 実験結果

4.1 人物領域抽出

MRRF 法の有効性を確かめるために、複雑背景下における人形を用いた実験を行った。図 2(a)に示すように背景には左半分にはテクスチャが多く含まれ、右半分にはテクスチャがほとんど含まれない画像を用いた。(b)のような対象画像が得られた場合、RRF 法による抽出結果は(f)、MRRF 法による結果は(g)のように得られた。また、それぞれの抽出結果において人形領域外で生じたノイズの pixel 数を比較した結果、表 1 のような結果を得た。これにより MRRF 法では従来法に比べ 65[%]の改善結果が得られ、テクスチャの少ない背景での抽出精度が明らかに向上したことが確かめられた。

4.2 姿勢認識

本提案手法の有効性を確かめるために屋外での姿勢認識実験を行った(図 3)。実験には 5 人に 4 種類の姿勢(直立、立ったまま物を拾う、座っている、座ったまま物を拾う)をとってもらい、それぞれ 3 方向からの画像を 3 枚ずつ、即ち 1 人につき 36 枚、合計 180 枚の画像を用いた。実験には leave-one-out 法を用い 4 人分の画像 144 枚をデータベースとし、残り 1 人の画像 36 枚に対して認識を行い、それを合計 5 組繰り返し認識率を平均した。

その結果、表 2 のような認識結果が得られた。いずれの方向の結果も 80~100%の認識結果を得た。誤認識の原因としては、カメラから対象までの距離によって生じる抽出領域の大小による誤認識を抑制するために抽出領域を 32×32 のサイズに正規化を行った際、「立ったまま拾う」と「座ったまま拾う」などの画像は正面方向以外の画像では非常に類似した画像となったことなどが挙げられる。しかし、これらは正規化のサイズを大きくするなどにより改善できると考えられる。

5. おわりに

本実験では MRRF 法を用いて、複雑背景下における人物領域の抽出と姿勢認識を行い、有効な結果を得た。本提案手法では認識のためにデータベースを作成する必要があり、データベースの情報量が大きくなるにつれ認識にかかる時間も長くなるが、我々はこれらを解決する手法[6]の研究を行っている。また、抽出に際しては固定背景画像を使用している為、背景の固定物体が動く、またはカメラのぶれなどにより背景が変化してしまう場合には本提案手法は適さないという問題があるが、我々はこのような背景の変化に対応した背景作成の手法[7]も研究している。これらの手法を併用することにより、以上に挙げた問題は解決されると考えられる。

今後は提案手法を用いて、動作の認識を行う予定である。本研究の一部は科学研究費(19700175)の助成を受けた。

参考文献

- [1] Satoh, Y., Kaneko, S., Niwa, Y. and Yamamoto, K.: Robust object detection using a radial reach filter, *IEICE Trans. Inform. Syst.*, Vol.J86-D-II, No.5, pp.616-624, 2003.
- [2] Tan, J. K., Houman, S. and Ishikawa, S.: Human motion representation using eigenspace, *Proc. of TENCON2005*, 4 pages in CD-ROM, Melbourne, 2005.
- [3] 島田竜也, 河口尚広, 加賀健太, 山田博三, 森晃徳: 屋内侵入者検知のための動的背景抽出法, *電子情報通信学会論文誌*, Vol.J88-D-II, no.10, pp.2054-2068, 2005.
- [4] Stauffer, C. and Grimson, W.E.L.: Adaptive background mixture models for real-time tracking, *Proc. of Computer Vision and Pattern Recognition.*, Vol.2, pp.246-252, 1999.
- [5] 飯島俊匠, 川嶋稔夫, 青木由直: 日常視点映像列の解析による動作抽出手法, *信学技報*, PRMU2001-50, pp.83-90, 2001.
- [6] タンジュークイ, 河野憲亮, 石川聖二, 金亨燮, 四宮孝史: モーションデータベースを用いた人の動作の高速認識, *画像電子学会誌*, Vol.36, no. 5, pp.738-746, 2007.
- [7] Sakai, Y., Tan, J. K. and Ishikawa, S.: Extracting a Human Area by Background Detection, *SICE-ICCAS2006*, 4pages in CD-ROM, pp.2296-2299, 2006.

MRRF 法を用いた複雑背景下における人物の姿勢認識

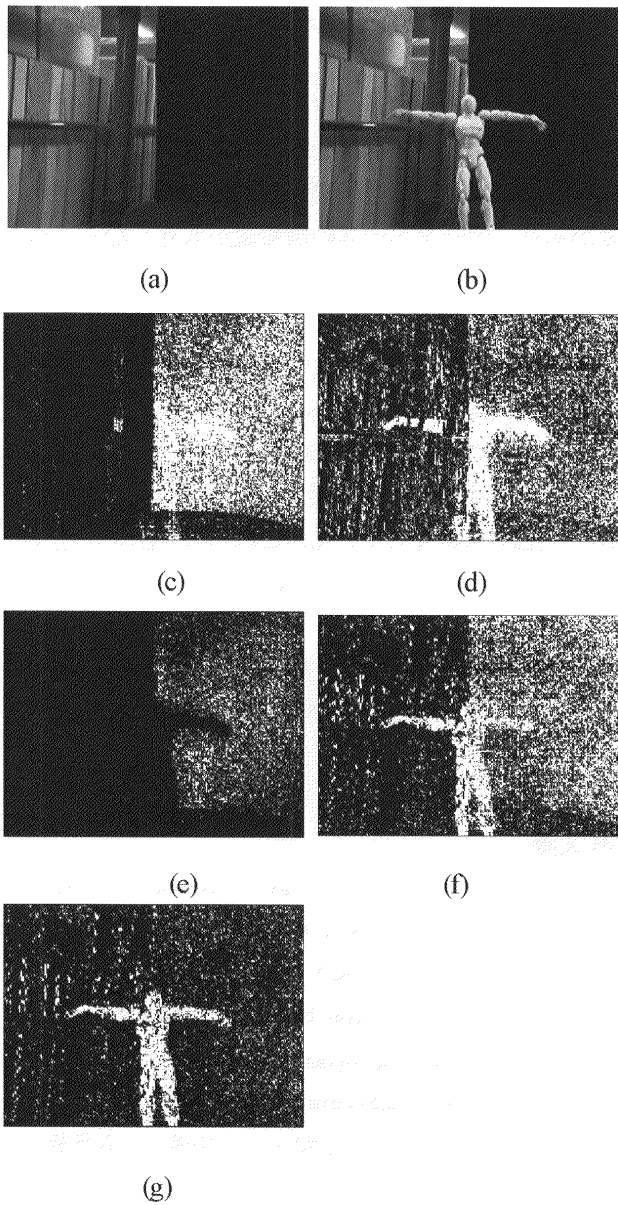


図2 人物領域抽出：(a) 背景画像, (b) 対象画像, (c) σ_{RL} により得られた画像, (d) σ_{RR} により得られた画像, (e) σ_{RL} and σ_{RR} により得られた画像, (f) RRF 画像 (g) MRRF 画像

表1 ノイズ pixel 数と改善率

RRF 画像	MRRF 画像	改善率(%)
18483[pixel]	6455[pixel]	65

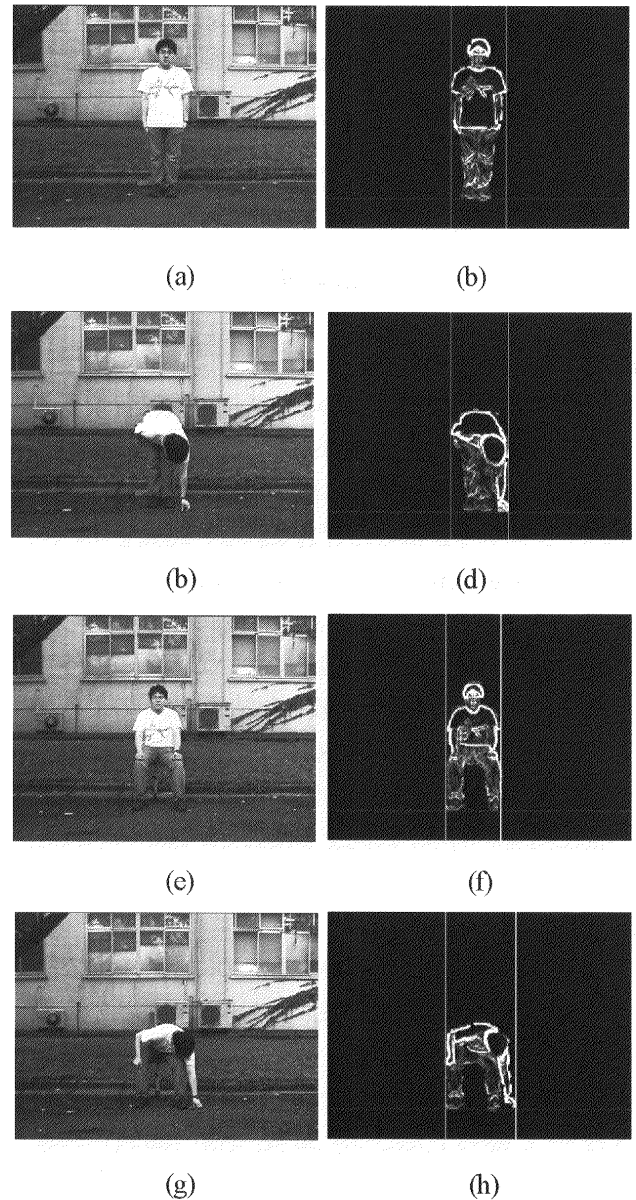
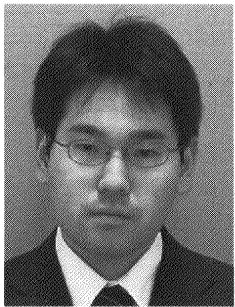


図3 屋外環境における姿勢認識: (a) 対象画像 (直立), (b) 抽出画像, (c) 対象画像 (直立しながら拾う), (d) 抽出画像, (e) 対象画像 (座る), (f) 抽出画像, (g) 対象画像 (座りながら拾う), (h) 抽出画像

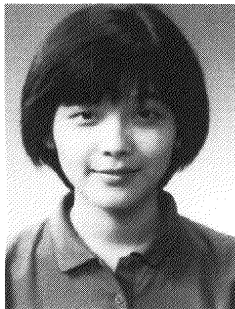
表2 認識率

方向	直立	立ったまま拾う	座る	座ったまま拾う
正面	100	100	100	100
側面	90	80	100	100
斜め	100	100	100	80
合計(%)	97	93	100	93



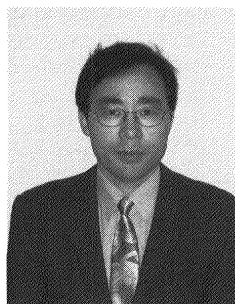
大場 統 (おおば おさむ)

九州工業大学工学部機械知能工学科卒。
現在、同大学大学院工学研究科博士前期
課程機械知能工学専攻在学中。画像処理
及び動作認識に興味を持つ。計測自動制
御学会学生会員。



タン ジュークイ

九州工業大学大学院設計生産工学専攻
博士後期課程修了。博士工学。同大SVBL
ポストドク研究員を経て、現在、同大工
学部機械知能工学科助教。平11、計測自
動制御学会九州支部奨励賞、平16、情報
処理学会九州支部第13回奨励賞、平17、
知能生命・ロボット国際シンポジウム若
手奨励賞受賞。主に画像計測・解析、動作認識、コンピュジョン
ジョンなどに関する研究に従事。IEEE、計測自動制御学会、情
報処理学会等会員。



金 亨燮 (きむ ひょんそふ)

1994年九州工業大学工学部卒。1996年
同大学大学院工学研究科博士前期課程
修了。1997年より九州工業大学工学部助
手(機械知能工学科)。2005年、同大学
工学部講師を経て、現在、同大工学部准
教授。画像計測、対称性解析、医用画像
処理、パターン認識に関する研究に従事。
博士(工学)。電子情報通信学会、日本医用画像工学会、バイオ
メディカルファジィシステム学会、医用画像情報学会、IEEE 各
会員。



石川 聖二 (いしかわせいじ)

東京大学工学部計数工学科卒。東京大学
大学院工学系研究科博士課程了。工学博
士。九州工業大学情報工学科講師を経て、
現在、九州工業大学工学部教授(機械知
能工学科)。この間、連合王国シェフィ
ールド大学客員研究員、オランダ王国ユ
トレヒト大学客員教授。3次元画像計測、
医学画像処理の研究に従事。計測自動制御学会、電子情報通信学
会、画像電子学会、IEEE 各会員。