

〔非公開〕

TR-C-0109

手振り画像データベースの

構築と認識手法の検証

山口 晃
Akira YAMAGUCHI

内海 章
Akira UTSUMI

宮里 勉
Tutomu MIYASATO

1 9 9 4 . 3 . 1 0

A T R 通信システム研究所

手振り画像データベースの構築と認識手法の検証

山口 晃 内海 章 宮里 勉

1995年2月24日

1 まえがき

臨場感通信会議システム [1] などの仮想空間操作システムではマンマシンインターフェースとして手振りが利用される例が多い。そのようなシステムにおいては、手振り入力デバイスとして多くの場合「データグローブ」のような装着型の装置が用いられている。しかしこのような装置は装着に手間がかかるなど使用者への負担が大きく、非装着型の装置の実現が望まれる。我々は複数のカメラをもちいた画像処理ベースの手振りの認識装置について研究を進めている。

我々の手法では、手のひらが平面形状であることを仮定することで、長袖の衣服を着用した場合のように手の平領域が腕の部分から分割されて得られる場合については、既に安定した認識動作が得られている [2][3]。我々はシステムの自由度を増すために手のひら領域と腕領域の分割を仮定しない手法について現在開発を進めている。

従来の認識手法では、最初に手の輝度情報から手の領域だけを切り出して2値化し、手の平の重心位置を検出していた。しかし領域が分割されていない場合には、この重心位置の検出が正しく行なえない。そこで、重心位置を示す特徴量として画像のスケルトンに着目し、得られた重心候補 (スケルトン) について、種々の特徴量から重心らしさを評価して重心位置を決定することとした [4]。その際、重心に加えて射影の極小点を手首候補とし、重心-手首の組を評価の対象としている。候補点群の中から4つの特徴量に基づいて、真の重心と手首を推定する。重心位置の決定ができれば、その後の処理については従来の手法がそのまま適用できる。

本報告では、重心手首推定の過程においてどの特徴量が正確な認識を行なうのにより必要かの検証を行う。

2 手のモデル

本システムでは、手の位置および姿勢を重心と手のひらの方向によって、また各指の曲げを重心と各指先点の位置関係によって表現する [5]。図1において、 O_h は手の重心を、 θ_{wh} は手のひらの方向を示す。これらは画像中の手のひら領域の重心、手のひら部分のエッジの主方向でそれぞれ近似する。複数カメラからの情報を組み合わせれば、手の位置と方向が算出できる。また、手のひらの特徴点は、同一平面内にあることを仮定し、指先点の移動は求めた平面上への投影点の動きとして表現する。

3 システム構成

システムの構成図は図2に示される通りとなっている。まず、入力画像中の肌色領域を、色情報および輝度情報を用いて領域分割する。ここでは、人間の手よりも大きな領域は得られないと仮定している。主軸検出部においては、Sobel フィルタの結果から平均エッジ方向を求め、手の方向とする。ここで得られた方向に基づき、指先が上方を向くよう画像を回転変換する。この

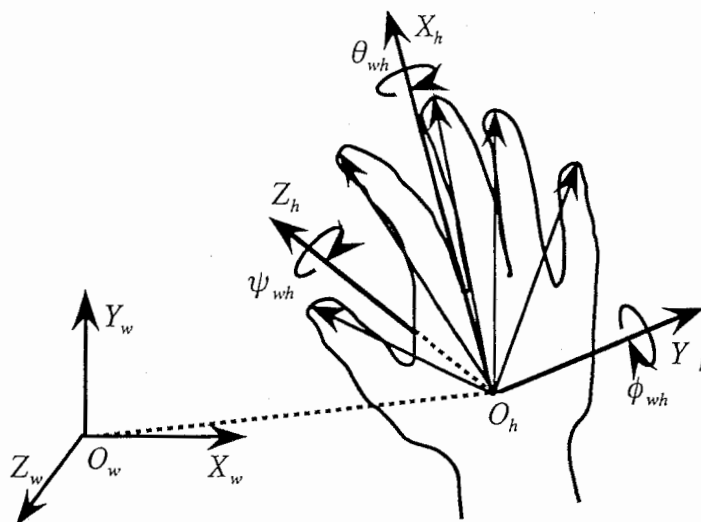


図 1: 手形状モデル

操作については、腕の部分が含まれると誤差が大きくなるため、手首位置確定後に手首より先の領域のみを対象として再度行なう。重心・手首処理部では、得られた領域画像とそれまでの重心点の軌跡から、現フレームにおける重心の3次元位置を決定する。これについては次節において詳しく述べる。

最終的な回転変換後の画像からラベリング処理により、指先候補点を抽出し、その3次元復元点について可能な組合せの中から、重心からの距離、3次元復元精度、画像上の指の並び方の3つの条件により選択し、決定する [3]。

4 スケルトン画像による重心探索手法

4.1 重心、手首候補点の推定

はじめに左右の画像を2値化して、手領域を取り出す。そして手領域の主軸を求め、主軸が垂直方向になるように手領域を回転させる。回転後、手領域の各点について一番近い境界線との距離を求める（距離変換）。変換値の極大点（スケルトン）を重心の候補点、また回転後の手領域の水平方向画素数が極小になる部分の垂直方向位置を手首候補点として抽出する（図3）。なお、計算量削減のため、手首候補点が重心より上にある組は候補から除いている。

手の平の重心、手首の候補点が求まると、その候補点の組合せに関して3次元復元を行ない、後述する4つの特徴量を求める。現段階ではこの4つの特徴量が独立であると仮定して、4つの特徴量の積が最大となる組をもって推定値としている。手首の位置が明らかになればその部分で切断して手のひら領域のみを対象とできるため、より正確に主軸検出を行なうことができる。

4.2 4つの特徴量

手の重心、手首の候補点の組合せについて以下の特徴量を求め、それぞれの特徴量が真の候補点に近くなるほど大きくなるように0~1の間で正規化し、正規化後の特徴量の積が最大になるものを真の重心と手首としている。

- 特徴量1

左右の画像について重心のスケルトンの大きさ(大きいほど大)と手首の細さ(細いほど大)を求め、それらの積を特徴量1とする。

- 特徴量2

重心と手首の距離が一定であることを利用し、事前情報(重心-手首距離)との差を特徴量2とする。

- 特徴量3

重心候補点を3次元復元する時の左右のカメラの視線の距離を復元精度と考え、特徴量3とする。

- 特徴量4

手首候補点についての特徴量3と同じ方法で求めた復元精度を特徴量4とする。

5 データベースの構築

特徴量の分布を求めるため、手の動きとそれを撮影した画像について、データベースを構築した。手の動きに関しては磁気センサを手の平の中心に固定させて、手の平の3次元の位置と姿勢を求めた。ここで手振りは、手の形状、手の始点とその時の姿勢、手の終点とその時の姿勢を無作為に選出したものである。

手の重心と手首の位置は、センサ座標系においてあらかじめ求めておいたセンサの計測値とのオフセットに基づいて、各時点で算出する。

6 特徴量の分布

より有効な特徴量の利用法を検討するため、重心と手首候補点の特徴量の分布を求めた。

6.1 分布図

データベースの画像を前章で述べた方法で処理を行ない、候補点を求めた。この時の推定真値の特徴量と候補点全体の特徴量分布をそれぞれ図4, 図5に示す。ここで推定真値とは、重心と手首の候補点の組合せ(候補点全体)の中から、前節で求めた真値に一番近い候補点の組合せを選んでいる。図4と図5を比較してみるとわかるように、推定真値と候補点全体とではかなり分布が異なっていることがわかる。

6.2 特徴量の相関係数

5つの手振りに関して、4つの特徴量の相関を求めた。これを表1に示す。

特徴量1と3、1と4についてはわずかだが相関がある。そこで特徴量1を左右(カメラ)の重心のスケルトン、左右の手首の細さに分解して求めてみた。これをそれぞれ特徴量5,6,7,8とする。そして、特徴量3,4と5,6,7,8間で相関を求めたものを表2に示す。

表2を見るとわかるように特徴量3と5,6に相関があり、特徴量4と7,8に相関があることがわかる。これからいえることは、重心のスケルトンが大きいときは重心の復元精度がよいということである。この事は手首についてもいえる。

手振りを認識するにあたって、同じ比率で特徴量を使用する場合、互いに相関がない特徴量を使用した方がより良い識別が得られると思われる。よって1または3,4の特徴量を認識する際にははずした方がよいと思われる。

表 1: 特徴量 1,2,3,4 の相関

特徴量		手振り 1	手振り 2	手振り 3	手振り 4	手振り 5
1	2	-0.0890	0.0935	-0.1225	0.0393	-0.3109
1	3	-0.1503	-0.6525	-0.1779	-0.6733	-0.6099
1	4	-0.5121	-0.4723	-0.1414	-0.4312	-0.4152
2	3	0.1851	-0.0954	0.1624	0.0760	0.1013
2	4	0.0690	-0.0562	-0.0225	-0.0164	0.4542
3	4	-0.2610	0.3189	0.2107	0.3244	0.3535

表 2: 特徴量 3,4 と 5,6,7,8 の相関

特徴量		手振り 1	手振り 2	手振り 3	手振り 4	手振り 5
3	5	-0.2129	-0.7094	-0.0255	-0.7073	-0.7241
4	5	0.1738	-0.2875	0.1984	-0.5489	-0.3780
3	6	-0.6791	-0.3373	-0.6913	-0.5461	-0.4819
4	6	0.0906	-0.0588	-0.1466	-0.0179	-0.0534
3	7	0.0711	-0.2238	0.1626	0.1261	0.0107
4	7	-0.3113	-0.7961	-0.5002	-0.3306	-0.4073
3	8	0.2616	-0.3163	-0.0124	0.0752	-0.1183
4	8	-0.6569	-0.8345	0.2925	-0.1589	-0.0989

7 認識手法の改善法

前節で得られたデータを用いれば、特徴量ベクトル α の全候補についての分布 $P(\alpha)$ と真の重心-手首の組についての分布 $P(\alpha|T)$ を近似的に求めることができる。これにより、重心位置推定に MAP 推定法を適用できると考える。

すなわち画像処理の最中に特徴量のパターンが得られた時、それが真 (T) である確率をベイズ公式 (1) により計算し、確率が一番高いものを真の重心と手首とすればよい。現在、データベースから求めた頻度分布をスプライン補間することで確率密度を近似する手法について検討を行っている。

$$P(T|\alpha) = \frac{P(T)P(\alpha|T)}{P(\alpha)} \quad (1)$$

ここで全候補中の真の候補である確率 $P(T)$ は定数と考える。

8 まとめ

- 手の挙動データを画像が対応した手振りに関するデータベースを構築した。
- 推定真値と候補点全体に関する分布を求めた。
- 特徴量 1 と 3,4 間には相関があるので、どちらかのみ使用することが望ましいと言える。

- 現状の候補点検出では、真値付近に候補点を得られない場合があります改善が必要である。

謝辞

この報告書を書くにあたって常に御指導してくださった内海章氏、裏隠居宏氏ならびに知能処理研究室の方々に多大の感謝の意を表します。

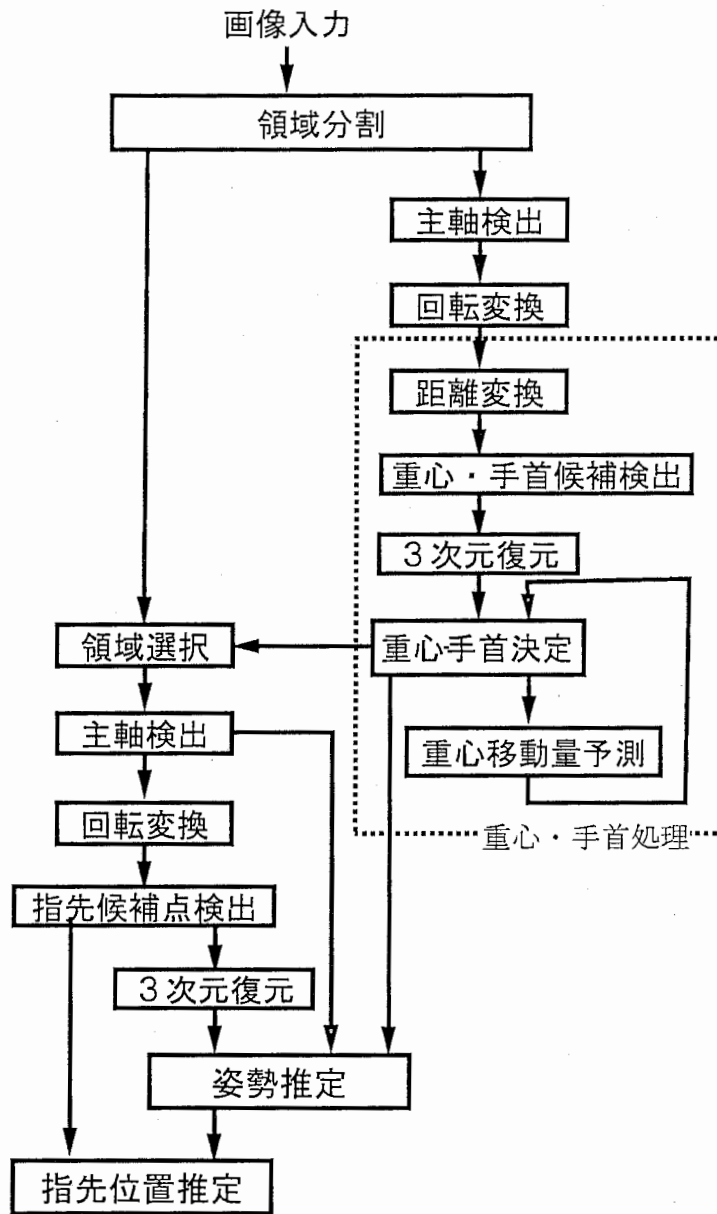


図 2: システム構成図

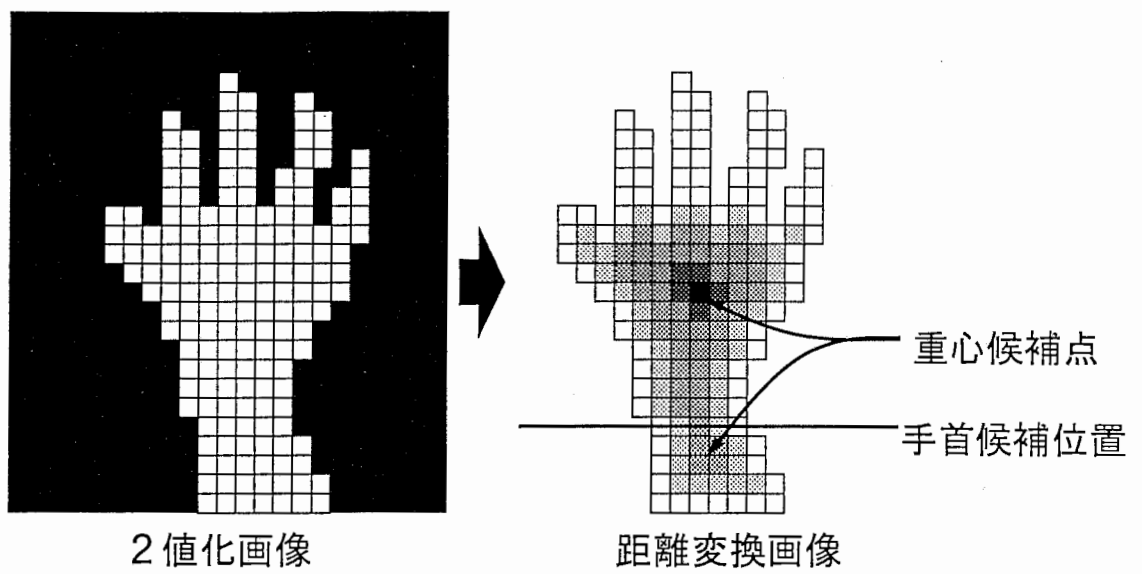
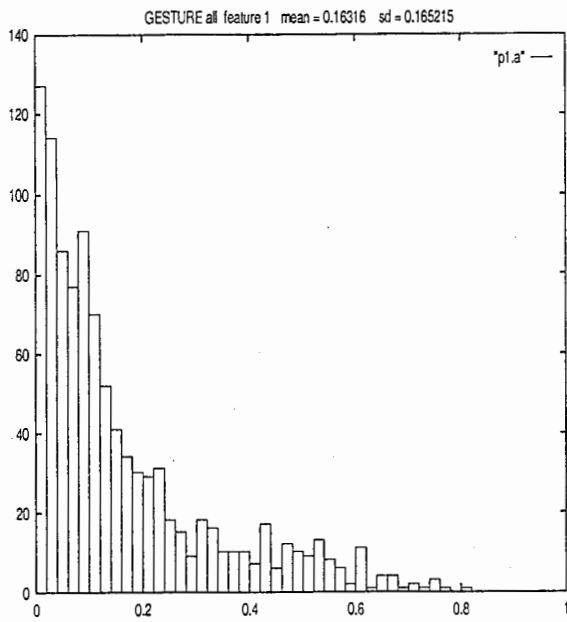
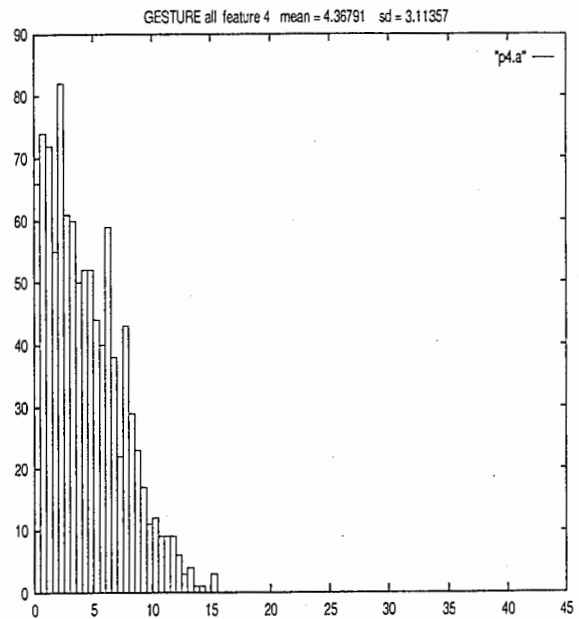


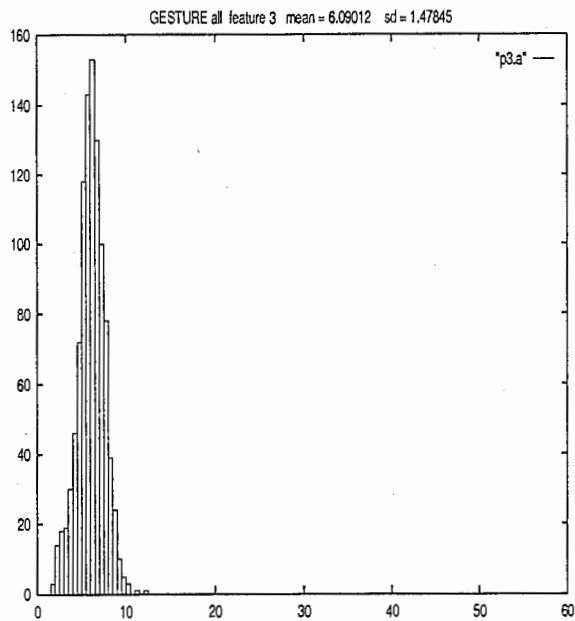
図 3: 手領域の 2 値化とスケルトン



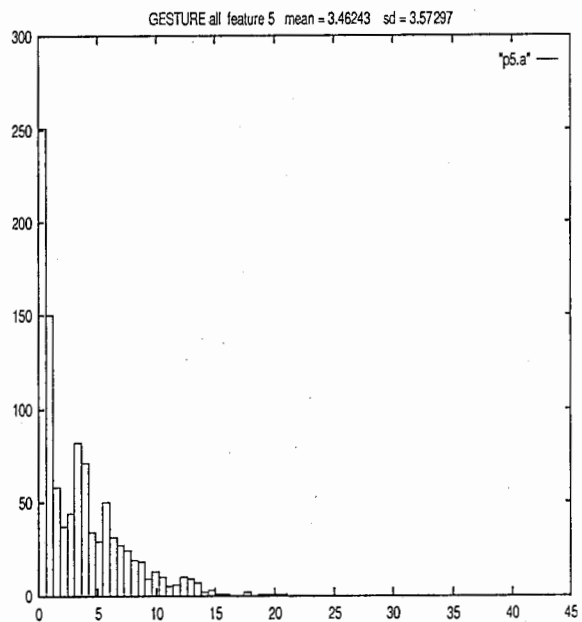
(a) 推定真値における特徴量 1 の分布



(c) 推定真値における特徴量 3 の分布

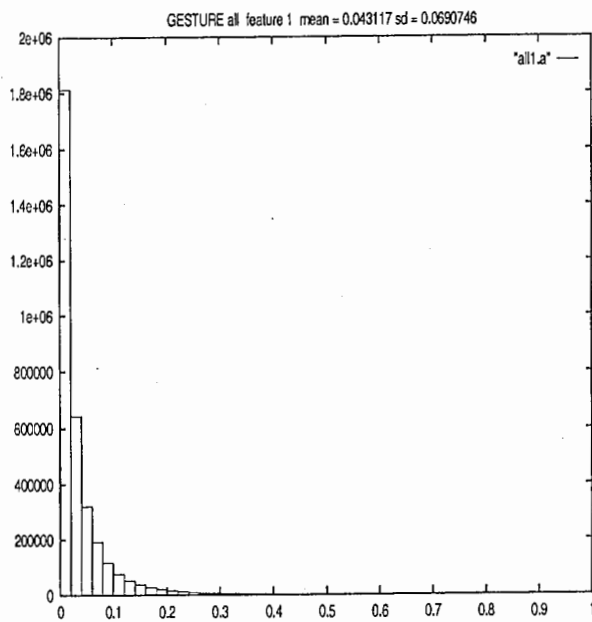


(b) 推定真値における特徴量 2 の分布

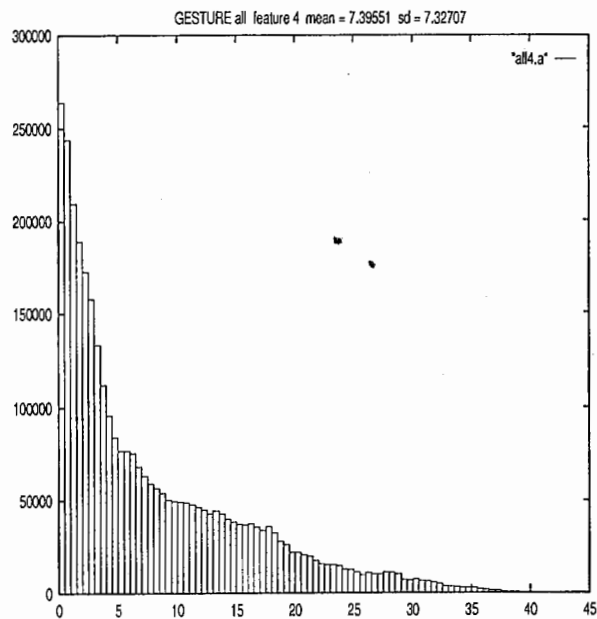


(d) 推定真値における特徴量 3 の分布

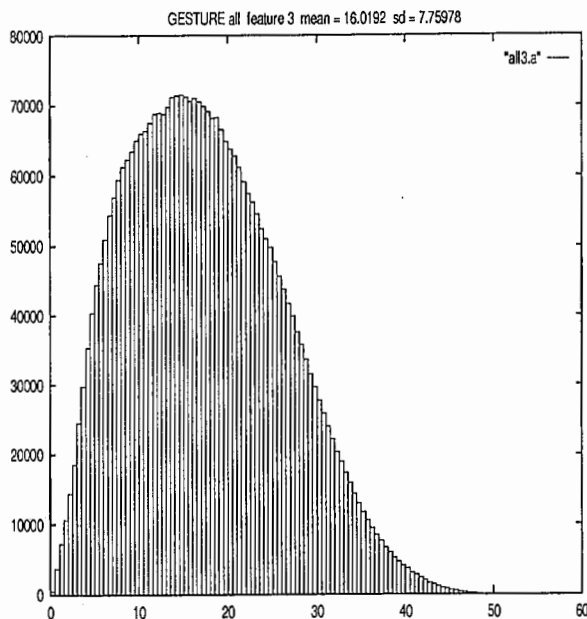
図 4: 推定真値の特徴量分布



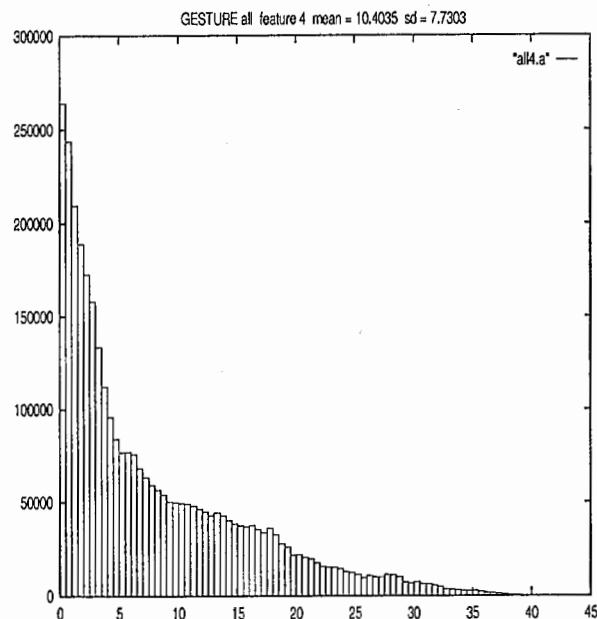
(a) 全候補点における特徴量1の分布



(c) 全候補点における特徴量3の分布



(b) 全候補点における特徴量2の分布



(d) 全候補点における特徴量4の分布

図 5: 全候補点の特徴量分布

参考文献

- [1] 岸野: “臨場感通信”, テレビジョン学会誌, Vol.46, No.6, pp.698-702 (1992).
- [2] 石淵ほか, “画像処理を用いた実時間手振り推定とヒューマンインターフェースへの応用”, 電気情報通信学会論文誌原稿”.
- [3] Koichi Ishibuchi, Haruo Takemura, Fumio Kishino, “Real time hand gesture recognition using 3D prediction model”, IEEE SMC, 1993.
- [4] 内海 章, “着衣によらない実時間手振り認識”, 特許出願資料, 1995.
- [5] 大西, 竹村, 岸野, “手振り認識のための手形状記述について”, テレビジョン学会技術報告 Vol.15, No.25, pp.25-29 (1991).