

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4060261号  
(P4060261)

(45) 発行日 平成20年3月12日(2008.3.12)

(24) 登録日 平成19年12月28日(2007.12.28)

(51) Int. Cl.		F I			
<b>G06T</b>	<b>7/00</b>	<b>(2006.01)</b>	G06T	7/00	300F
<b>G06F</b>	<b>3/033</b>	<b>(2006.01)</b>	G06F	3/033	310Y
<b>G06N</b>	<b>3/08</b>	<b>(2006.01)</b>	G06N	3/08	Q
<b>G06T</b>	<b>1/00</b>	<b>(2006.01)</b>	G06T	1/00	340Z

請求項の数 8 (全 18 頁)

(21) 出願番号	特願2003-357082 (P2003-357082)	(73) 特許権者	503360115
(22) 出願日	平成15年10月16日(2003.10.16)		独立行政法人科学技術振興機構
(65) 公開番号	特開2005-122492 (P2005-122492A)		埼玉県川口市本町4丁目1番8号
(43) 公開日	平成17年5月12日(2005.5.12)	(74) 代理人	100091443
審査請求日	平成17年3月22日(2005.3.22)		弁理士 西浦 ▲嗣▼晴
		(72) 発明者	星野 聖
			茨城県つくば市竹園3-102-103
		(72) 発明者	小渡 悟
			沖縄県浦添市屋富祖3-24-3
		審査官	松尾 俊介

最終頁に続く

(54) 【発明の名称】 画像に含まれる身体形状を判定する方法及び非接触型ポインティング・デバイスの実現方法

(57) 【特許請求の範囲】

【請求項1】

予め所定の複数種類の身体動作についての学習用の動画像の各フレームについて対数極座標変換を行って対数極座標画像を取得し、前記各フレームについての前記対数極座標画像を複数の画素からなるセルに切り分けて各セルについての高次局所自己相関特徴を抽出し、抽出した複数の高次局所自己相関特徴に基づいて前記各フレームに示された身体形状の高次局所自己相関パターンを抽出し、前記高次局所自己相関パターンに基づいて前記動画像に含まれる各フレームに示された前記身体形状を判定するための判定基準を定める学習ステップと、

実際の身体動作についての動画像の各フレームについて対数極座標変換を行って実際の対数極座標画像を取得し、前記各フレームについての前記対数極座標画像を複数の画素からなるセルに切り分けて各セルについての実際の高次局所自己相関特徴を抽出し、抽出した複数の実際の高次局所自己相関特徴に基づいて前記各フレームに示された身体形状の実際の高次局所自己相関パターンを抽出し、前記実際の高次局所自己相関パターンと前記判定基準を用いて前記フレームに示された前記身体形状を判定する判定ステップとをコンピュータにより実施し、

前記学習ステップの前記高次局所自己相関特徴及び前記判定ステップの前記実際の高次局所自己相関特徴を抽出するための特徴抽出ステップでは、前記セルに含まれる点画像または線画像成分を強調するための複数種類の局所パターンを予め決めておき、前記セルの画素に含まれる情報と前記複数種類の局所パターンに示された前記情報との演算条件とに

基づいて、1つの前記セルの情報について前記局所パターンの数に対応した複数の演算値を求め、該複数の演算値により前記高次局所自己相関特徴を構成し、

前記高次局所自己相関パターンを抽出する相関パターン抽出ステップでは、前記フレームから切り出した複数の前記セルについて求めた前記複数の演算値を前記局所パターンに対応して加算して得た複数の加算値により前記高次局所自己相関パターンを抽出することによりコンピュータ上で画像に含まれた身体形状を判定する方法において、

前記特徴抽出ステップでは、前記セルの画素に含まれる情報として濃度を用い、前記複数種類の局所パターンの前記演算条件として前記セルに含まれる画素間の濃度差を強調する条件を用いることを特徴とする画像に含まれた身体形状を判定する方法。

【請求項2】

前記複数種類の局所パターンの前記演算条件には、前記1つの画素の前記濃度を示す値を累乗する条件が含まれている請求項1に記載の画像に含まれた身体形状を判定する方法。

【請求項3】

前記セルを  $n \times n$  個 ( $n$  は2以上の整数) の画素の集合から構成し、

前記フレームから前記セルを切り分ける際に、前に切り分けたセルと後から切り分けるセルとが一部重なるように切り分けることを特徴とする請求項1に記載の画像に含まれた身体形状を判定する方法。

【請求項4】

前記判定ステップでは、ニューラルネットを用いて前記高次局所相関パターンを自己組織化することにより前記身体形状を判定することを特徴とする請求項1または2に記載の画像に含まれる身体形状を判定する方法。

【請求項5】

請求項1, 2, 3または4に記載の画像に含まれる身体形状を判定する方法をコンピュータで実施して、人間の手形形状を判定し、コンピュータ上で前記手形形状をポインティングとすることを特徴とする非接触型ポインティング・デバイスの実現方法。

【請求項6】

予め所定の複数種類の身体動作についての学習用の動画像の各フレームについて対数極座標変換を行って対数極座標画像を取得し、前記各フレームについての前記対数極座標画像を複数の画素からなるセルに切り分けて各セルについての高次局所自己相関特徴を抽出し、抽出した複数の高次局所自己相関特徴に基づいて前記各フレームに示された身体形状の高次局所自己相関パターンを抽出し、前記高次局所自己相関パターンに基づいて前記動画像に含まれる前記各フレームに示された前記身体形状を判定するための判定基準を定める学習ステップと、

実際の身体動作についての動画像の各フレームについて対数極座標変換を行って実際の対数極座標画像を取得し、前記各フレームについての前記対数極座標画像を複数の画素からなるセルに切り分けて各セルについての実際の高次局所自己相関特徴を抽出し、抽出した複数の実際の高次局所自己相関特徴に基づいて前記各フレームに示された身体形状の実際の高次局所自己相関パターンを抽出し、前記実際の高次局所自己相関パターンと前記判定基準を用いて前記フレームに示された前記身体形状を判定する判定ステップと、

前記対数極座標画像に基づいて前記身体動作のジェスチャの始点と終点とを認識するジェスチャ認識ステップとをコンピュータにより実施し、

前記学習ステップの前記高次局所自己相関特徴及び前記判定ステップの前記実際の高次局所自己相関特徴を抽出するための特徴抽出ステップでは、前記セルに含まれる点画像または線画像成分を強調するための複数種類の局所パターンを予め定めておき、前記セルの画素に含まれる情報と前記複数種類の局所パターンに示された前記情報との演算条件とに基づいて1つの前記セルの情報について前記局所パターンの数に対応した複数の演算値を求め、該複数の演算値により前記高次局所自己相関特徴を構成し、

前記高次局所自己相関パターンを抽出する相関パターン抽出ステップでは、前記フレームから切り出した複数の前記セルについて求めた前記複数の演算値を前記局所パターンに

10

20

30

40

50

対応して加算して得た複数の加算値により前記高次局所自己相関パターンを抽出し、

前記判定ステップの判定結果と前記ジェスチャ認識ステップの認識結果とに基づいて前記身体動作により示されたポインティングをコンピュータ上で判別する非接触型ポインティング・デバイスの実現方法であって、

前記特徴抽出ステップでは、前記セルの画素に含まれる情報として濃度を用い、前記複数種類の局所パターンの前記演算条件として前記セルに含まれる画素間の濃度差を強調する条件を用いることを特徴とする非接触型ポインティング・デバイスの実現方法。

【請求項 7】

前記複数種類の局所パターンの前記演算条件には、前記 1 つの画素の前記濃度を示す値をべき乗する条件が含まれている請求項 6 に記載の非接触型ポインティング・デバイスの実現方法。

10

【請求項 8】

前記判定ステップでは、ニューラルネットを用いて前記高次局所相関パターンを自己組織化することにより前記身体形状を判定することを特徴とする請求項 7 に記載の非接触型ポインティング・デバイスの実現方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、画像に含まれる身体形状を判定する方法及びこの方法を用いて非接触型ポインティング・デバイスを実現する方法に関するものである。

20

【背景技術】

【0002】

マウス、トラックボール、タブレットなどのポインティング・デバイスは、Graphical User Interface (以下、GUI) を構成する上で不可欠な要素であり、その中でも現在は操作が直感的で分かりやすいということでマウスが広く利用されている。しかし、人間から計算機への意思伝達を自然に行なえるマン・マシンインターフェースという観点からみると、マウスをはじめとする一般的なポインティング・デバイスは、操作を行なうにはユーザがデバイスに直接触れる必要がある。一方、コンピュータの利用分野は、仮想環境、モバイル環境、ユビキタス環境などと実生活全般に大きく拡大しつつあり、また、計算機の急速な普及に伴いユーザがより良い操作性や自然な操作感覚をえられるインターフェースが求められている。これらの環境で用いられるポインティング・デバイスは、ユーザを空間内に固定するものであってはならない。また、モバイル環境においては、計算機の小型化により機器の表面積が減少するため、従来のユーザの指先やスタイラスペンなどが機器に接することによるポインティング・デバイスを実装することは困難になってきている。一般にポインティング・デバイスは、ポインティング機構とスイッチ機構から成り立つことから、これらの機構をユーザの身体動作から推定できれば、上記の制限から解放されるものと思われる。

30

【0003】

ユーザの身体動作からポインティング・デバイスを実現する手法としては、腕の筋収縮信号を用いたEMG制御型デバイス(辻敏夫、福田修、村上満、金子真、“ニューラルネットを利用したEMG制御型ポインティングデバイス、”計測自動制御学会論文集、vol. 37、no. 5、pp. 425 - 431、2001.) [非特許文献1]、両足で板を傾斜させて操作する両足操作型デバイス(久米祐一郎、井上啓、“両足操作型ポインティングデバイスの検討、”映像情報メディア学会誌、vol. 54、no. 6、pp. 871 - 874、2000.) [非特許文献2]、3系統のセンサを組み合わせた「Ubi-Finger」(塚田浩二、安村通晃、“Ubi-Finger: モバイル指向ジェスチャ入力デバイスの研究、”情報処理学会論文誌、vol. 43、no. 12、pp. 3675 - 3684、2002.) [非特許文献3]などが提案されている。その他にも身振り手ぶりなどのジェスチャ識別までを含めると、データグローブ、磁気式モーションキャプチャ、加速度センサ、関節角度検出センサといったセンサを体に装着することで身体

40

50

動作を計測することでジェスチャ識別を行なう各種の手法が提案されている（澤田秀之、橋本周司、“加速度センサを用いたジェスチャ認識と音楽制御への応用、”信学論A、vol.79-A、no.2、pp.452-459、1996.）[非特許文献4]、（宮尾淳一、“手話学習システムのための手話単語特徴に基づく教授法”、信学論D-I、vol.J83-D-I、no.10、pp.1120-1128、2000.）[非特許文献5]。これらはセンサを直接体に装着することと専用の機材を用いることで処理の高速性、安定性、計測精度を実現できるが、特殊な機材を装着することになるためユーザに対して拘束感や不自然な使用感を与えてしまい、快適とは言いがたい。これに対して非接触型の手法としては、超音波の位相差を用いた位置計測手法（野中秀俊、伊達惇、“超音波の位相差を利用したポインティング装置（SPD）の開発、”計測自動制御学会論文集、vol.29、no.7、pp.735-744、1993.）[非特許文献6]、複数のカメラを用いる手法（渡辺博己、本郷仁志、安本護、山本和彦、“マルチカメラを用いた全方位ポインティングジェスチャの方向推定、”電学論C、vol.121、no.9、pp.1388-1394、2001.）[非特許文献7]などがある。しかしこれらは特殊な機材を使用することになるため、一般のユーザが利用するのは難しい。通常のカメラを用いて追跡領域の低次のモーメント特徴を用いる手法（高松亮、佐藤誠、“最適視点視野による掌の追跡と手指形状認識に基づくポインティングデバイスの提案、”ヒューマンインターフェース学会、vol.1、no.1、pp.45-52、1999.）[非特許文献8]も提案されているが、低次のモーメントは対象の大きさや各方向の広がりなどの情報程度しか持たないため、対象が回転すると広がり方向が変わり誤認識を起こす、細かな形状が扱えないなどの欠点を有する。そのため、一般のユーザが快適に利用することを考慮すると、このときのポインティング・デバイスは、非接触でリアルタイム処理が可能で、取り付け箇所や空間を自由に設定できること、また、十分に小型で軽量で安価であることが必要である。

#### 【0004】

このような事情から、発明者等は非接触型デバイスである単眼カメラを用い、画像からユーザの手の位置と形状を推定することで、ユーザに対して拘束感や不自然な使用感を与えないポインティング・デバイスの構築を検討した。本システムでは、取り込み画像を直交座標系から対数極座標系に変換することで、画像のデータ量を減らし計算コストを削減することで、通常のカメラ以外の専用ハードウェアを用いずにリアルタイム性を実現し、対数極座標空間の高次局所自己相関特徴を用いることで背景の変化や手の回転に対する頑健性を実現する方法を先に提案した（小渡悟、星野聖、“単眼動画像からの手の位置と形状の推定に基づくポインティングデバイスの提案、”電子情報通信学会技術研究報告、HIP2002-76、pp.79-84、2003.）[非特許文献9]。

【非特許文献1】辻敏夫、福田修、村上満、金子真、“ニューラルネットを利用したEMG制御型ポインティングデバイス、”計測自動制御学会論文集、vol.37、no.5、pp.425-431、2001.

【非特許文献2】久米祐一郎、井上啓、“両足操作型ポインティングデバイスの検討、”映像情報メディア学会誌、vol.54、no.6、pp.871-874、2000.

【非特許文献3】塚田浩二、安村通晃、“Ubi-Finger：モバイル指向ジェスチャ入力デバイスの研究、”情報処理学会論文誌、vol.43、no.12、pp.3675-3684、2002.

【非特許文献4】澤田秀之、橋本周司、“加速度センサを用いたジェスチャ認識と音楽制御への応用、”信学論A、vol.79-A、no.2、pp.452-459、1996.

【非特許文献5】宮尾淳一、“手話学習システムのための手話単語特徴に基づく教授法”、信学論D-I、vol.J83-D-I、no.10、pp.1120-1128、2000.

【非特許文献6】野中秀俊、伊達惇、“超音波の位相差を利用したポインティング装置（SPD）の開発、”計測自動制御学会論文集、vol.29、no.7、pp.735-

10

20

30

40

50

744、1993.

【非特許文献7】渡辺博己、本郷仁志、安本護、山本和彦、“マルチカメラを用いた全方位ポインティングジェスチャの方向推定、”電学論C、vol.121、no.9、pp.1388-1394、2001.

【非特許文献8】高松亮、佐藤誠、“最適視点視野による掌の追跡と手指形状認識に基づくポインティングデバイスの提案、”ヒューマンインターフェース学会、vol.1、no.1、pp.45-52、1999.

【非特許文献9】小渡悟、星野聖、“単眼動画像からの手の位置と形状の推定に基づくポインティングデバイスの提案、”電子情報通信学会技術研究報告、HIP2002-76、pp.79-84、2003.

【発明の開示】

【発明が解決しようとする課題】

【0005】

しかしながら非特許文献9で提案した方法では、高次局所自己相関特徴を抽出するステップでは、画像から切り分けるセルに含まれる画素の濃度を予め定めた基準と比較して、その基準より大きいかが否かにより“0”か“1”の2値で表現していたため、設計者が適切な閾値を設定する手間が必要であった。また抽出のために使用する局所パターンとして、点画像または線画像成分を強調するための複数種類の局所パターンを予め定めているが、従来の局所パターンでは点画像または線画像成分を十分に強調することができなかった。そのため従来の方法では、身体形状の判定精度が必ずしも高くなく、また識別の不安定

【0006】

本発明の目的は、従来よりも身体形状の判定精度が高い画像に含まれた身体形状を判定する方法及び該方法を用いた非接触型ポインティング・デバイスの実現方法を提供することにある。

【0007】

本発明の他の目的は、従来の方法よりも身体形状の判定精度が高く且つ識別の安定性が高い、画像に含まれた身体形状を判定する方法及び該方法を用いた非接触型ポインティング・デバイスの実現方法を提供することにある。

【課題を解決するための手段】

【0008】

本発明の画像に含まれた身体形状を判定する方法では、学習ステップと判定ステップとをコンピュータにより実施する。学習ステップでは、まず予め所定の複数種類の身体動作についての学習用の動画像の各フレームについて対数極座標変換を行って対数極座標画像を取得する。そして各フレームについての対数極座標画像を複数の画素からなるセルに切り分けて各セルについての高次局所自己相関特徴を抽出し、抽出した複数の高次局所自己相関特徴に基づいて各フレームに示された身体形状の高次局所自己相関パターンを抽出し、高次局所自己相関パターンに基づいて動画像に含まれる各フレームに示された身体形状を判定するための判定基準を定める。また判定ステップでは、実際の身体動作についての動画像の各フレームについて対数極座標変換を行って実際の対数極座標画像を取得する。そして学習ステップと同様に、各フレームについての対数極座標画像を複数の画素からなるセルに切り分けて各セルについての実際の高次局所自己相関特徴を抽出し、抽出した複数の実際の高次局所自己相関特徴に基づいて各フレームに示された身体形状の実際の高次局所自己相関パターンを抽出し、実際の高次局所自己相関パターンと判定基準を用いてフレームに示された身体形状を判定する。

【0009】

学習ステップの高次局所自己相関特徴及び判定ステップの実際の高次局所自己相関特徴を抽出するための特徴抽出ステップでは、まずセルに含まれる点画像または線画像成分を強調するための複数種類の局所パターンを予め定めておき、セルの画素に含まれる情報と複数種類の局所パターンに示された情報との演算条件とに基づいて1つのセルの情報につ

10

20

30

40

50

いて局所パターンの数に対応した複数の演算値を求め、該複数の演算値により高次局所自己相関特徴を構成する。そして高次局所自己相関パターンを抽出する相関パターン抽出ステップでは、フレームから切り出した複数のセルについて求めた複数の演算値を局所パターンに対応して加算して得た複数の加算値により高次局所自己相関パターンを抽出する。

特に、本発明においては、特徴抽出ステップにおいて、セルの画素に含まれる情報として濃度を用いる。そして複数種類の局所パターンの演算条件としてセルに含まれる画素間の濃度差を強調する条件を用いる。従来は、セルに含まれる画素の濃度を予め定めた基準と比較して、その基準より大きいか否かにより“0”か“1”の2値で表現していたのに対して、本発明ではセルの画素に含まれる情報として濃度をそのまま用いるので、情報は2値ではなく多値となる。その上で、これら多値と積算される複数種類の局所パターンの演算条件としてセルに含まれる画素間の濃度差を強調する条件を用いると、多値の濃度の差はより大きなものとなり、身体形状の判定精度を高めることができる。

#### 【0010】

ここで複数種類の局所パターンの演算条件には、1つの画素の濃度を示す値を累乗する条件を含めるのが好ましい。このように累乗する条件を含めると、濃度差はより顕著なものとなり、確実に判定精度を高めることができる。また局所パターンの種類を増やすことができるので、その分判定の精度と安定性を増大させることができる。

#### 【0011】

具体的には、セルを $n \times n$ 個( $n$ は2以上の整数)の画素の集合から構成し、フレームからセルを切り分ける際に、前に切り分けたセルと後から切り分けるセルとが一部重なるように切り分けるのが好ましい。このようにすると処理すべきデータ量は多くなるものの、フレーム内をより細かく精査することになるので、判定精度を更に高めることができる。

#### 【0012】

また判定ステップでの判定手法は任意である。しかし判定ステップでは、ニューラルネットを用いて高次局所相関パターンを自己組織化することにより身体形状を判定すると、従来よりも判定精度及び安定性を共に高めることができる。

#### 【0013】

上記の本発明の画像に含まれる身体形状を判定する方法をコンピュータ上で実施して、人間の手形形状を判定し、手形形状をポインティングとすると、手形を利用した非接触型ポインティング・デバイスをより高い精度で実現できる。

#### 【発明の効果】

#### 【0014】

本発明によれば、従来の方法に比べて、身体形状の判定の精度と安定性を高めることができる。

#### 【発明を実施するための最良の形態】

#### 【0015】

図1は、本発明の方法を用いて手の形状及び動作(身体形状)により非接触型ポインティング・デバイスを実現する場合のステップをブロック図形式で示したものであり、図2は、本発明の画像に含まれる身体形状を判定する方法をコンピュータを用いて実現する場合に用いるソフトウェアのアルゴリズムを示すフローチャートである。

#### 【0016】

図1のステップS1は学習ステップであり、S2は判定ステップであり、S3はジェスチャ認識ステップである。学習ステップS1においては、固定単眼カメラ(CCDカメラ)で撮影された時系列画像(学習用の動画)の各フレームの画像を入力として、ステップS102で対数極座標変換を行う。この対数極座標変換ステップS102は、フレームの画像を網膜と大脳視覚領との写像関係を近似した対数極座標画像に変換するものである。Schwartzはこの網膜-皮質間の関係は、対数極座標変換(Log Polar Mapping; 以下、LPMと略す)によって近似できる(E. L. Schwartz, "Computational anatomy and functional a

10

20

30

40

50

rchitecture of striate cortex: a spatial mapping approach to perceptual coding, "Vision Research, vol. 20, no. 8, pp. 645 - 668, 1980.)としており、本願明細書においても、このLPMを用いて対数極座標画像を生成する。

【0017】

対数極座標変換画像の生成は以下のようにして行う。すなわち直交画像での座標  $I(x, y)$  を複素平面  $Z$  とし、そのときの複素平面上の点を  $z = x + iy$  と表現する。つぎにLPMの座標  $L(p, q)$  も同様に複素平面  $W$  とし、そのときの複素平面上の点を  $w = p + iq$  と表現する。このとき、LPMは式(1)となる。

10

【0018】

$$w = \log(z + \dots) \quad \dots (1)$$

ここで、 $\dots$  は原点が特異点となるのをさけるための並進量である。LPMにより、原画像  $I$  は角度方向、半径方向の各成分に分解される。図3からも分かるように、半径方向のサンプリングは対数的なものとなるため、周辺部の情報は中心部に比べると大まかに捉えられている。そのため、中心部で高解像度を保ちつつ、周辺部にいくにつれ解像度を下げていくため、空間の状況を大まかに捉えデータ量を大幅に削減し、また処理時間を減少させることができる。LPMの例として、図4(a)の画像を図4(b)に示す点においてサンプリングを行なった結果を図4(c)に示す。またその逆変換を図4(d)に示す。入力画像からLPMを行う際、各画素がLPM画像のどの画素に対応しているかを計算する必要がある。本実施の形態では、入力画像の画像サイズは一定のため、変換処理の扱いを容易にするために、あらかじめ前処理の段階で各画素の対応点を求め、ルックアップテーブルを作成する。このようにすることで変換処理の扱いが容易になり、計算時間の短縮になる。

20

【0019】

本実施の形態では、上記のようにして求めた各フレームについての対数極座標画像を、ステップS103で複数の画素からなるセルに切り分けて各セルについての高次局所自己相関特徴を抽出する。概念的に説明すれば、図5に示すようにフレームの対数極座標画像1から  $n \times n$  (本実施の形態では  $3 \times 3$ ) の画素を切り出す。例えば、図5の例では、第1のセルを構成する画素として画素(1, 2, 3, 101, 102, 103, 201, 202, 203)を切り出した後、第2のセルを構成する画素として画素(2, 3, 4, 102, 103, 104, 202, 203, 204)を切り出す。このように各セルを切り出す際に、一部の画素が重なるようにすると、形状の検出精度が高くなる。本実施の形態では、図6に概念的に示すように、各セルの画素に含まれる情報として濃度を用いる。図6の各画素内の数字は濃度の程度を示している。そして学習ステップS1の高次局所自己相関特徴を抽出するための特徴抽出ステップS104では、まずセルに含まれる点画像または線画像成分を強調するための複数種類の局所パターンを予め決めておく。図7は本実施の形態で用いる複数種類の局所パターンとして35種類の局所パターンM1~M35の例を示している。図7の局所パターンでは、35種類の局所パターンの演算条件として、セルに含まれる画素間の濃度差を強調する条件を用いている。例えば、局所パターンM1の中心の画素に対応する部分の数字「1」は、セルの対応画素の濃度を1累乗することを意味し、局所パターンM2の中心の画素に対応する部分の数字「11」は、セルの対応画素の濃度を二乗することを意味し、局所パターンM2の中心の画素に対応する部分の数字「11」は、セルの対応画素の濃度を二乗することを意味し、局所パターンM7の中心の画素に対応する部分の数字「111」は、セルの対応画素の濃度を三乗することを意味する。

30

40

【0020】

高次局所自己相関特徴を得るためには、各セルの画素に含まれる情報(図6に示すような数値情報として表示された濃度)と35種類の局所パターンM1~M35に示された情報との演算条件(図7の局所パターン中の累乗)とに基づいて1つのセルの情報について

50

局所パターンの数に対応した複数の演算値を求める。ここで求める演算値とは、図6に示す9つの画素の濃度 $N_1$ を局所パターン $M_1$ の対応する位置の累乗数で累乗して得た値を加算した値(2<sup>1</sup>)である。図6に示す9つの画素の濃度 $N_1$ と局所パターン $M_8$ の演算値は、(2<sup>2</sup> + 3<sup>1</sup>)である。この演算を図8に示すように便宜状  $N_1^{M_1}$ と表記するものとする。具体的には、図6に示す9つの画素の濃度 $N_1$ と35種類の局所パターン $M_1 \sim M_{35}$ の演算値をそれぞれ求め、この演算値を各セルについての高次局所自己相関特徴として抽出する。そしてこのようにして抽出した複数の高次局所自己相関特徴に基づいて各フレームに示された身体形状の高次局所自己相関パターンを抽出する。具体的には、図8に示すように、各セルについての高次局所自己相関特徴を加算した値により高次局所自己相関パターンを抽出する。図8は、高次局所自己相関パターンの作り方を概念的に示している。そして抽出した複数の高次局所自己相関特徴に基づいて各フレームに示された身体形状の高次局所自己相関パターンを抽出し、高次局所自己相関パターンに基づいて動画像に含まれる各フレームに示された身体形状を判定するための判定基準を定める(ステップS104)。

10

## 【0021】

例えば、マウス様の機能を実現するには、ポインティング、マウスの右ボタン、左ボタンを判別しなければならない。また、ジェスチャの入力キーとなる手形状を加えると最低でも4クラスを識別しなければならない。本実施の形態では、例えば手形状として図9の4種類を設定するものとする。図9の(a)はポインティング、(b)はマウスの右ボタン、(c)は左ボタン、(d)はジェスチャの入力キーに対応する。本例では、これらのジェスチャについてそれぞれ判定基準を定める。

20

## 【0022】

なお前述の対数極座標画像Lからの特徴抽出には高次局所自己相関関数(栗田多喜夫、“柔らかな情報処理のための統計的手法の応用に関する研究、”電総研研究報告、vol. j957、pp. 126 - 140、1993.)を用いることができる。これは画面内の対象画像を $f(r)$ とすると、N次自己相関関数は、変位方向( $a_1, a_2, \dots, a_N$ )に対して式(2)として定義される。

## 【数1】

$$x^N(a_1, a_2, \dots, a_N) = \int f(r)f(r+a_1)\dots f(r+a_N)dr \quad (2)$$

30

## 【0023】

ここでは高次自己相関係数の次数Nを2とし、また、変位方向を参照点rの周りの局所的な3×3画素の領域に限定した。

## 【0024】

局所パターンとしては、図10に示す公知の局所パターンを用いてもよい。図10の例では、平行移動による等価な特徴を除いているため、特徴の数は25個になる。図中の黒で塗りつぶした部分は、局所パターンの対応する画素の位置に対応する。局所パターンの対応する画素の値の積を全画素に対して足し合わせることで、各セルについての高次局所自己相関特徴が抽出される。

40

## 【0025】

図1に示すように、実際の判定を行うためには、動画像を入力し(ステップS201)、各フレームについての対数極座標画像を複数の画素からなるセルに切り分けて(図2のステップS21)各セルについての実際の高次局所自己相関特徴を抽出する(ステップS202)。そして抽出した複数の実際の高次局所自己相関特徴に基づいて動画像入力各フレームに示された身体形状の実際の高次局所自己相関パターンを抽出して順次記憶し(ステップS203:図2のステップS22)、実際の高次局所自己相関パターンと判定

50



基準を用いてフレームに示された身体形状を判定する（ステップS204）。

【0026】

本実施の形態では、図2に示すように、身体形状の判定にニューラルネットの自己組織化を用いる。図11は、ニューラルネットの自己組織化を説明するために用いる図である。図11において、入力層はn個のニューロンを持ち、本実施の形態では局所自己相関関数の局所パターン数と同じく $n = 35$ となる。入力層と結合層のあいだには、0から1までの間で連続量で値が変動する結合荷重Wがある。各入力層からの入力は結合加重により重み付け線形加算され結合層ニューロンの出力となる。出力層ニューロンの数は弁別したいパターン数であり、10動作を識別できるようにする場合には、 $m = 10$ とする。

【0027】

最初に、結合加重の初期値を0から1までのあいだでランダムに与える。次に、結合加重の更新を、以下の手順で行う。

【0028】

(1) 次式により、入力ベクトル $v$ と結合加重ベクトル $W$ のユークリッド距離が最小となる $c$ を計算し、それを最適ニューロン $c$ として選ぶ（図2のステップS25）。

【数2】

$$c = \min \{ \| v - W_m \| \} \quad (3)$$

【0029】

(2) 最適ニューロン $c$ を中心として近傍 $d$ にあるニューロンを近傍ニューロン $N_c$ とする。 $d$ を求める式は、以下の通りである。

【0030】

$$d = d(0) \cdot (1.0 - t / rlen) \quad \dots (4)$$

ここで、 $t$ は経過回数、 $rlen$ は総学習回数（100,000）、 $d(0)$ は初期値（=2）である。

【0031】

(3) 結合加重 $W$ を、次式に従って更新する（図2のステップS26）。

【数3】

$$\begin{aligned} W_i(t+1) &= W_i(t) + a(t) (x(t) - W_i(t)) && \text{if } N_i \in N_c(t) \\ W_i(t+1) &= W_i(t) && \text{others} \end{aligned} \quad (5)$$

【0032】

ここで、 $a(t)$  ( $0 < a(t) < 1$ ) は時間とともに単調に減衰する学習パラメータであり、次式に従うものとする。

【0033】

$$a(t) = a(0) \cdot (1.0 - t / rlen) \quad \dots (6)$$

ここで、 $t$ は経過回数、 $rlen$ は総学習回数（100,000）、 $a(0)$ は初期値（=0.02）である。

【0034】

(1) から (3) までの操作を多数回（100,000）繰り返す（図2のステップS27）。

【0035】

(4) 弁別したいパターン（クラス。本件では10種類）ごとに最も多くの入力データとマッチする出力層ニューロンを見つけ、クラスを割り振る。

【0036】

(5) 続いて、以下の手順により教師信号を与えてパターン分類能力を付加する（図2のステップS28及びS29）。

【0037】

10

20

30

40

50

$W_c(t+1) = W_c(t) + \beta(t)(v(t) - W_c(t)) \dots (7)$   
 入力ベクトル  $v$  と結合加重ベクトル  $W$  が同じクラスに分類されていたら下記のようになる。

【数4】

$$\begin{aligned} W_c(t+1) &= W_c(t) - \beta(t)(v(t) - W_c(t)) && \text{同じクラスに分類されない場合} \\ W_i(t+1) &= W_i(t) && W_c \text{以外の結合加重 } W_i \end{aligned} \quad (8)$$

【0038】

10

ここで、 $\beta(t)$  は学習パラメータであり、次式に従うものとする。

【0039】

$$\beta(t) = \beta(t-1) / (1 + s(t) \cdot \beta(t-1)) \dots (9)$$

ここで、 $s(t)$  は分類が正しければ +1、正しくなければ -1 である。

【0040】

十分なサンプル数で上記手続き(5)の操作を繰り返すことにより、パターン弁別能力が獲得される。

【0041】

ニューラルネットを用いずに、各クラスにおいて抽出した特徴をもとに重回帰式を求め、識別には重回帰式の目的変数が最大値のクラスを選択し、その最大値が閾値以上のときに、そのクラスに決定するようにしてもよい。

20

【0042】

下記の表1は、非特許文献9に記載の従来の方法により図9のパターンの識別実験を行った結果を示している。また下記の表2は、上記ニューラルネットを用いた実施の形態により図9のパターンの識別実験を行った結果を示している。実験では、手の大きさを変え、また対数極座標リサンプリングの解像度も4種類異ならせて行った。表1及び表2の平均及び標準偏差を対比すると明らかなように、上記実施の形態の方法によれば、精度(識別力)が平均70%台から90%以上に向上した。また識別の不安定さ(ばらつき)が1/3以下に小さくなった。なおパターンを増やしてもこれらの結果の傾向に大きな差は生じない。

30

【表1】

手の大きさ	1	2	3	4	5	平均	標準偏差
LPMリサンプリングの解像度							
120×120画素	70.6	83.8	76.5	82.3	79	78.4	5.22
60×60画素	80.4	74.1	78.7	70.9	61	73	7.7
40×40画素	82.1	79.6	59.6	67.1	66	70.9	9.58
30×30画素	79.7	80.7	71	77.9	69	75.7	5.31
平均	78.2	79.6	71.5	74.6	68.8	74.5	
標準偏差	5.17	4.05	8.54	6.83	7.59		6.66

40

【表 2】

手の大きさ LPMリサンプリング の解像度	1	2	3	4	5	平均	標準偏差
120×120画素	84.1	91.5	92.3	93.6	90.8	90.5	3.7
60×60画素	87.3	95.5	95.4	94.1	90.3	92.5	3.6
40×40画素	88.4	95.2	94.9	93.8	92.4	92.9	2.77
30×30画素	88.6	95.4	95.1	95.3	91.6	93.2	3.02
平均	87.1	94.4	94.4	94.2	91.3	92.3	
標準偏差	2.08	1.94	1.43	0.76	0.92		2.25

10

## 【0043】

前述の対数極座標変換による処理の利点は、少ない画素数で高解像度、広域の作業範囲を得るだけでなく、スケーリング不変性、変換中心に対する回転不変性を持っており、また、画像のデータ量を減らし画像処理の計算時間を減少させることも期待できる点である。しかし、欠点として非均一なサンプリングなため平行移動に対して画像の形が大きく変わってしまうことから、動的視覚処理には不向きであるとされてきた (Richard Wallace, Ping-Wen Ong, Ben Bederson, and Eric L. Schwartz, "Space Variant Image Processing" International Journal of Computer Vision, vol. 13, no. 1, pp. 71-90, 1994.)。この欠点を解決するために、対数極座標上で連続した画像からオプティカルフローを求め、その運動方向が2で変化することを利用して平行移動を求める手法 (Noboru Okajima, Hiroki Nitta and Wataru Mitsuhashi "Motion Estimation and Target Tracking in The Log-Polar Geometry") や、直交座標系における2次元フーリエ変換と同等の結果を対数極座標系から算出できる指数チャープ変換 (Giorgio Bonmassar, and Eric L. Schwartz, "Space-Variant Fourier Analysis: The Exponential Chirp Transform," IEEE Pattern Analysis and Machine Vision, vol. 19, no. 10, pp. 1080-1089, 1997) を用いることで、通常の2次元フーリエ変換と同じように平行移動に対して不変なパラメータを抽出する手法 (野村智宣, 三橋渉, "中心窩センサの設計と物体認識への応用," 信学技報 EID2000-313, pp. 77-82, 2001.) が提案されている。しかし、前者は通常の動きベクトルの算出が、画像が変形を伴わずに動くことを前提としている。そのため対数極座標上の画像が物体の平行移動により大きく歪むことから、対象物体が大きく移動すると動きベクトルが正しく算出できず、平行移動のパラメータをうまく算出できないという欠点がある。また、後者は通常の変換より処理時間がかかるため、実時間の処理には不向きである。

20

30

40

## 【0044】

そこで本実施の形態では、図1に示すように、対数極座標画像から時間差分 (S301)、空間差分 (ステップ S302)、肌色情報 (S303) を基に輪郭画像を生成し、この輪郭画像の重心をもとに手領域の位置を推定する (S304, S305)。また、手領域の過去の位置情報から移動先を推定する (S306)。そして、肌色情報を用いて抽出した手領域から高次局所自己相関特徴を算出し、重回帰分析により識別を行なう。また、手領域の軌跡を辞書内のデータと比較することでジェスチャ認識を行う。これより、平行移動量をチャープ変換などの複雑な計算をせずに算出することで処理の高速化が図れ、また、詳細な形状に関する情報が捨象された重心を位置の推定に用いることで、対象物体の急激な変化による追跡精度の悪化の可能性を低減させる。また、肌色領域から抽出した色

50

情報と背景差分を用いることで、背景領域の類似した色情報をもつ物体を除去し、対象物体を抽出することができる。

【0045】

手領域の位置推定について詳しく説明する。まず人物の手領域を抽出し追跡を行うために、時間差分、空間差分、肌色情報を用いる（ステップS301～S303）。背景と移動領域の分離は、時系列画像中の移動領域は観測された輝度値の時間的变化に対応するため、時間差分を用いることで容易に行える。しかし、時間差分の結果は、背景と移動領域の輝度値の差に依存してしまうため、照明の変化に影響を受けやすい。そのため、画像中の輝度値の変化が大きい部分は、移動領域のエッジ付近にあることを考慮し、照明の変化に影響されがたい空間差分により得られるエッジ情報を併用する。また、ここでは手領域の抽出が目的であるため、その大きな特徴である肌色情報を用いる（ステップS304）。以下、入力画像より手指領域の抽出方法について述べる。入力画像より作成した時刻  $t$  での対数極座標画像を得る。入力画像より作成した時刻  $t$  での対数極座標画像を  $L(p, q, t)$  とし、この  $L(p, q, t)$  の時間差分により得られる画像  $L_1(p, q, t)$ 、空間差分により得られる画像を  $L_2(p, q, t)$ 、肌色領域  $L_3(p, q, t)$  を次のように計算する。時間差分画像  $L_1(p, q, t)$  は、式(10)に示すように連続した2フレーム間の差分をとることにより得られる。これより時刻  $t$  の静止領域と運動領域の分割が行われる。

【数5】

$$L_1(p, q, t) = \begin{cases} 1 & , \quad |L(p, q, t) - L(p, q, t-1)| \geq th_t \\ 0 & , \quad otherwise \end{cases} \quad (10)$$

【0046】

ただし、 $th_t$  は輝度値が変化したかを決定する閾値である。

【0047】

空間差分画像  $L_2(p, q, t)$  は、時刻  $t$  における画像に対して、式(11)に示す  $3 \times 3$  の Sobel フィルタを施すことで、画像中のエッジを抽出することで得る。

【数6】

$$L_2(p, q, t) = \begin{cases} 1 & , \quad \sqrt{LHS(p, q, t)^2 + LVS(p, q, t)^2} \geq th_s \\ 0 & , \quad otherwise \end{cases} \quad (11)$$

【0048】

ただし、 $th_s$  はエッジであるかを決定する閾値である。また  $L_2(p, q, t)$ 、 $LHS(p, q, t)$ 、 $LVS(p, q, t)$  は、それぞれ時刻  $t$  における座標  $(p, q)$  の空間差分値、 $p$  方向 Sobel オペレータによる値、 $q$  方向 Sobel オペレータによる値を表している。入力画像は RGB 表色系で表現されているが、この RGB 表色系では、各値間の相関が高いため、明るさの変化に大きく影響を受けるという問題がある。そのため本稿では、明るさの変動に影響を受けず、RGB それぞれの値が1対1で対応づけられる  $L^*u^*v^*$  表色系 (CIE 1976 表色系) に変換し、その中の  $u$ 、 $v$  平面を用い、肌色領域の平均値  $M(u, v)$ 、分散共分散行列  $C$  を求める。これより、肌色領域  $L_3(p, q, t)$  は、式(12)に示すようになる。背景領域の壁やダンボールなどの肌色領域が存在する場合にはその領域も抽出されてしまうが、前述の差分画像との論理積をとることで背景領域として静止している壁やダンボールの領域を除去することができる。

【数 7】

$$L_3(p, q, t) = \begin{cases} 1 & , (L_c - M)^T \mathbf{C}^{-1} (L_c - M) \geq th_c \\ 0 & , otherwise \end{cases} \quad (12)$$

【0049】

ただし、 $th_c$  は肌色であるかを決定する閾値である。このように式 (10) から式 (12) より得られた情報から式 (13) に示す輪郭画像を得る。 10

【数 8】

$$L_d(p, q, t) = \begin{cases} 1 & , \sum_{i=1}^3 L_i(p, q, t) = 3 \\ 0 & , otherwise \end{cases} \quad (13)$$

20

【0050】

この得られた輪郭画像  $L_d$  から重心を算出することにより、時刻  $t$  における入力画像  $I$  中の手の位置  $p_{xy}(t)$  が算出される。

【0051】

次に、手領域の重心位置を追跡する方法について述べる。時刻  $t$  における手領域の重心位置を  $p_{xy}(t)$ 、手領域の移動速度を  $v(t)$ 、加速度を  $a(t)$ 、時刻  $t-1$  時による時刻  $t$  での手領域の推定重心位置を  $\hat{p}_{xy}(t)$  とする。ただし、時刻  $t=0$  では手領域の重心位置、推定重心位置を取り込み画像の中心点とし、速度、加速度はともに 0 と仮定する。時刻  $t$  における手領域の推定重心位置  $\hat{p}_{xy}(t)$  は、式 (14) として求まる。この推定重心位置  $\hat{p}_{xy}(t)$  は、直交画像から LPM 画像に変換する場合の対数極座標空間の中心点となる。ただし、式中の  $t$  はフレーム間隔を表す。 30

【数 9】

$$\hat{p}_{xy}(t) = p_{xy}(t-1) + v(t-1) \cdot \Delta t \quad (14)$$

【0052】

ただし、実際の手領域の重心位置  $p_{xy}(t)$  と推定重心位置  $\hat{p}_{xy}(t)$  が一致しない場合は、時刻  $t-1$  から時刻  $t$  までの間に、式 (15) に示す加速度が生じたものとし、移動速度  $v(t)$  は式 (16) とする。 40

【数 10】

$$a(t-1) = \frac{2}{(\Delta t)^2} (\hat{p}_{xy}(t) - p_{xy}(t)) \quad (15)$$

$$v(t) = v(t-1) + a(t-1) \cdot \Delta t \quad (16)$$

【0053】

次にジェスチャ推定について説明する。

【0054】

50

ジェスチャ動作は、図12(a)のCircle、(b)のTriangle、(c)のRectangle、(d)のCrossの4種類とした。手指によるジェスチャ認識の場合、一連の手指動作をジェスチャ動作区間とその他の動作区間に分割しなければならない。この場合、連続DPマッチングを用いてジェスチャ区間を指定せずにジェスチャのスポッティング認識を行う手法(西村拓一、向井理朗、野崎俊輔、岡隆一、“動作者適応のためのオンライン教示可能なジェスチャ動画像のスポッティング認識システム、”電子情報通信学会論文誌D-II, vol. J81-D-II, no. 8, pp. 1822-1830, 1998.)、手指の動作が極小となる時点分割位置とする手法(佐川浩彦、竹内勝、“手話認識における手動作セグメンテーション方式、”ヒューマンインタフェースシンポジウム'99, pp. 749-754, 1999.)などがある。計算機の10  
 入力デバイスとしてのインターフェースとしてジェスチャ認識を用いることを考慮すると、任意の動作から対象とするジェスチャを認識するよりは、ユーザが入力を意図したときにジェスチャを認識をしたほうが誤動作などを招かなくて都合がよい。そこで本手法では、ユーザが手の動きが極小となり、手形状がシステムに登録されているジェスチャ入力キーと一致したときを動作の開始、または、終了の点として扱い、その間の区間をジェスチャ区間とみなし単純DPマッチングによる照合を行う。一般に画像からの動作区間の推定は、手指の動作が極小となる時点を検出することは難しいが、本手法では動作区間の推定情報として手指動作が極小であるという情報に加え、そのときの手指形状がシステムに登録されているかという情報を加えることで、手指動作の動作区間の決定を実現している。ジェスチャ推定の特徴ベクトルSは、ジェスチャ区間の手指の追跡軌跡から20  
 得られる軌道ベクトルを用いる。時刻tにおける手の位置を $p \times y(t) = (x_t, y_t)$ とすると、時刻tの特徴ベクトル $s(t)$ は式(17)となる。

【数11】

$$s(t) = (v(t), \theta(t)) \quad (17)$$

$$v(t) = (x_t - x_{t-1}, y_t - y_{t-1})$$

$$\theta(t) = \cos^{-1} \frac{v(t) \cdot v(t-1)}{|v(t)| |v(t-1)|}$$

30

【0055】

これより、ジェスチャの推定は、特徴ベクトル $\{S: s_1(v, \theta), s_2(v, \theta), \dots, s_i(v, \theta), \dots, s_N(v, \theta)\}$ と辞書内の参照特徴ベクトル $\{T: t_1(v, \theta), t_2(v, \theta), \dots, t_j(v, \theta), \dots, t_M(v, \theta)\}$ とのコストをDPマッチングにより算出し、その中から最小のものを選択することにより行う。

【0056】

ステップS308の認識結果と、ステップS204の判別結果とは、例えばパソコンにインストールされた各種のアプリケーションの選択や起動のために利用可能である。

【図面の簡単な説明】

40

【0057】

【図1】本発明の方法を用いて手の形状及び動作(身体形状)により非接触型ポインティング・デバイスを実現する場合のステップをブロック図形式で示したものである。

【図2】画像に含まれる身体形状を判定する方法をコンピュータを用いて実現する場合に用いるソフトウェアのアルゴリズムを示すフローチャートである。

【図3】対数極座標変換を説明するために用いる図である。

【図4】(a)乃至(d)は、対数極座標変換の実際の例を説明するために用いる図である。

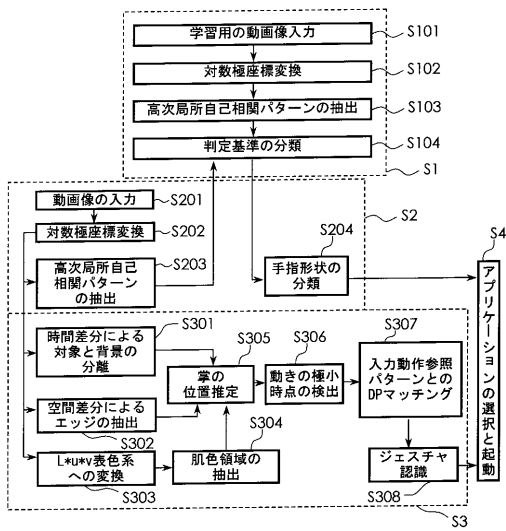
【図5】フレームからセルを切り出す方法を説明するために用いる図である。

【図6】濃度を示す情報を含む1つのセルを示す図である。

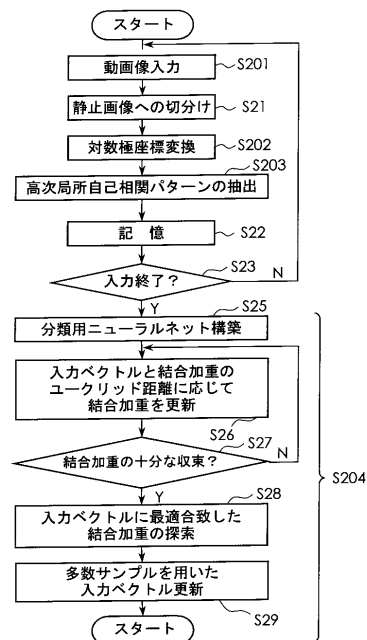
50

- 【図7】 35種類の局所パターンの例を示す図である。
  - 【図8】 高次局所自己相関パターンの概念を示す図である。
  - 【図9】 (a)乃至(d)は、手形の例を示す図である。
  - 【図10】 25種類の他の局所パターンの例を示す図である。
  - 【図11】 ニューラルネットの自己組織化を説明するための図である。
  - 【図12】 (a)乃至(d)は、ジェスチャのパターンを示す図である。
- 【符号の説明】
- 【0058】
- 1 フレーム
  - N1 セル
  - M1 ~ M35 局所パターン

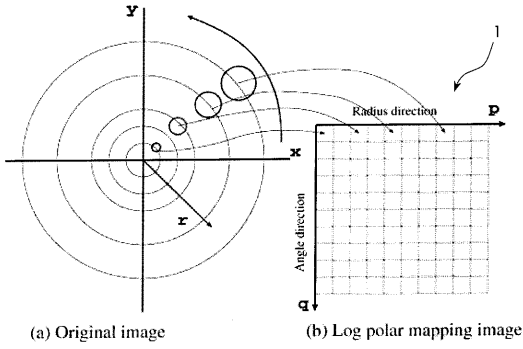
【図1】



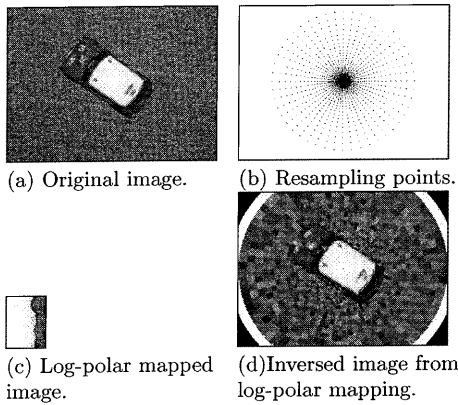
【図2】



【 図 3 】



【 図 4 】



【 図 5 】

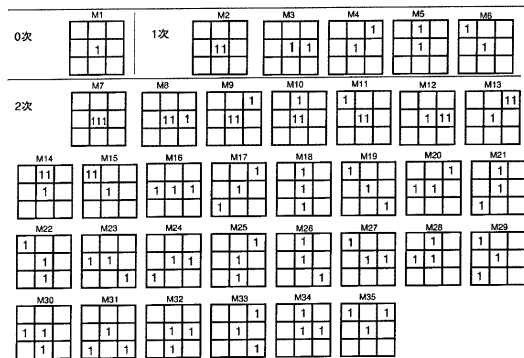
1	2	3	4	5	6	7
101	102	103	104	105	106	107
201	202	203	204	205	206	207
301	302	303	304	305	306	307

【 図 6 】

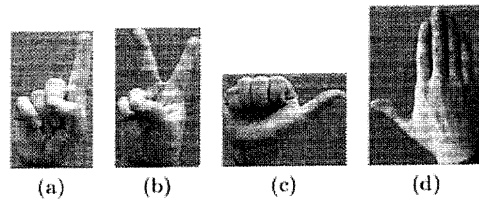
N1

0	0	1
0	2	3
1	3	4

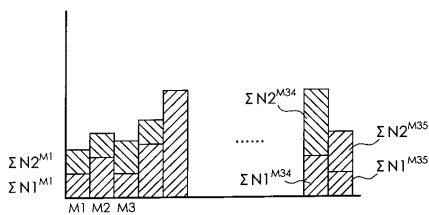
【 図 7 】



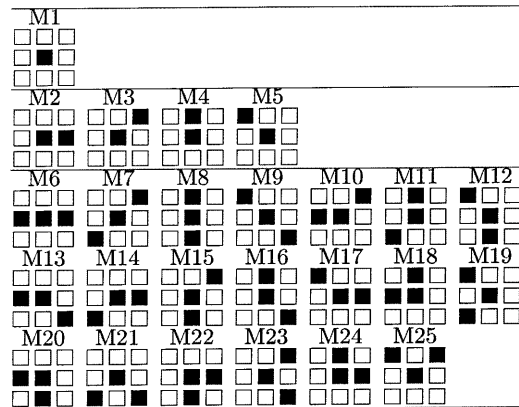
【 図 9 】



【 図 8 】

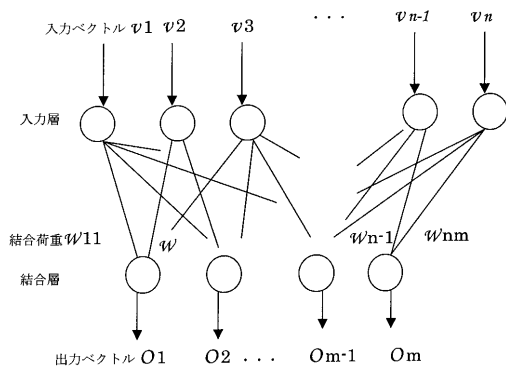


【 図 10 】





【図 1 1】



【図 1 2】



---

フロントページの続き

(56)参考文献 特開平09-153137(JP,A)  
特開平08-226961(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06T	7/00
G06F	3/033
G06N	3/08
G06T	1/00