

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2005-78376

(P2005-78376A)

(43) 公開日 平成17年3月24日(2005.3.24)

(51) Int.C1.⁷

G06T 1/00

B25J 19/04

G06T 7/00

F 1

G06T	1/00	340A	3C007
G06T	1/00	510	5B057
B25J	19/04		5L096
G06T	7/00	100C	
G06T	7/00	300D	

テーマコード(参考)

審査請求 未請求 請求項の数 21 O L (全 32 頁)

(21) 出願番号

特願2003-307924 (P2003-307924)

(22) 出願日

平成15年8月29日 (2003.8.29)

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(74) 代理人 100067736

弁理士 小池 晃

(74) 代理人 100086335

弁理士 田村 榮一

(74) 代理人 100096677

弁理士 伊賀 誠司

(72) 発明者 森貞 英彦

東京都品川区北品川6丁目7番35号 ソニーブルーバード内

Fターム(参考) 3C007 CS08 KS11 KS36 KT01 KT11
KT15 WA04 WA14 WB17

最終頁に続く

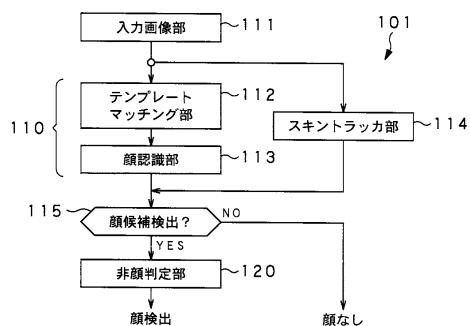
(54) 【発明の名称】 対象物検出装置、対象物方法、及びロボット装置

(57) 【要約】

【課題】 演算量を増大させることなく且つ簡便な方法にて、対象物を誤って検出した場合の誤検出を低減することができる対象物検出装置及び対象物検出方法、及びロボット装置を提供する。

【解決手段】 顔検出装置101は、入力画像からテンプレートマッチングでおおまかに顔候補を選択し、サポートベクタマシンにより顔認識して顔候補を検出する従来と同様の顔検出部110と、顔検出部110にて一旦顔であるとされた顔候補から顔でない非顔候補を検出して排除する非顔判定部120と、非顔判定後の顔領域をトラッキングするスキントラッカ部114とからなる。非顔判定部120は、入力画像から算出した顔との想定距離と距離センサにより測定した測定距離との差が大きい場合、顔候補の色の分散値が小さい場合、肌色領域の占有率が大きい場合、所定期間経過後の顔領域の大きさ変化が大きい場合等に非顔であるとして顔候補から排除する。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

入力画像から検出対象となる対象物を検出する対象物検出装置において、
上記対象物であると判定した対象物候補を検出する対象物検出手段と、
上記対象物候補から対象物ではないと判定した非対象物候補を検出し、該非対象物候補
を該対象物候補から除いて出力する非対象物検出手段とを有する
ことを特徴とする対象物検出装置。

【請求項 2】

上記検出対象は、顔領域であって、
上記対象物検出手段は、上記入力画像から所定の顔条件を満たす領域を顔である顔候補 10
と判定し、該顔候補を検出する顔検出手段であり、
上記非対象物検出手段は、上記顔候補から、所定の非顔条件を満たすものを顔ではない
非顔候補と判定し、該非顔候補を該顔候補から除いて出力する非顔検出手段である
ことを特徴とする請求項 1 記載の対象物検出装置。

【請求項 3】

検出対象物との間の距離を測定する距離測定手段を有し、
上記非顔検出手段は、上記入力画像に対する上記顔候補の大きさから検出対象物との間
の距離である想定距離を算出し、該想定距離と上記距離測定手段により測定された測定距離
との差が所定の閾値以上である場合に、上記顔候補を上記非顔候補と判定する
ことを特徴とする請求項 2 記載の対象物検出装置。 20

【請求項 4】

上記非顔検出手段は、上記顔候補の色空間上における分散値を算出し、該分散値が所定
の閾値未満である場合に、上記顔候補を上記非顔候補と判定する
ことを特徴とする請求項 2 記載の対象物検出装置。

【請求項 5】

上記非顔検出手段は、現在の入力画像における上記顔候補の大きさと、所定時間前の入
力画像における上記顔候補の大きさとの差が所定の閾値以上である場合に、上記顔候補を
上記非顔候補と判定する
ことを特徴とする請求項 2 記載の対象物検出装置。

【請求項 6】

上記非顔検出手段は、上記顔領域に対する肌色領域の占有率を算出し、該肌色領域の占
有率が所定の閾値以上である場合に、上記顔候補を上記非顔候補と判定する
ことを特徴とする請求項 2 記載の対象物検出装置。 30

【請求項 7】

上記非顔検出手段により非顔候補が除かれた顔候補のトラッキングをするトラッキング
手段を有し、
上記非顔検出手段は、所定のタイミングにて上記トラッキング手段のトラッキング対象
となっている顔候補から上記非顔候補を検出する
ことを特徴とする請求項 2 記載の対象物検出装置。

【請求項 8】

上記顔検出手段は、上記入力画像と、平均的な顔画像を示す所定サイズのテンプレート
との相関値を求めるマッチング結果生成手段と、該相関値に基づき選択された入力画像領域
からパターン認識により上記顔候補を識別する識別手段とを有する
ことを特徴とする請求項 2 記載の対象物検出装置。 40

【請求項 9】

入力画像から検出対象となる対象物を検出する対象物検出方法において、
上記対象物であると判定した対象物候補を検出する対象物検出工程と、
上記対象物候補から対象物ではないと判定した非対象物候補を検出し、該非対象物候補
を該対象物候補から除いて出力する非対象物検出工程とを有する
ことを特徴とする対象物検出方法。 50

【請求項 10】

上記検出対象は、顔領域であって、

上記対象物検出工程は、上記入力画像から所定の顔条件を満たす領域を顔である顔候補と判定し、該顔候補を検出する顔検出工程であり、

上記非対象物検出工程は、上記顔候補から、所定の非顔条件を満たすものを顔ではない非顔候補と判定し、該非顔候補を該顔候補から除いて出力する非顔検出工程である

ことを特徴とする請求項9記載の対象物検出方法。

【請求項 11】

上記非顔検出工程では、上記入力画像に対する上記顔候補の大きさから検出対象物との間の距離である想定距離を算出し、該想定距離と検出対象物との間の距離を測定する距離測定手段により測定された測定距離との差が所定の閾値以上である場合に、上記顔候補を上記非顔候補と判定する

ことを特徴とする請求項10記載の対象物検出方法。

【請求項 12】

上記非顔検出工程では、上記顔候補の色空間上における分散値を算出し、該分散値が所定の閾値未満である場合に、上記顔候補を上記非顔候補と判定する

ことを特徴とする請求項10記載の対象物検出方法。

【請求項 13】

上記非顔検出工程では、現在の入力画像における上記顔候補の大きさと、所定時間前の入力画像における上記顔候補の大きさとの差が所定の閾値以上である場合に、上記顔候補を上記非顔候補と判定する

ことを特徴とする請求項10記載の対象物検出方法。

【請求項 14】

上記非顔検出工程では、上記顔領域に対する肌色領域の占有率を算出し、該肌色領域の占有率が所定の閾値以上である場合に、上記顔候補を上記非顔候補と判定する

ことを特徴とする請求項10記載の対象物検出方法。

【請求項 15】

上記非顔検出工程にて非顔候補が除かれた顔候補のトラッキングをするトラッキング工程と、

所定のタイミングにて上記トラッキングの対象となっている顔候補から上記非顔候補を検出する非顔再検出工程とを有する

ことを特徴とする請求項10記載の対象物検出方法。

【請求項 16】

入力画像から検出対象となる対象物を検出する対象物検出装置を搭載し自律的に動作するロボット装置において、

上記対象物検出装置は、

上記対象物であると判定した対象物候補を検出する対象物検出手段と、

上記対象物候補から対象物ではないと判定した非対象物候補を検出し、該非対象物候補を該対象物候補から除いて出力する非対象物検出手段とを有する

ことを特徴とするロボット装置。

【請求項 17】

上記対象物検出装置は、上記入力画像から検出対象である顔領域を検出するものであつて、

上記対象物検出手段は、上記入力画像から所定の顔条件を満たす領域を顔である顔候補と判定し、該顔候補を検出する顔検出手段であり、

上記非対象物検出手段は、上記顔候補から、所定の非顔条件を満たすものを顔ではない非顔候補と判定し、該非顔候補を該顔候補から除いて出力する非顔検出手段である

ことを特徴とする請求項16記載のロボット装置。

【請求項 18】

検出対象物との間の距離を測定する距離測定手段を有し、

10

20

30

40

50

上記非顔検出手段は、上記入力画像に対する上記顔候補の大きさから検出対象物との間の距離である想定距離を算出し、該想定距離と上記距離測定手段により測定された測定距離との差が所定の閾値以上である場合に、上記顔候補を上記非顔候補と判定する

ことを特徴とする請求項17記載のロボット装置。

【請求項19】

上記非顔検出手段は、上記顔候補の色空間上における分散値を算出し、該分散値が所定の閾値未満の場合に、上記顔候補を上記非顔候補と判定する

ことを特徴とする請求項17記載のロボット装置。

【請求項20】

上記非顔検出手段は、現在の入力画像における上記顔候補の大きさと、所定時間前の入力画像における上記顔候補の大きさとの差が所定の閾値以上である場合に、上記顔候補を上記非顔候補と判定する

ことを特徴とする請求項17記載のロボット装置。

【請求項21】

上記非顔検出手段は、上記顔領域に対する肌色領域の占有率を算出し、該肌色領域の占有率が所定の閾値以上である場合に、上記顔候補を上記非顔候補と判定する

ことを特徴とする請求項17記載のロボット装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、検出対象となる例えば顔や物体等の対象物を検出する対象物検出装置、対象物検出方法、及びこれを搭載したロボット装置に関し、特に、検出対象以外のものを検出してしまう誤検出の低減を図った対象物検出装置、対象物検出方法、及びこれを搭載したロボット装置に関する。

【背景技術】

【0002】

電気的又は磁気的な作用を用いて人間（生物）の動作に似た運動を行う機械装置を「ロボット装置」という。我が国においてロボット装置が普及し始めたのは、1960年代末からであるが、その多くは、工場における生産作業の自動化・無人化等を目的としたマニピュレータ及び搬送ロボット装置等の産業用ロボット装置（Industrial Robot）であった。

【0003】

最近では、人間のパートナーとして生活を支援する、すなわち住環境その他の日常生活上の様々な場面における人的活動を支援する実用ロボット装置の開発が進められている。このような実用ロボット装置は、産業用ロボット装置とは異なり、人間の生活環境の様々な局面において、個々に個性の相違した人間、又は様々な環境への適応方法を自ら学習する能力を備えている。例えば、犬又は猫のように4足歩行の動物の身体メカニズム及びその動作を模した「ペット型」ロボット装置、或いは、2足直立歩行を行う人間等の身体メカニズム及びその動作をモデルにしてデザインされた「人間型」又は「人間形」ロボット装置（Humanoid Robot）等のロボット装置は、既に実用化されつつある。

【0004】

これらのロボット装置は、産業用ロボット装置と比較して、例えばエンターテインメント性を重視した様々な動作等を行うことができるため、エンターテインメントロボット装置と呼称される場合もある。また、そのようなロボット装置には、CCD（Charge Coupled Device）カメラ及びマイクロホン等の各種外部センサが搭載され、これら外部センサの出力に基づいて外部状況を認識して、外部からの情報及び内部の状態に応じて自律的に動作するものがある。

【0005】

ところで、かかるエンターテインメント型のロボット装置において、対話中にその相手となる人間の顔や、移動中に視界内に入る人間の顔を検出して、その人間の顔を見ながら

10

20

30

40

50

対話や動作を行うことができれば、人間が普段行う場合と同様に、その自然性から考えて最も望ましく、エンターテインメントロボット装置としてのエンターテインメント性をより一層向上させ得るものと考えられる。

【0006】

従来、動画のような複雑な画像シーンの中から色や動きを使うことなく画像信号に基づく濃淡パターンのみを使って、人間の顔を検出する方法が数多く提案されている。

【0007】

これらの顔検出方法としては、固有顔、ニューラル・ネットワーク及びサポートベクタマシン（SVM：Support Vector Machine）等のパターン認識の手法を利用して、予め顔パターンを学習させて識別器を生成する方法が挙げられる。

10

【0008】

【特許文献1】特開2002-42116号公報

【発明の開示】

【発明が解決しようとする課題】

【0009】

しかし、このパターン識別器を生成する方法によると、ロボット装置は、膨大なデータ量である顔画像データに対して学習によるパターン識別を行うにあたって、環境の変化や自己の姿勢及び表情の変化に対してロバスト性を示すが、その分当該パターン識別に要する演算量が増加するため、演算処理に要する時間が膨大になってしまふ。

20

【0010】

実際に、撮像画像の中から人間の顔画像を検出するプロセス（以下、これを顔検出タスクという。）においては、当該撮像画像の中から顔画像を切り出しながら識別を行うため、撮像画像全体を様々なスケールでスキャンすることとなる。このため一回一回のパターン識別に要する演算処理を極力少なくすることが極めて重要となる。

【0011】

例えば、サポートベクタマシンによるパターン認識を用いた顔検出タスクの場合、撮像画像の中から400（=20×20）画素程度の切り出し画像から得られる400次元のベクトルに対して、数百通りのサポートベクタ（400次元）との内積演算が必要となる。これを大きさ（W、H）からなる全画面中で行うとなると、当該内積演算を（W-20+1）×（H-20+1）回繰り返さなくてはならないため、膨大な量の演算処理となる。

30

【0012】

また、顔検出タスクをロボット装置に利用する場合には、動画の中から十分に早く顔画像を検出しなければ、リアルタイム性が要求されるロボット装置の行動としてフィードバックすることが極めて困難となる。更に、ロボット装置の内部のCPUは、顔検出タスク以外にも常時実行しているタスクが多くあるため、これらのタスクに対して演算能力を費やしている分、当該顔検出タスクに全ての演算能力を費やすことは極めて困難である。

【0013】

このような顔認識のみならず、物体等の対象物を入力画像から認識するような画像認識器においては、検出対象物との類似性やその特徴点等に基づく所定の条件を満たした場合にある対象物を検出対象物であるとして認識する。ここで、対象物の検出処理においては、上記所定の条件を厳しくした場合には、実際には対象物が存在するのにも拘わらず検出を見逃しきることになり、また上記所定の条件を緩くすると、対象物とは異なる物体を検出対象物であるとして検出する誤検出が多発するというトレードオフの関係がある。従つて、認識器の性能、入力画像を取得するためのカメラの性能、認識時の演算処理量等から最適な条件設定を行う必要がある。

40

【0014】

しかしながら、上述したように、特にリソースが限られたロボット装置等においては演算処理量を増加することは好ましくなく、また、画像認識器において見逃しをなくそうとすると誤検出が多発してしまう。

【0015】

50

本発明は、このような従来の実情に鑑みて提案されたものであり、演算量を増大させることなく且つ簡便な方法にて、対象物を誤って検出する誤検出を低減することができる対象物検出装置、対象物検出方法、及びロボット装置を提供することを目的とする。

【課題を解決するための手段】

【0016】

上述した目的を達成するために、本発明に係る対象物検出装置は、入力画像から検出対象となる対象物を検出する対象物検出装置において、上記対象物であると判定した対象物候補を検出する対象物検出手段と、上記対象物候補から対象物ではないと判定した非対象物候補を検出し、該非対象物候補を該対象物候補から除いて出力する非対象物検出手段とを有することを特徴とする。

10

【0017】

本発明においては、一度検出した対象物の中から非対象物を検出して排除することにより、例えば対象物の検出条件を緩くしてある程度の誤検出を許したとしてもこの非対象物検出手段により誤検出を排除することができるので、演算処理量を低減しつつ高精度の認識装置と同程度に誤検出を低減することができる。

【0018】

また、上記検出対象は、顔領域であって、上記対象物検出手段は、上記入力画像から所定の顔条件を満たす領域を顔である顔候補と判定し、該顔候補を検出する顔検出手段であり、上記非対象物検出手段は、上記顔候補から、所定の非顔条件を満たすものを顔ではない非顔候補と判定し、該非顔候補を該顔候補から除いて出力する非顔検出手段とすることにより、従来の顔検出装置の後段に、非対象物検出手段としての非顔検出手段を設けることにより、特に検出が難しい顔検出における誤検出を低減することができる。

20

【0019】

更に、検出対象物との間の距離を測定する距離測定手段を有し、上記非顔検出手段は、入力画像に対する上記顔候補の大きさから検出対象物との間の距離である想定距離を算出し、該想定距離と上記距離測定手段により測定された測定距離との差が所定の閾値以上である場合に、上記顔候補を上記非顔候補と判定することができ、検出された顔領域が人間の顔として想定できる大きさに比して非常に大きかったり、又は反対に小さかったりする場合は、顔でないと判定することができる。

30

【0020】

更にまた、上記非顔検出手段は、上記顔候補の色空間上における分散値を算出し、該分散値が所定の閾値未満である場合に、上記顔候補を上記非顔候補と判定することができ、ある程度色の分散がない場合には、単なる腕や足等の顔以外の肌色領域であったり、又は肌色の壁であったりする可能性が高く、これを顔ではないと判定することができる。

【0021】

また、上記非顔検出手段は、現在の入力画像における上記顔候補の大きさと、所定時間前の入力画像における上記顔候補の大きさとの差が所定の閾値以上である場合に、上記顔候補を上記非顔候補と判定することができ、短時間で大きさが著しく変化するような領域は顔ではないと判断することができる。

40

【0022】

更に、上記非顔検出手段は、上記顔領域に対する肌色領域の占有率を算出し、該肌色領域の占有率が所定の閾値以上である場合に、上記顔候補を上記非顔候補と判定することができ、人間の顔には、目や口等の肌色以外の領域が含まれるのにも拘わらず、肌色占有率が大きすぎる場合は、単なる腕や足等の顔以外の肌色領域や、肌色の壁などを検出しているものとして、非顔であると判定することができる。

【0023】

本発明に係る対象物検出方法は、入力画像から検出対象となる対象物を検出する対象物検出方法において、上記対象物であると判定した対象物候補を検出する対象物検出工程と、上記対象物候補から対象物ではないと判定した非対象物候補を検出し、該非対象物候補を該対象物候補から除いて出力する非対象物検出工程とを有することを特徴とする。

50

【 0 0 2 4 】

また、上記検出対象は、顔領域であって、上記対象物検出工程は、上記入力画像から所定の顔条件を満たす領域を顔である顔候補と判定し、該顔候補を検出する顔検出工程であり、上記非対象物検出工程は、上記顔候補から、所定の非顔条件を満たすものを顔ではない非顔候補と判定し、該非顔候補を該顔候補から除いて出力する非顔検出工程とすることができます。

【 0 0 2 5 】

本発明に係るロボット装置は、入力画像から検出対象となる対象物を検出する対象物検出装置を搭載し自律的に動作するロボット装置において、上記対象物検出装置は、上記対象物であると判定した対象物候補を検出する対象物検出手段と、上記対象物候補から対象物ではないと判定した非対象物候補を検出し、該非対象物候補を該対象物候補から除いて出力する非対象物検出手段とを有することを特徴とする。10

【 0 0 2 6 】

本発明においては、限られたリソースのロボット装置において、対象物が検出された後に誤検出防止手段を設けることにより、演算量の増加を最低限とし、簡単且つ高速に入力画像から対象物を検出する際の誤検出を防止する。

【 発明の効果 】**【 0 0 2 7 】**

本発明によれば、顔領域等の対象物検出装置において、検出対象物として一旦検出した対象物から、対象物ではないと判断できる非対象物を検出し、対象物から排除する非対象物検出手段を設けることにより、前段の対象物検出過程にて検出されてしまう誤検出を排除することができる。特に、複雑な対象物である顔は検出が難しく、誤検出が生じるが非対象物検出手段としての非顔検出手段を設けることにより、演算処理を増大させることなく、簡単且つ高速に誤検出を発見して排除することができる。更に、リソースが限られたロボット装置等にこのような非対象物検出手段を有する対象物検出装置を搭載すれば、顔等の検出対象物の誤検出を低減して更にエンターテイメント性を向上させることができる。20

【 発明を実施するための最良の形態 】**【 0 0 2 8 】**

以下、本発明を適用した具体的な実施の形態について、図面を参照しながら詳細に説明する。この実施の形態は、本発明を、検出対象とは異なるものを検出してしまった誤検出（誤認識）を低減した対象物検出装置のうち、特に顔検出を行う際の誤検出を低減する顔検出装置及びこれを搭載したロボット装置に適用したものである。30

【 0 0 2 9 】

ここでは、先ず、本実施の形態における対象物検出装置としての顔検出装置について説明し、次にこれを搭載したロボット装置について説明する。また、ここでは、顔検出の誤検出を低減する方法について説明するが、本発明は、顔検出に限ったものではなく、対象物を検出する一般的な画像認識における誤検出を低減する極めて効果的な手法として使用できるものである。

【 0 0 3 0 】**(1) 顔検出装置**

通常、顔検出装置は、入力画像（撮像画像）から人物の顔（顔領域）を検出するものである。この顔検出タスクは、大きく分けて2つの工程を有する。すなわち、入力画像に対してテンプレートマッチングを行って顔の候補を抽出する第1の工程と、この顔候補からサポートベクタマシン（SVM）等のパターン認識等により顔であるか否かの判定する第2の工程である。40

【 0 0 3 1 】

これに対し、本実施の形態における顔検出装置は、顔検出した後、すなわち、通常の顔検出タスクに対し、後工程（第3の工程）として、顔ではない領域を検出する非顔判定部を設け、これにより前段で検出された顔候補から顔ではないと判定した非顔候補を検出す50

る工程を設けるものである。そして、この非顔候補を顔候補から除くことで、誤検出の低減を図る。

【0032】

ロボット装置等のCPU及びメモリ等のリソースが限られたシステムにおいて、リアルタイムに顔検出を行うためには、顔検出タスクにおける演算量を削減することが必要である。しかしながら、第2の工程で行われるSVM等を使用する顔認識は膨大な演算量を要する。従って、顔検出タスクにおいて、顔検出性能を保持しつつ顔検出に要する演算量を削減するためには、第1の工程において顔候補を絞り込むことも有効である。

【0033】

ここで、第1の工程において、テンプレートマッチングを行って顔候補を抽出する際、例えはマッチングの相関値に対して閾値を設けて顔候補を抽出する方法がある。この方法では、顔候補の見逃しを軽減しようとした場合、相関値の閾値を上げる方法又はマッチングの際の間引きを減らす方法等をとることができるもの、例えは相関値の閾値を下げる多々の顔候補が抽出されるため、計算量が増大してしまい、ロボット装置等のリソースの限られた環境においては好ましくない場合がある。一方、閾値を上げると、顔候補の数を減らすことができるものの、本来顔である画像も候補から取り除いてしまい、顔画像を見逃してしまう場合がある。

【0034】

そこで、本願発明者等は、本来顔である画像を顔候補から見逃すことなく、演算量の増大を防止する、すなわち、顔候補の数を有効に絞り込むことができる方法を提案している（特願2003-133601号）。具体的には、第1の工程では、入力画像と、入力画像をスキャンして所定画素ずつずらしながら移動させた、平均的な顔画像を示す所定サイズのテンプレートとの相関値の集合であるマッチング結果を生成し、このマッチング結果における相関値の局所最大値に基づき顔領域を抽出する。顔領域は、マッチング結果を例えはテンプレートと同一サイズの複数の領域に分割して各分割領域毎に少なくとも相関値の最大値を顔候補として抽出する。そして、第2の工程では、上記顔候補として抽出された点に相当する入力画像領域からSVM等により顔領域（顔画像）を識別（判定）するものである。

【0035】

ところで、既に説明したように、顔認識処理は処理が膨大であり、認識処理の処理量をできるだけ減らしつつ、検出性能を向上する努力がなされている。しかしながら、ロボット装置のようなリソースが限られた装置においては、特に上記第2の工程等のパターン認識における認識率を上げることは容易ではなく、従って、例えは腕や足といった肌色部分や、肌色と同じような色の壁等を顔と間違える等の誤認識が生じる場合がある。

【0036】

そこで、本願発明者等が鋭意実験した結果、ある程度の誤認識を許容したとしても、顔検出処理の後段に、非顔判定部（non face判定部）を設けることで、顔であると判定し検出した顔領域（顔候補）のうち誤検出の非顔領域を効率的に取り除くことができ、結果として顔検出率を格段に向上させることができることを見出した。

【0037】

認識技術における問題点として、本来検出しなければならない検出対象物がないものとして見落としてしまう場合（以下、見逃しという。）と、検出対象物とは異なる対象物を検出対象物として認識し検出してしまった場合（以下、誤検出という。）とがある。本発明は、このうち、本来検出したい検出対象物以外のものを見誤って検出した場合である誤検出を低減するものである。ここで、見落としと、誤検出とはトレードオフの関係があり、認識条件を緩めれば当然、見逃しを減らすことができる代わりに、誤検出が増えることになる。

【0038】

例えは、上述したように、通常は、第1の工程にて顔と推定される領域を大まかに選択し、第2の工程によりそれらの領域から顔領域を抽出する、すなわち第2の工程にて顔領

10

20

30

40

50

域を絞り込むものであるが、例えばこの第2の工程における検出条件を緩めると、演算量を減らすことができるものの顔として検出される領域が増え誤検出も増加することになる。

【0039】

そこで、本実施の形態においては、従来顔検出処理が終了していた第2の工程の後工程に、誤検出防止手段としての非顔判定部により非顔判定処理を実行する第3の工程を設けることで、迅速且つ簡便に誤検出を排除するものである。

【0040】

図1は、本実施の形態における顔検出装置における顔検出処理を機能別に示すプロック図である。図1に示すように、顔検出装置101は、入力画像を供給する入力画像部111と、入力画像に対してテンプレートマッチングを行っておおまかな顔候補を抽出するテンプレートマッチング部112と、サポートベクタマシンSVMにより抽出されたおおまかな顔候補を識別して、顔であると判定した顔候補を抽出する顔識別部113とからなる顔検出部110と、顔検出部110にて顔候補が検出されたか否かを判定する顔検出判定部115と、顔検出部110にて顔候補が検出された場合に、その顔候補から、顔ではないと判定した非顔候補を検出し、顔候補から非顔候補を除いて出力する非顔検出手段としての非顔判定部120とを有する。

【0041】

また、この顔検出装置101は、非顔判定部120により、顔候補から非顔候補が排除された顔領域を追跡(トラッキング)するスキントラッカ部114を有している。このスキントラッカ部114のトラッキング対象となっている顔領域は、所定のタイミングにて非顔判定部120に送られ本当に顔であるか否か再度非顔判定が行われる。

【0042】

ここで、上記テンプレートマッチング部112、顔認識部113等の各処理についての一具体例は後述するが、非判定部120の前段に設ける顔検出部110としては、テンプレートマッチングやサポートベクタマシンによる顔認識等に限らず、他の公知の顔検出方法を使用した顔検出装置であってもよいことは勿論である。どのような顔検出部110に対しても、その後段に非顔判定部120を設けることで誤検出を低減することができる。

【0043】

非顔判定部120により非顔候補が排除された顔領域には、例えば顔であることを示すフラグ等を使用し、以降のフレームにおいては、スキントラッカ部114にてその顔領域のトラッキングを行うことができる。トラッキングは、例えば顔として検出された顔領域近傍の肌色領域を検出し、これをトラッキングするようにすればよい。また、スキントラッカ部114にてトラッキング中の顔領域は、所定のタイミングか、又は常に非顔判定部120に送られ非顔判定される。

【0044】

以下、非顔判定部120における具体的な処理方法について詳細に説明する。本実施の形態における非顔判定部120が実行する非顔判定の条件としては、例えば以下の4つの条件が挙げられる。

第1の条件：入力画像内に検出された顔候補の領域の大きさから算出した、実際の対象物(顔)までとの間の想定距離と、距離センサにて得られた測定距離との差が所定値以上である場合

第2の条件：入力画像内に検出された顔候補の領域が、色空間においてある一定以上分散していない場合

第3の条件：入力画像内に検出された顔候補の領域の大きさが、所定時間内に所定以上変化した場合

第4の条件：入力画像内に検出された顔候補の領域における肌色領域の占有率が一定以上大きい場合

【0045】

以上の条件に示すように、本実施の形態における非顔判定条件は、いずれも顔ではない

10

20

30

40

50

ものを抽出するための条件である。これらの条件を少なくとも 1 以上満たす場合を、誤検出とし、検出された領域（顔候補）から除くことで、本来は顔ではないにも拘わらず顔として検出されてしまったものを排除することができる。また、これら非顔判定のための処理は、上述のサポートベクタマシン SVM 等の演算処理に比して格段に演算量が少ないものであり、全ての条件判定を行っても短時間で判定終了可能であるため、ロボット装置等のリソースが限られた環境においても容易に実行可能なものである。従って、この非顔判定部 120 を設けることにより、前段の顔検出部 110 の顔検出においてある程度の誤検出を許容することができ、例えばサポートベクタマシン SVM 等の演算処理を減らすことができる。なお、以上の条件のうち複数を同時に満たす場合にのみ非顔であると判定してもよく、また認識精度や環境に応じて判定条件を増減させたり、判定条件に優先順位を設けたりしてもよい。

10

【0046】

以上の非顔判定処理（non-face 判定処理）は、プログラムコード形式で下記のように記載することができる。

【0047】

```
if( (condition1 == false) ||
    (condition2 == false) ||
    (condition3 == false) ||
    (condition4 == false) ){
    NONFACE ;
}
else{
    FACE;
}
```

20

【0048】

次に、上述した第 1 乃至第 4 の条件について説明する。先ず、第 1 の条件（condition1）については、先ず、測定距離センサ（position sensing device：P S D）により、対象となる人物までの間の距離が測定可能か否かを判定し、測定可能であると判断された場合に、測定距離 L2 を測定する。次に、入力画像内にて検出された顔候補となっている顔領域のサイズから想定される想定距離 L2 を算出する。例えば、人物の顔の大きさの平均値と、入力画像内の顔領域の大きさとに基づき想定距離 L2 を算出する。

30

【0049】

そして、想定距離 L1 / 測定距離 L2 = a（以下、第 1 の判定値という。）を算出し、第 1 の判定値 a が第 1 の閾値 th0 より大きいか、又は第 2 の閾値 th1 より小さい場合に、判定対象の入力画像の顔領域は偽（false）、すなわち、顔ではないとされ、第 1 の判定値 a が第 1 の閾値 th1 以上第 2 の閾値以下である場合に、判定対象の入力画像の顔領域は真（true）、すなわち、顔である判定する。そして、第 1 の判定値 a が第 1 の閾値 th0 より大きいか、又は第 2 の閾値 th1 より小さいという第 1 の条件を満たす顔候補（顔領域）が、この前段の顔検出部により検出された顔候補に含まれる非顔候補、すなわち誤検出として除かれる。

40

【0050】

具体的に、例えば、第 1 の閾値 th0 = 0.4、第 2 の閾値 th1 = 2.5 等とすることができ、

$2.5 < a = \text{想定距離 } L1 / \text{測定距離 } L2$ 、又は $0.4 > a = \text{想定距離 } L1 / \text{測定距離 } L2$

の場合、顔でないと判定される。

すなわち、 $0.4 < a = \text{想定距離 } L1 / \text{測定距離 } L2 < 2.5$

の条件を満たす顔候補以外の顔候補は、顔でない非顔候補と判定される。

【0051】

この第 1 の条件を満たすか否かの非顔判定方法は、プログラムコード形式で下記のよう

50

に記載することができる。

【 0 0 5 2 】

```
a = supposed_distance / measured_distance

if( is_measured_distance == true ){
    if( (a > th0) || (th1 > a) )
        condition_1 = false;
    else
        condition_1 = true;
}
else{
    condition_1 = true;
}
```

10

Where

```
is_measured_distance:= 顔領域の距離が測定可能かどうかを示す
supposed_distance:= 顔領域の画面上のサイズから計算される想定距離 L 1
measured_distance:= 測定距離センサから求められる距離 L 2
```

Example:

```
th0 = 2.5;
th1 = 0.4;
```

20

【 0 0 5 3 】

なお、閾値 th_0 、 th_1 は、入力画像を撮影するカメラや、測定距離センサの精度等に応じて適宜設定するものとする。

【 0 0 5 4 】

次に、第 2 の条件 (condition2) について説明する。人間の顔にはある程度、色の分散がある。すなわち、全面が肌の色一色の均一な領域ではなく、目や口等が含まれるためである。従って、顔領域においてある程度の色の分散がないものは顔ではないと判定することができる。色空間における分散は、YUV (YCbcCr) 又はRGB空間内において、1フレーム内の各成分の分散値の平均、及び / 又はその複数フレーム間の平均から求めることができる。

30

【 0 0 5 5 】

図 2 は、第 2 の条件を説明するための図であって、顔候補として検出されている顔領域内の画素 $P(i, j)$ を示す図である。図 2 において、水平方向（紙面左方向）を X、垂直方向（紙面下方向）を Y とする。

【 0 0 5 6 】

図 2 に示すように、入力画像 121 において、検出された顔候補 F_c に含まれる各画素 $P(i, j)$ を R 成分 ($R_{i,j}$)、G 成分 ($G_{i,j}$)、B 成分 ($B_{i,j}$) に分離する。そして、顔候補 F_c が占める顔領域における R 成分の分散 (Var_R)、G 成分の分散 (Var_G)、B 成分の分散 (Var_B) を算出する。

40

【 0 0 5 7 】

そして、これら RGB 成分の分散値の和 / 3 (第 2 の判定値) が第 3 の閾値 th_2 未満である場合に顔でないと判定する。すなわち、分散を示す第 2 の判定値が第 3 の閾値 th_2 以上ではない場合は、顔ではないと判断される。具体的に、第 3 の閾値 th_2 は、例えば 450 等と設定することができる。

【 0 0 5 8 】

ここで、本実施の形態においては、現フレームから求めた R、G、B の各成分の分散値の平均 (第 1 の平均値) が上記第 3 の閾値 th_2 以上ない場合に顔ではないと判定すると共に、時間軸上で連続する複数フレームの分散値の平均 (第 2 の平均値) が第 4 の閾値 th_3 以上ない場合に顔ではないと判定することもできる。この複数フレームの分散値の平

50

均である第2の平均値は、上記第1の平均値を加算してフレーム数で割ったものを使用すればよい。

【0059】

このように複数フレーム間での第2の平均値を使用して非顔判定することで、1フレームの判定のみよりも誤差を低減し、更に正確に非顔候補であるものを排除することができる。また、このように、1フレーム内の分散値の平均（第1の平均値）に合わせて複数フレーム間の分散値の平均（第2の平均値）を使用する場合、第1の平均値の閾値として使用する第3の閾値 th_2 を、第2の平均値の閾値として使用する第4の閾値 th_3 より大きくし、その判定基準を厳しくすることで更に正確に非顔検出することができる。

【0060】

この第2の条件を満たすか否かの非顔判定方法のうち、1フレームにおける非顔判定方法は、プログラムコード形式で下記のように記載することができる。

【0061】

```
if( ((Var_R + Var_G + Var_B))/3 < th2 ){
    condition_2 = false;
}
else
    condition_2 = true;
}
```

10

20

Where

```
Var_R = <顔領域のR成分の分散>;
Var_G = <顔領域のG成分の分散>;
Var_B = <顔領域のB成分の分散>;
```

Example:

```
th2 = 450;
```

【0062】

次に、第3の条件（condition3）について説明する。人間がある一定時間内に移動できる距離には限界がある。そこで、一定時間内に顔領域の大きさの変化率が所定値以上であった場合には顔ではないと判断することができる。

30

【0063】

この第3の条件は、スキントラッカ部114にて顔領域を追跡（追尾）している場合に使用することができる。スキントラッカ部114は、非顔判定部120により非顔候補が除かれ、顔と判断された顔領域を検出すると、その肌色領域を検出し、例えば、肌色領域の重心が入力画像の中心にくるようにして、肌色領域がカメラの視野内に入るよう追跡する。また、顔検出部110が複数フレームを処理して顔検出するようなものではある場合には、顔検出部110の結果に本第3の条件（condition3）を適用することもできる。

【0064】

この顔領域の大きさ判定の方法としては、例えば、スキントラッカ部1114がトラッキングしている現在のフレームにおける顔領域のサイズ S_{f1} と、比較対象となる所定時間前のフレームにおける顔領域のサイズ S_{f2} との差を求め、これを上記所定時間で除した値と第5の閾値 th_4 とを比較することにより行うことができる。そして、この所定時間の顔領域の変化が第5の閾値 th_4 より大きい場合は、顔でないと判断することができる。

40

【0065】

すなわち、所定時間、例えば6フレーム前に検出された顔領域の大きさと、現在の顔領域の大きさとの差が大きければ、人間はそのように不自然な動きをしないものとの仮定に基づき、顔ではないと判断することができる。

【0066】

また、顔検出部110が複数フレームから顔候補を検出するものである場合には、顔検

50

出部 110 から比較対象となる顔領域サイズ Sf1、Sf2 を非顔判定部 120 に供給して非顔判定するようにしてもよい。

【0067】

この第 3 の条件を満たすか否かの非顔判定方法は、プログラムコード形式で下記のように記載することができる。

【0068】

```
if( (size_of_face - previous_size_of_face)/frame_interval > th3 ){
    condition_3 = false;
}
else{
    condition_3 = true;
}
```

10

Where:

size_of_face := 現在の顔領域のサイズ(in pixel)
 previous_size_of_face := 前フレームの顔領域のサイズ(in pixel)
 frame_interval := 判定に使うframe間隔(Ex. 6 frame)

Example:

th3= 5;

20

【0069】

次に、第 4 の条件 (condition4) について説明する。スキントラッカ部 114 は、顔領域を検出すると、例えばその肌色領域を検出してトラッキングを行うことができる。この第 4 の条件は、このスキントラッカ部 114 のように肌色領域を検出する場合に使用することができる。また、顔検出部 110 にて肌色領域を検出し、検出された肌色領域から顔候補を認識するような場合には、この第 4 の条件にて非顔判定することができる。

【0070】

第 2 の条件において説明したように、人間の顔には目や口が含まれているため、100% 肌色ではない。そこで、顔領域にある一定以上の割合の肌色領域が含まれる場合は、これを顔でないと判定することができる。

30

【0071】

具体的には、顔検出部 110 にて顔候補とされ、非顔検出部 120 にて非顔候補とされず、真の顔領域として検出された領域の大きさと、この顔領域のトラッキングを行うスキントラッカ部 114 にて検出した肌色領域の大きさとから、顔領域の画素内の肌色占有率 Cf を求める。そして、この肌色占有率 Cf が例えば Cf 95 % であるとき、その顔領域は、例えば肌色に近い色の壁や、肌色であっても顔ではない腕や足等であって、顔ではないと判断することができる。

【0072】

ここで、上述のスキントラッカ部 114 は、顔検出部 110 により顔検出され、非顔判定部 120 により非顔候補が排除された顔領域の肌色領域をトラッキングするものであるが、非顔検出部 120 は、肌色占有率が例えば 95 % より小さいものは非顔と判定せず、この顔候補が他の条件にも当てはまらない場合は真の顔領域として検出される。このような場合、スキントラッカ部 114 が肌色領域であれば全てトラッキングするようにした場合には、非顔検出部 120 において、肌色占有率が例えば 70 % より大きく、且つ 95 % 未満であるものを顔候補とし、肌色占有率が 70 % 以下、95 % 以上であるものを非顔候補として判定するようにしてもよい。または、スキントラッカ部 114 にて顔領域と肌色領域とから肌色占有率を測定し、肌色占有率が 70 % 以上である場合にのみトラッキングするようにしてもよい。また、例えば顔検出部 110 が、肌色領域を抽出し、肌色領域から顔候補を検出するような検出部からなる場合であって、肌色占有率が 70 % 以上であることを検出条件に含むような場合は、一度非顔判定され、スキントラッカ部 114 が肌色

40

50

領域のトラッキングを開始した以降に、肌色占有率が 70 % 以下、 95 % 以上であるものを非顔候補として判定するようにしてもよい。

【 0 0 7 3 】

この第 4 の条件を満たすか否かの非顔判定方法は、プログラムコード形式で下記のように記載することができる。

【 0 0 7 4 】

```
if( (current_skin_rate > th4) ){
    condition_4 = false;
}
else{
    condition_4 = true;
}
```

10

Where

current_flesh_rate:=顔領域の画素内の肌色占有率 (%)

Example:

th4 = 95;

【 0 0 7 5 】

以上説明した 4 つの条件の少なくとも 1 つを満たした場合には、顔検出部 110 にて検出した顔領域（顔候補）は、顔ではない非顔候補と判断し、顔候補から除外することができる。ここで、上述したように、これら 4 つの条件は、計算処理的には顔検出部に比して格段に計算量が少なく、顔検出部 110 の検出条件を緩めてある程度誤検出を多くしたとしても、これら 4 つの条件全てを満たすか否かの非顔判定部 120 の処理を実行することで、結果的に顔検出の誤検出を低減し、且つ演算処理量を増大させることがない。

20

【 0 0 7 6 】

また、本実施の形態においては、非顔検出部 120 により顔検出部 110 の検出結果である顔候補の非顔判定をして非顔候補を排除し、またスキントラッカ部 114 のトラッキング対象となっている領域の非顔判定をして非顔候補を排除することにより、誤検出を取り除くものであるが、顔検出部 110 は、テンプレートマッチングや SVM を使用した顔検出（顔認識）ではなく他の認識方法であってもよく、例えば肌色領域を抽出しておき、更に、目、鼻、口等の特徴点を検出し、その特徴量や位置関係等に基づき顔候補を検出してもよい。また、顔検出部 110 及びスキントラッカ部 114 の何れか一方と非顔検出部 120 とを組み合わせた顔検出装置としてもよい。

30

【 0 0 7 7 】

更に、ここでは、検出された顔候補から顔ではない非顔候補を検出するものとしたが、非対象物判定部を設けることで、検出対象が顔ではない物体であっても、同様の効果を奏する。特に、顔検出のように、対象物として複雑な物体を検出するような場合に、上述の顔検出部に相当する対象物検出部の後段に、上述の非顔検出部に相当する非対象物検出部を設けることにより、演算処理を増加させることなく、簡単且つ高速に対象物検出の誤検出を排除することができる。なお、例えばボール等、検出対象として単純な物体を検出するような場合には、このような非対象物検出部を設けて、非対象物を排除する処理を設けるよりも、対象物を検出する検出処理を高精度に行つた方が効率的となる。

40

【 0 0 7 8 】

(2) 口ボット装置

次に、上述した本実施の形態における顔検出装置、すなわち従来の顔検出部に加えて非顔判定部を有する顔検出装置を搭載するに好適な口ボット装置について説明する。上述した如く、口ボット装置は、リソースが限られており、またエンターテイメント用とされる場合は、顔検出のみならず、ユーザと対話したり、ユーザの指示に従ったり同時に進行べき処理が膨大である。このような口ボット装置には、上述したような計算量が少なく、且

50

つ誤検出が少ない顔検出装置を搭載することが好ましい。

【0079】

(2-1) ロボット装置のハードウェア構成

図3に示すように、本実施の形態によるロボット装置1は、4足歩行の脚式移動ロボット装置とされ、胴体部ユニット2の前後左右にそれぞれ脚部ユニット3A, 3B, 3C, 3Dが連結されると共に、胴体部ユニット2の前端部に頭部ユニット4が連結されて構成されている。

【0080】

胴体部ユニット2には、図4に示すように、CPU(Central Processing Unit)10、DRAM(Dynamic Random Access Memory)11、フラッシュROM(Read Only Memory)12、PC(Personal Computer)カードインターフェース回路13及び信号処理回路14が内部バス15を介して相互に接続されることにより形成されたコントロール部16と、このロボット装置1の動力源としてのバッテリ17とが収納されている。また、胴体部ユニット2には、ロボット装置1の向きや動きの加速度を検出するための角速度センサ18及び加速度センサ19なども収納されている。10

【0081】

また、頭部ユニット4には、外部の状況を撮像するとともに、周囲の明るさを検出するためのCCD(Charge Coupled Device)カメラ20と、前後に倒すといった使用者からの物理的な働きかけを検出するためのタッチセンサ21と、前方に位置する物体までの距離を測定するための距離センサ22と、外部音を集音するためのマイクロホン23と、各種音声を出力するためのスピーカ24と、頭部ユニット4内に格納可能とされたヘッドライト25と、ロボット装置1の「目」に相当するLED(Light Emitting Diode)(図示せず)などがそれぞれ所定位置に配置されている。なお、ロボット装置1においては、タッチセンサ21以外にも、胴体部ユニット2及び頭部ユニット4の所定位置に複数のタッチセンサが配置されている。20

【0082】

さらに、各脚部ユニット3A～3Dの関節部分や各脚部ユニット3A～3D及び胴体部ユニット2の各連結部分、並びに頭部ユニット4及び胴体部ユニット2の連結部分などにはそれぞれ自由度数分のアクチュエータ26₁～26_n及びポテンショメータ27₁～27_nが配設されている。例えば、アクチュエータ26₁～26_nはサーボモータを構成として有している。サーボモータの駆動により、脚部ユニット3A～3Dが制御されて、目標の姿勢或いは動作に遷移する。30

【0083】

そして、これら角速度センサ18、加速度センサ19、タッチセンサ21、距離センサ22、マイクロホン23、スピーカ24、及び各ポテンショメータ27₁～27_nなどの各種センサ並びにヘッドライト25、LED及び各アクチュエータ26₁～26_nは、それぞれ対応するハブ28₁～28_nを介してコントロール部16の信号処理回路14と接続され、CCDカメラ20及びバッテリ17は、それぞれ信号処理回路14と直接接続されている。

【0084】

信号処理回路14は、上述の各センサから供給されるセンサデータや画像データ及び音声データを順次取り込み、これらをそれぞれ内部バス15を介してDRAM11内の所定位置に順次格納する。また信号処理回路14は、これと共にバッテリ17から供給されるバッテリ残量を表すバッテリ残量データを順次取り込み、これをDRAM11内の所定位置に格納する。40

【0085】

このようにしてDRAM11に格納された各センサデータ、画像データ、音声データ及びバッテリ残量データは、この後CPU10がこのロボット装置1の動作制御を行う際に利用される。

【0086】

実際上 C P U 1 0 は、ロボット装置 1 の電源が投入された初期時、胴体部ユニット 2 の図示しない P C カードスロットに装填されたメモリカード 2 9 又はフラッシュ R O M 1 2 に格納された制御プログラムを P C カードインターフェース回路 1 3 を介して又は直接読み出し、これを D R A M 1 1 に格納する。

【 0 0 8 7 】

また、C P U 1 0 は、この後上述のように信号処理回路 1 4 より D R A M 1 1 に順次格納される各センサデータ、画像データ、音声データ及びバッテリ残量データに基づいて自己及び周囲の状況や、使用者からの指示及び働きかけの有無などを判断する。

【 0 0 8 8 】

ここで、例えば本実施の形態における顔検出タスクを実行する際には、C C D により撮像された入力画像から顔候補を検出し、距離センサ等の値を使用し、上述した4つの条件に当てはめて非顔候補を検出し、顔候補から非顔候補を排除したものを顔領域として出力する。10

【 0 0 8 9 】

さらに、C P U 1 0 は、この判断結果及びD R A M 1 1 に格納した制御プログラムに基づいて続く行動を決定すると共に、当該決定結果に基づいて必要なアクチュエータ 2 6₁ ~ 2 6_n を駆動させることにより、頭部ユニット 4 を上下左右に振らせたり、各脚部ユニット 3 A ~ 3 D を駆動させて歩行させるなどの行動を行わせる。

【 0 0 9 0 】

また、この際 C P U 1 0 は、必要に応じて音声データを生成し、これを信号処理回路 1 4 を介して音声信号としてスピーカ 2 4 に与えることにより当該音声信号に基づく音声を外部に出力させたり、上述の L E D を点灯、消灯又は点滅させる。また、C P U 1 0 は、後述するように、C C D カメラ 2 0 を用いて周囲の明るさを検出させ、その検出結果に応じてヘッドライト 2 5 を点灯させる。20

【 0 0 9 1 】

このようにしてこのロボット装置 1 においては、自己及び周囲の状況や、使用者からの指示及び働きかけに応じて自律的に行動し得るようになされている。

【 0 0 9 2 】

(2 - 2) 制御プログラムのソフトウェア構成

ここで、ロボット装置 1 における上述の制御プログラムのソフトウェア構成は、図 5 に示すようになる。この図 5 において、デバイス・ドライバ・レイヤ 3 0 は、この制御プログラムの最下位層に位置し、複数のデバイス・ドライバからなるデバイス・ドライバ・セット 3 1 から構成されている。この場合、各デバイス・ドライバは、C C D カメラ 2 0 (図 4) やタイマ等の通常のコンピュータで用いられるハードウェアに直接アクセスすることを許されたオブジェクトであり、対応するハードウェアからの割り込みを受けて処理を行う。30

【 0 0 9 3 】

また、ロボティック・サーバ・オブジェクト 3 2 は、デバイス・ドライバ・レイヤ 3 0 の最下位層に位置し、例えば上述の各種センサやアクチュエータ 2 6₁ ~ 2 6_n 等のハードウェアにアクセスするためのインターフェースを提供するソフトウェア群でなるパーティカル・ロボット装置 3 3 と、電源の切換えなどを管理するソフトウェア群でなるパワーマネージャ 3 4 と、他の種々のデバイス・ドライバを管理するソフトウェア群でなるデバイス・ドライバ・マネージャ 3 5 と、ロボット装置 1 の機構を管理するソフトウェア群でなるデザインド・ロボット装置 3 6 とから構成されている。40

【 0 0 9 4 】

マネージャ・オブジェクト 3 7 は、オブジェクト・マネージャ 3 8 及びサービス・マネージャ 3 9 から構成されている。オブジェクト・マネージャ 3 8 は、ロボティック・サーバ・オブジェクト 3 2 、ミドル・ウェア・レイヤ 4 0 、及びアプリケーション・レイヤ 4 1 に含まれる各ソフトウェア群の起動や終了を管理するソフトウェア群であり、サービス・マネージャ 3 9 は、メモリカード 2 9 (図 4) に格納されたコネクションファイルに記

10

20

30

40

50

述されている各オブジェクト間の接続情報に基づいて各オブジェクトの接続を管理するソフトウェア群である。

【0095】

ミドル・ウェア・レイヤ40は、ロボティック・サーバ・オブジェクト32の上位層に位置し、画像処理や音声処理などのこのロボット装置1の基本的な機能を提供するソフトウェア群から構成されている。また、アプリケーション・レイヤ41は、ミドル・ウェア・レイヤ40の上位層に位置し、当該ミドル・ウェア・レイヤ40を構成する各ソフトウェア群によって処理された処理結果に基づいてロボット装置1の行動を決定するためのソフトウェア群から構成されている。

【0096】

なお、ミドル・ウェア・レイヤ40及びアプリケーション・レイヤ41の具体的なソフトウェア構成をそれぞれ図6に示す。

【0097】

ミドル・ウェア・レイヤ40は、図6に示すように、騒音検出用、温度検出用、明るさ検出用、音階認識用、距離検出用、姿勢検出用、タッチセンサ用、動き検出用及び色認識用の各信号処理モジュール50～58並びに入力セマンティクスコンバータモジュール59などを有する認識系60と、出力セマンティクスコンバータモジュール68並びに姿勢管理用、トラッキング用、モーション再生用、歩行用、転倒復帰用、ライト点灯用及び音再生用の各信号処理モジュール61～67などを有する出力系69とから構成されている。

【0098】

認識系60の各信号処理モジュール50～58は、ロボティック・サーバ・オブジェクト32のバーチャル・ロボット装置33によりDRAM11(図4)から読み出される各センサデータや画像データ及び音声データのうちの対応するデータを取り込み、当該データに基づいて所定の処理を施して、処理結果を入力セマンティクスコンバータモジュール59に与える。ここで、例えば、バーチャル・ロボット装置33は、所定の通信規約によって、信号の授受或いは変換をする部分として構成されている。

【0099】

入力セマンティクスコンバータモジュール59は、これら各信号処理モジュール50～58から与えられる処理結果に基づいて、「うるさい」、「暑い」、「明るい」、「ボールを検出した」、「転倒を検出した」、「撫でられた」、「叩かれた」、「ドミソの音階が聞こえた」、「動く物体を検出した」又は「障害物を検出した」などの自己及び周囲の状況や、使用者からの指令及び働きかけを認識し、認識結果をアプリケーション・レイヤ41(図4)に出力する。

【0100】

アプリケーション・レイヤ41は、図7に示すように、行動モデルライブラリ70、行動切換モジュール71、学習モジュール72、感情モデル73及び本能モデル74の5つのモジュールから構成されている。

【0101】

行動モデルライブラリ70には、図8に示すように、「バッテリ残量が少なくなった場合」、「転倒復帰する」、「障害物を回避する場合」、「感情を表現する場合」、「ボールを検出した場合」などの予め選択された複数の条件項目にそれぞれ対応させて、それぞれ独立した行動モデル70₁～70_nが設けられている。

【0102】

そして、これら行動モデル70₁～70_nは、それぞれ入力セマンティクスコンバータモジュール59から認識結果が与えられたときや、最後の認識結果が与えられてから一定時間が経過したときなどに、必要に応じて後述のように感情モデル73に保持されている対応する情動のパラメータ値や、本能モデル74に保持されている対応する欲求のパラメータ値を参照しながら続く行動をそれぞれ決定し、決定結果を行動切換モジュール71に出力する。

10

20

30

40

50

【0103】

なお、この実施の形態の場合、各行動モデル $70_1 \sim 70_n$ は、次の行動を決定する手法として、図9に示すような1つのノード（状態）NODE₀～NODE_nから他のどのノードNODE₀～NODE_nに遷移するかを各ノードNODE₀～NODE_nに間を接続するアーケARC₁～ARC_nに対してそれぞれ設定された遷移確率P₁～P_nに基づいて確率的に決定する有限確率オートマトンと呼ばれるアルゴリズムを用いる。

【0104】

具体的に、各行動モデル $70_1 \sim 70_n$ は、それぞれ自己の行動モデル $70_1 \sim 70_n$ を形成するノードNODE₀～NODE_nにそれぞれ対応させて、これらノードNODE₀～NODE_nごとに図10に示すような状態遷移表80を有している。

10

【0105】

この状態遷移表80では、そのノードNODE₀～NODE_nにおいて遷移条件とする入力イベント（認識結果）が「入力イベント名」の列に優先順に列記され、その遷移条件についてのさらなる条件が「データ名」及び「データ範囲」の列における対応する行に記述されている。

【0106】

したがって、図10の状態遷移表80で表されるノードNODE₁₀₀では、「ボールを検出（BALL）」という認識結果が与えられた場合に、当該認識結果と共に与えられるそのボールの「大きさ（SIZE）」が「0から1000」の範囲であることや、「障害物を検出（OBSTACLE）」という認識結果が与えられた場合に、当該認識結果と共に与えられるその障害物までの「距離（DISTANCE）」が「0から100」の範囲であることが他のノードに遷移するための条件となっている。

20

【0107】

また、このノードNODE₁₀₀では、認識結果の入力がない場合においても、行動モデル $70_1 \sim 70_n$ が周期的に参照する感情モデル73及び本能モデル74にそれぞれ保持された各情動及び各欲求のパラメータ値のうち、感情モデル73に保持された「喜び（JOY）」、「驚き（SURPRISE）」若しくは「悲しみ（SUDNESS）」のいずれかのパラメータ値が「50から100」の範囲であるときには他のノードに遷移することができるようになっている。

30

【0108】

また、状態遷移表80では、「他のノードへの遷移確率」の欄における「遷移先ノード」の行にそのノードNODE₀～NODE_nから遷移できるノード名が列記されていると共に、「入力イベント名」、「データ値」及び「データの範囲」の列に記述された全ての条件が揃ったときに遷移できる他の各ノードNODE₀～NODE_nへの遷移確率が「他のノードへの遷移確率」の欄内の対応する箇所にそれぞれ記述され、そのノードNODE₀～NODE_nに遷移する際に出力すべき行動が「他のノードへの遷移確率」の欄における「出力行動」の行に記述されている。なお、「他のノードへの遷移確率」の欄における各行の確率の和は100[%]となっている。

【0109】

したがって、図9の状態遷移表80で表されるノードNODE₁₀₀では、例えば「ボールを検出（BALL）」し、そのボールの「SIZE（大きさ）」が「0から1000」の範囲であるという認識結果が与えられた場合には、「30[%]」の確率で「ノードNODE₁₂₀（node 120）」に遷移でき、そのとき「ACTION1」の行動が出力されることとなる。

40

【0110】

各行動モデル $70_1 \sim 70_n$ は、それぞれこのように状態遷移表80として記述されたノードNODE₀～NODE_nがいくつも繋がるようにして構成されており、入力セマンティクスコンバータモジュール59から認識結果が与えられたときなどに、対応するノードNODE₀～NODE_nの状態遷移表を利用して確率的に次の行動を決定し、決定結果を行動切換モジュール71に出力するようになされている。

50

【0111】

図7に示す行動切換モジュール71は、行動モデルライブラリ70の各行動モデル70₁～70_nからそれぞれ出力される行動のうち、予め定められた優先順位の高い行動モデル70₁～70_nから出力された行動を選択し、当該行動を実行すべき旨のコマンド（以下、これを行動コマンドという。）をミドル・ウェア・レイヤ40の出力セマンティクスコンバータモジュール68に送出する。なお、この実施の形態においては、図8において下側に表記された行動モデル70₁～70_nほど優先順位が高く設定されている。

【0112】

また、行動切換モジュール71は、行動完了後に¹⁰出力セマンティクスコンバータモジュール68から与えられる行動完了情報に基づいて、その行動が完了したことを学習モジュール72、感情モデル73及び本能モデル74に通知する。

【0113】

一方、学習モジュール72は、入力セマンティクスコンバータモジュール59から与えられる認識結果のうち、「叩かれた」や「撫でられた」など、使用者からの働きかけとして受けた教示の認識結果を入力する。

【0114】

そして、学習モジュール72は、この認識結果及び行動切換モジュール71からの通知に基づいて、「叩かれた（叱られた）」ときにはその行動の発現確率を低下させ、「撫でられた（誉められた）」ときにはその行動の発現確率を上昇させるように、行動モデルライブラリ70における対応する行動モデル70₁～70_n²⁰の対応する遷移確率を変更する。
。

【0115】

他方、感情モデル73は、「喜び（joy）」、「悲しみ（sadness）」、「怒り（anger）」、「驚き（surprise）」、「嫌悪（disgust）」及び「恐れ（fear）」の合計6つの情動について、各情動ごとにその情動の強さを表すパラメータを保持している。そして、感情モデル73は、これら各情動のパラメータ値を、それぞれ入力セマンティクスコンバータモジュール59から与えられる「叩かれた」と「撫でられた」などの特定の認識結果と、経過時間及び行動切換モジュール71からの通知などに基づいて周期的に更新する。
。

【0116】

具体的には、感情モデル73は、入力セマンティクスコンバータモジュール59から与えられる認識結果と、そのときのロボット装置1の行動と、前回更新してからの経過時間などに基づいて所定の演算式により算出されるそのときのその情動の変動量をE[t]、現在のその情動のパラメータ値をE[t]、その情動の感度を表す係数をk_eとして、(1)式によって次の周期におけるその情動のパラメータ値E[t+1]を算出し、これを現在のその情動のパラメータ値E[t]と置き換えるようにしてその情動のパラメータ値を更新する。また、感情モデル73は、これと同様にして全ての情動のパラメータ値を更新する。

【0117】

【数1】

$$E[t+1] = E[t] + k_e \times \Delta E[t] \quad \cdots (1)$$

【0118】

なお、各認識結果や出力セマンティクスコンバータモジュール68からの通知が各情動のパラメータ値の変動量E[t]にどの程度の影響を与えるかは予め決められており、例えば「叩かれた」といった認識結果は「怒り」の情動のパラメータ値の変動量E[t]に大きな影響を与え、「撫でられた」といった認識結果は「喜び」の情動のパラメータ値の変動量E[t]に大きな影響を与えるようになっている。

【0119】

10

20

30

40

50

ここで、出力セマンティクスコンバータモジュール68からの通知とは、いわゆる行動のフィードバック情報（行動完了情報）であり、行動の出現結果の情報であり、感情モデル73は、このような情報によっても感情を変化させる。これは、例えば、「吠える」といった行動により怒りの感情レベルが下がるといったようなことである。なお、出力セマンティクスコンバータモジュール68からの通知は、上述した学習モジュール72にも入力されており、学習モジュール72は、その通知に基づいて行動モデル70₁～70_nの対応する遷移確率を変更する。

【0120】

なお、行動結果のフィードバックは、行動切換モジュール71の出力（感情が付加された行動）によりなされるものであってもよい。

10

【0121】

一方、本能モデル74は、「運動欲(exercise)」、「愛情欲(affection)」、「食欲(appetite)」及び「好奇心(curiosity)」の互いに独立した4つの欲求について、これら欲求ごとにその欲求の強さを表すパラメータを保持している。そして、本能モデル74は、これらの欲求のパラメータ値を、それぞれ入力セマンティクスコンバータモジュール59から与えられる認識結果や、経過時間及び行動切換モジュール71からの通知などに基づいて周期的に更新する。

【0122】

具体的には、本能モデル74は、「運動欲」、「愛情欲」及び「好奇心」については、認識結果、経過時間及び出力セマンティクスコンバータモジュール68からの通知などに基づいて所定の演算式により算出されるそのときのその欲求の変動量をI[k]、現在のその欲求のパラメータ値をI[k]、その欲求の感度を表す係数k_iとして、所定周期で(2)式を用いて次の周期におけるその欲求のパラメータ値I[k+1]を算出し、この演算結果を現在のその欲求のパラメータ値I[k]と置き換えるようにしてその欲求のパラメータ値を更新する。また、本能モデル74は、これと同様にして「食欲」を除く各欲求のパラメータ値を更新する。

20

【0123】

【数2】

$$I[k+1] = I[k] + k_i \times \Delta I[k]$$

・・・(2)

30

【0124】

なお、認識結果及び出力セマンティクスコンバータモジュール68からの通知などが各欲求のパラメータ値の変動量I[k]にどの程度の影響を与えるかは予め決められており、例えば出力セマンティクスコンバータモジュール68からの通知は、「疲れ」のパラメータ値の変動量I[k]に大きな影響を与えるようになっている。

【0125】

なお、本実施の形態においては、各情動及び各欲求（本能）のパラメータ値がそれぞれ0から100までの範囲で変動するように規制されており、また係数k_e、k_iの値も各情動及び各欲求ごとに個別に設定されている。

40

【0126】

一方、ミドル・ウェア・レイヤ40の出力セマンティクスコンバータモジュール68は、図6に示すように、上述のようにしてアプリケーション・レイヤ41の行動切換モジュール71から与えられる「前進」、「喜ぶ」、「鳴く」又は「トラッキング（ボールを追いかける）」といった抽象的な行動コマンドを出力系69の対応する信号処理モジュール61～67に与える。

【0127】

そしてこれら信号処理モジュール61～67は、行動コマンドが与えられると当該行動コマンドに基づいて、その行動を行うために対応するアクチュエータ26₁～26_n（図

50

4)に与えるべきサーボ指令値や、スピーカ24(図4)から出力する音の音声データ及び又は「目」のLEDに与える駆動データを生成し、これらのデータをロボティック・サーバ・オブジェクト32のバーチャル・ロボット装置33及び信号処理回路14(図4)を順次介して対応するアクチュエータ26₁～26_n又はスピーカ24又はLEDに順次送出する。

【0128】

このようにしてロボット装置1においては、制御プログラムに基づいて、自己(内部)及び周囲(外部)の状況や、使用者からの指示及び働きかけに応じた自律的な行動を行うことができるようになされている。このような自律的動作が可能なロボット装置に上述した誤検出を低減した顔検出装置を搭載することで、リアルタイムに正確な顔検出を可能とし、ユーザとの円滑なコミュニケーションを実現することができる。10

【0129】

(3)顔検出部の具体例

以下、上述した本発明の実施の形態における顔検出装置における非顔判定部120の前段にて顔候補を抽出するための顔検出部110の一具体例について説明する。この顔検出部は、ロボット装置等が撮像した入力画像に対して、テンプレートマッチングを行っておまかに顔候補を抽出することでこの後段におけるSVM等の演算量を削減しつつ、顔候補の中から顔であるか否かの顔検出を行うことにより、ロボット装置の顔検出のリアルタイム性を向上させたものである(特願2003-133601号参照)。

【0130】

上述したように、テンプレートマッチングを行って顔候補を抽出し(第1の工程)、この顔候補の中からSVM等により顔領域を判定して(第2の工程)顔領域を検出する方法において、第1の工程においては、単純に正規化相関値の代償により顔候補を決定すると、顔候補の見逃しを軽減しようとした場合、閾値を上げる方法又は間引きを減らす方法をとることができるもの、閾値を下げると演算量が増大してしまい、ロボット装置等のリソースの限られた環境においては好ましくない場合がある。一方、閾値を上げると、第2の工程において顔判定するための候補画像が減るため、演算量を減らすができるものの、本来顔である画像も候補画像から取り除いてしまい、顔画像を見逃してしまう場合がある。

【0131】

そこで、本願発明者等は、テンプレートと同一サイズの顔領域(顔画像)が入力画像内に存在する場合、この顔画像とテンプレートとの相関をとれば、テンプレートサイズ近傍では最も相関値が大きくなることに着目し、顔領域の候補を絞り込む際に、局所的な絞り込みを行うアルゴリズムを使用することにより、本来顔である画像を見逃すことなく顔候補画像を低減して後段の第2の工程にて顔判定する計算量を低減する方法を見出した。具体的には、入力画像と所定サイズの平均顔のテンプレートとの正規化相関をとった相関値の集合であるマッチング結果における相関値の局所最大値に基づき候補となる顔領域を抽出するものである。以下、この顔検出方法について詳細に説明する。30

【0132】

本実施の形態における顔検出装置は、例えば上述した動物を模したロボット装置1等のCPU及びメモリ等のリソースが限られたシステムにおいて、性能を保持しつつ顔検出に要する計算量を削減する場合に好適である。40

【0133】

(3-1)顔検出タスク機能に関する信号処理回路14の処理

本実施の形態におけるロボット装置1に搭載された顔検出タスク機能について説明する。上述した如く、ロボット装置1には、CCDカメラ20を介して内部メモリ(DRAM)11に記憶されるフレーム画像について、当該フレーム画像の中から人間の顔画像を検出する顔検出タスク機能が搭載されている。そしてこの顔検出タスク機能は、主制御部となる信号処理回路14における各種処理により実現されている。

【0134】

10

20

30

40

50

ここで、かかる顔検出タスク機能に関する信号処理回路 14 の処理内容を機能的に分類すると、図 11 に示すように、テンプレートサイズ及び入力画像スケール変換決定部 90 、ウィンドウ切り出し部 91 、テンプレートマッチング部 92 、スケール変換及び切り出し部 93 、前処理部 94 、及びパターン識別部 95 に分けることができる。

【0135】

入力画像スケール変換決定部 90 は、図 4 に示す CCD カメラ 20 からの画像信号に基づくフレーム画像をロボット装置等の内部メモリ (DRAM 11) から読み出して、当該フレーム画像を縮小率が相異なる複数のスケール画像（第 1 ~ 第 5 のスケール画像）に変換すると共に、後段のテンプレートマッチング部 92 で使用するテンプレートサイズ、すなわち、顔画像のサイズを選択する（以下、これを第 1 のテンプレートサイズという）。 10

【0136】

続くウィンドウ切出部 91 は、第 1 ~ 第 5 のスケール画像のうち、まず第 1 のスケール画像に対して、画像左上を起点として順に画像右下まで、適当な画素（例えば 2 画素）分を飛ばしながらスキャンするようにして、 $400 (= 20 \times 20)$ 画素の矩形領域でなるウィンドウ画像を順次切り出す。 20

【0137】

その際、ウィンドウ切出部 91 は、第 1 のスケール画像から切り出した複数のウィンドウ画像のうち先頭のウィンドウ画像を、スケール変換決定部 91 で選択した第 1 のテンプレートサイズのテンプレートと共に後段のテンプレートマッチング部 92 に送出する。 30

【0138】

テンプレートマッチング部 92 は、ウィンドウ切出部 91 から得られた先頭のウィンドウ画像について、第 1 のテンプレートサイズを有する平均的な顔画像をテンプレートとして大まかなマッチングをとり、当該テンプレートとの相関値（テンプレートのうちマッチングする画素数の割合）の集合をマッチング結果として当該ウィンドウ画像と共に DRAM 11 に格納する。 20

【0139】

すなわち、図 12 (a) に示すように、任意のスケール画像から切り出した、例えば高さ (y 軸方向の辺の長さ) hei_s × 幅 (x 軸方向の辺の長さ) wid_s のウィンドウ画像（スケール変換後の入力画像）W 2 について、図 12 (b) に示すように、例えば高さ hei_t × 幅 wid_s である第 1 のテンプレートサイズを有する平均的な顔画像であるテンプレート T 2 を使用し、ウィンドウ画像 W 2 をスキャンし、所定画素（例えば 1 画素）ずつずらしながら移動させたテンプレート T 2 と上記入力画像との相関値の集合であるマッチング結果を求める。このマッチング結果は、テンプレート T 2 の移動に伴い相関値が 2 次元に配列されたものであり、図 13 に示すように、当該相関値を表す高さ hei_r × 幅 wid_r のテンプレートマッチング結果画像 R 2 が得られる。ここで、テンプレートトレートマッチング結果画像 R 2 の高さ hei_r は、hei_s - (hei_t + 1) であり、画像 R 2 の幅 wid_s は、wid_s - (wid_t + 1) となる。 30

【0140】

次に、このテンプレートトレートマッチング結果画像 R 2 を所定のサイズ、例えば第 1 のテンプレートサイズと同一の大きさに分割し、各第 1 のテンプレートサイズに仕切られた分割領域毎に相関値の最大値を有する点（位置）を求め、これら各分割領域から得られた最大値を示す点のうち、所定の閾値以上のものを顔候補として抽出する。 40

【0141】

すなわち、平均顔のテンプレートを使用して正規化相関をしようした場合、必ずしも任意のパターンより、顔画像の方が相関値が高くなるという保証はないものの、テンプレートと同一のサイズの顔画像が存在する場合は、テンプレートサイズ近傍の大きさでは相関値が最大値をとることから、相関値が分割領域内で最大値となり、且つ所定の閾値以上の点を顔候補として抽出することにより、単にテンプレートマッチングの結果、相関値が所定の閾値以上であるものを顔候補として抽出する場合に比して、顔候補をより有効に絞り込むことができる。 50

【 0 1 4 2 】

次に、この顔候補として抽出された点を顔候補（第1の顔候補）と共に、この顔候補近傍の点も顔候補（第2の顔候補）として抽出し、後段のパターン認識部95において、SVMを使用して顔判定する。ここで、SVMは1画素ずれるのみで顔検出ができないほど敏感であるため、本実施の形態においては、顔候補の近傍も検索範囲（第2の顔候補）とすることにより、顔候補の近傍を重点的に検索して顔検出性能を向上することができる。ここで第1の顔候補として抽出された点の上下左右に隣接する8点全てを検索範囲とすると、後段の計算量が増加するため、所定の条件を満たす場合にのみ、検索範囲に指定することにより、後段の計算量の増加を抑えつつ検出性能を向上させることができる。

10

【 0 1 4 3 】

すなわち、第1の顔候補として抽出された点に隣接する点において、これらの点に対応するテンプレートサイズの入力画像領域において、肌色領域の占有率（肌色占有率）が所定の閾値以上である場合、又は、予め検出された（又は予め学習された）顔色情報を有し、この顔色領域の占有率（顔色占有率）が所定の閾値以上である場合にのみ、その点を検索範囲に設定することができる。ここで、肌色占有率は、例えば、肌色カラーテーブルを使用し、この肌色と比較することにより求めることができる。

【 0 1 4 4 】

この後、テンプレートマッチング部92は、ウィンドウ切出部91に対して、第1のスケール画像のうち先頭のウィンドウ画像のテンプレートマッチングが終了した旨を通知することにより、当該ウィンドウ切出部91から第1のスケール画像のうち次にスキャンされたウィンドウ画像をテンプレートマッチング部92に送出させる。

20

【 0 1 4 5 】

このようにテンプレートマッチング部92は、ウィンドウ切出部91において第1のスケール画像から切り出した全てのウィンドウ画像について、スキャン順にそれぞれ顔候補を検出する。

【 0 1 4 6 】

スケール変換及び切出部93は、各スケール画像毎に候補リストとしてDRAM11に格納された複数のスコア画像について、当該DRAM11から第1のスケール画像内でかつ候補リストの決定順に1つずつスコア画像を読み出して、前処理部94に送出する。

30

【 0 1 4 7 】

前処理部94は、スケール変換及び切出部93から得られたスコア画像について、矩形領域でなる当該スコア画像（400画素）から人間の顔画像とは無関係である背景部分に相当する4隅の領域を除去するために、当該4隅の領域を切り取ったマスクを用いて360画素分を抽出する。

【 0 1 4 8 】

そして前処理部94は、撮像時の照明により濃淡で表される被写体の傾き条件を解消すべく、当該抽出した360画素分のスコア画像のうち顔画像として最適な部位を基準とする平面を形成するように画素単位で濃淡値に補正をかける。

40

【 0 1 4 9 】

続いて前処理部94は、当該360画素分のスコア画像のコントラストを強調した結果をヒストグラム平滑化処理を行うことにより、CCDカメラ20のゲインや照明の強弱によらずに検出できるようにする。

【 0 1 5 0 】

次いで、前処理部94は、ガボア・フィルタリング処理を行うことにより、当該360画素分のスコア画像をベクトル変換し、得られたベクトル群を更に1つのパターンベクトルに変換する。人間の視覚細胞には、ある特定の方位に対して選択性を持つ細胞が存在することが既に判っている。これは、垂直の線に対して反応する細胞と、水平の線に反応する細胞で構成される。ガボア・フィルタリングは、これと同様に、方位選択性を持つ複数のフィルタで構成される空間フィルタである。

50

【0151】

そして、識別手段としてのパターン識別部95は、外部から供給される学習用のデータすなわち教師データを用いて、暫定的な識別関数を得た後、当該識別関数を前処理部94からパターンベクトルとして得られた360画素分のスコア画像に試して顔の検出を行う。そして、検出に成功したものを顔データとして出力する。また検出に失敗したものを非顔データとして学習データに追加して、更に学習をし直す。

【0152】

このパターン識別部95は、例えばパターン認識の分野で最も学習汎化能力が高いとされるサポートベクタマシンを用いて、前処理部94からパターンベクトルとして得られた360画素分のスコア画像に試して顔か否かの識別を学習しながら行う。

10

【0153】

サポートベクタマシン自体に関しては、例えばB.sholkopf外著の報告（B.Sholkopf、C.Burges、A.Smola、“Advance in Kernel Support Vector learning”、The MIT Press、1999.）を挙げることができる。本願出願人が行った予備実験の結果からは、サポートベクタマシンによる顔認識方法は、主成分分析（PCA）やニューラル・ネットワークを用いる手法に比べ、良好な結果を示すことが判っている。

【0154】

サポートベクタマシンは、識別関数に線形識別器（パーセプトロン）を用いた学習機械であり、カーネル関数を使うことで非線形空間に拡張することができる。また識別関数の学習では、クラス間分離のマージンを最大にとるように行われ、その解は2次数理計画法を解くことで得られるため、グローバル解に到達できることを理論的に保証することができる。

20

【0155】

すなわち、パターン識別部95は、前処理部94から与えられたスコア画像に基づくパターンベクトルについて、当該スコア画像内での顔データに相当する画像領域の位置（座標）及びその大きさ（画素数）と、当該スコア画像の切出し元となるスケール画像のフレーム画像に対する縮小率（すなわち上述の5段階のうちの該当する段階）と、テンプレートのサイズとをリスト化し、これをリストデータとしてDRAM11に格納する。

【0156】

この後、パターン識別部95は、スケール変換及び切出部93に対して、第1のスケール画像のうち最初のスコア画像の顔検出が終了した旨を通知することにより、当該スケール変換及び切出部93から第1のスケール画像のうち次のスコア画像を前処理部94に送出する。前処理部95は、当該スコア画像をパターンベクトルに変換してパターン識別部95に送出する。パターン識別部95は、パターンベクトルから得られた顔データに基づいてリストデータを生成してDRAM11に格納する。

30

【0157】

このようにスケール変換及び切出部93において、第1のスケール画像内で候補リストにある全てのスコア画像について、順番に前処理部94及びパターン識別部95の各処理を行うことにより、当該第1のスケール画像から撮像結果に存在する顔画像を検出することができる。

40

【0158】

すなわち、テンプレートマッチング部92は、入力画像スケール変換部90に対して、第1のスケール画像及び第1のテンプレートサイズのテンプレートを使用した顔検出が終了した旨を通知することにより、当該入力画像スケール変換部90から、第1のスケール画像及び第2のテンプレートサイズのテンプレートをウィンドウ切り出し部91に送出させる。第2のテンプレートサイズを使用した場合も、上述した第1のテンプレートサイズのテンプレートを使用した場合と同様の処理を行って全てのテンプレートサイズに対応する顔候補を検出した後、テンプレートマッチング部92は、入力画像スケール変換部90の第1のスケール画像に対して全てのテンプレートサイズを使用した顔検出が終了した旨を通知することにより、当該入力画像スケール変換決定部90から第2のスケール画像を

50

ウィンドウ切出部 9 1 に送出させる。

【 0 1 5 9 】

そして第 2 のスケール画像についても、上述した第 1 のスケール画像と同様の処理を行って、当該第 2 のスケール画像から全てのウィンドウ画像に対応する顔候補を検出した後、第 3 ~ 第 5 のスケール画像についても同様の処理を順次行う。

【 0 1 6 0 】

かくしてテンプレートマッチング部 9 2 は、撮像画像であるフレーム画像を 5 段階に縮小した第 1 ~ 第 5 のスケール画像及び複数のテンプレートサイズのテンプレートについて、それぞれ切出した複数のウィンドウ画像に対応する顔候補を抽出して、当該ウィンドウ画像と共に D R A M 1 1 に格納する。

10

【 0 1 6 1 】

ここで、本実施の形態においては、任意の大きさのテンプレートを使用することができるが、使用するテンプレートサイズを切り替えて、テンプレートサイズを選択することにより、入力画像に対して準備できる全てのテンプレートサイズに対して演算をする場合に比して、演算量を減らして高効率化することができる。例えば、一度顔が検出された場合に、次に顔検出する際はそのテンプレートサイズを使用することができる。また、例えば、ロボット装置に設けられた距離センサを使用し、この距離センサからの距離情報に基づき入力画像に含まれる対象物との間の距離を認識することにより、対象物の顔領域の大きさを予測してテンプレートサイズを選択する対象距離切り替え手段を設ける等することができ、目的に応じてテンプレートサイズを切り替えることができる。

20

【 0 1 6 2 】

(3 - 2) 顔検出部における動作

以上の構成において、ロボット装置 1 等においては、 C C D カメラ 2 0 により撮像したフレーム画像を縮小率が相異なる複数のスケール画像に変換した後、当該各スケール画像の中からそれぞれ所定サイズのウィンドウ画像を所定画素ずつずらすようにスキャンしながら 1 枚ずつ切り出す。

30

【 0 1 6 3 】

このウィンドウ画像について、平均的な顔画像を表すテンプレートを用いてマッチングをとて当該テンプレートとの相關値の集合であるマッチング結果画像を生成する。このように各スケール画像毎にそれぞれ全てのウィンドウ画像についてスキャン順にそれぞれマッチング結果画像を生成する。以下、マッチング結果画像から顔候補を検出する工程について詳細に説明する。

30

【 0 1 6 4 】

図 1 4 は、テンプレートマッチング部 9 2 において、テンプレートマッチング結果画像 R 2 から顔候補となる画素を検出する各処理工程を示すフローチャートである。図 1 4 に示すように、先ず、テンプレートマッチング結果画像 R 2 が入力されると、マッチング結果画像 R 2 をテンプレートサイズに分割し、その分割領域の 1 つ、例えば $0 \times wid_t - 1$ 、 $0 \times hei_t - 1$ において、最も相關値が高い点（座標）を抽出する（ステップ S P 1 1）。以下、マッチング結果画像 R 2 をテンプレートサイズに分割した領域を分割領域 r n 、分割領域 r n において、相關値が最も大きい点（座標）を local_max(x , y) という。ここでは、この各分割領域内において最も相關値が高い画素を抽出するが、本実施の形態においては、マッチング結果画像において分割された分割領域を左から右へ一行ずつ順に処理を行う場合について説明する。

40

【 0 1 6 5 】

次に、 local_max(x , y) が所定の閾値 (t h 5) より大きいか否かを判定し（ステップ S P 1 2）、大きい場合は、顔候補として追加する（ステップ S P 1 3）。上述した如く、入力画像スケール変換決定部 9 0 においては、スケールと共に入力画像に含まれると想定される顔の大きさのテンプレートサイズを選択するが、テンプレートサイズは異なる大きさの複数種類あり、複数種類ある各テンプレートサイズ毎にマッチング結果画像 R 2 を算出して顔候補を抽出すると、同一の点が抽出される場合がある。従って、ステップ

50

S P 1 3において、顔候補として同一の点がある場合、すなわち、異なるテンプレートサイズで顔候補を抽出した際に既に抽出されている場合はこの点は追加しない。

【 0 1 6 6 】

次に、顔候補として抽出された点に対応するテンプレートサイズの入力画像領域において、この画像領域内に含まれる肌色画素の占有率を求める。本実施の形態においては、肌色画素の占有率を求める際に、肌色カラーテーブル 1 0 0 を参照する。そして、この肌色画素占有率が所定の閾値 (t h 6) より大きいか否かを判定する (ステップ S P 1 4) 。大きい場合は、この local_max (x , y) の周辺、例えば上下左右の 8 近傍点を顔候補として追加する (ステップ S P 1 5) 。ここで、ステップ S P 1 3 と同様に、既にこれらの 8 近傍点が既に顔候補として抽出されている場合は、候補に追加しない。

10

【 0 1 6 7 】

ステップ S P 1 2 で local_max (x , y) が閾値 t h 5 未満だった場合、ステップ S P 1 4 で local_max (x , y) に相当する入力画像における肌色画素占有率が閾値 t h 6 未満であった場合、及びステップ S P 1 5 で顔候補の追加が終了した後は、いずれもステップ S P 1 6 に進み、次の顔候補を抽出するために次の分割領域に移り、処理を進める。

【 0 1 6 8 】

先ず、マッチング結果画像 R 2 において、x 方向にテンプレートサイズ分、すなわち、 wid_t だけずれた隣の分割領域に移る (ステップ S P 1 6) 。次に、 wid_t だけずれた x 座標 (x + wid_t) の分割領域において、その x 座標がマッチング結果画像の幅 (x 方向の辺) wid_r より大きい場合は、分割領域がマッチング結果画像に含まれないことを示し、次の行に移り、 0 x wid_t - 1 であって、 y 方向にテンプレートサイズ分、すなわち、 hei_t だけずれた隣の分割領域に移る (ステップ S P 1 8) 。次に、分割領域の y 座標がマッチング結果画像の高さ (y 方向の辺) hei_r より大きいか否かを判定し (ステップ S P 1 9) 、大きい場合は、マッチング結果画像における全ての分割領域の相関値の最大値を求めたことを示し、処理を終了する。

20

【 0 1 6 9 】

一方、ステップ S P 1 7 及びステップ S P 1 8 において、分割領域がマッチング結果画像に含まれると判定された場合は、再びステップ S P 1 1 に戻り、その分割領域内で最も高い相関値を有する点を抽出する。

30

【 0 1 7 0 】

本実施の形態においては、マッチング結果画像 R 2 をテンプレートサイズに区切った分割領域における相関値の最大値を求めていたため、ステップ S P 1 6 において、隣接する分割領域に移る場合は、x 方向に wid_t だけずれるものとしたが、マッチング結果画像 R 2 は、テンプレートサイズ以下のサイズであれば、任意の大きさに分割することができる。その際、分割する画像の大きさの幅 (x 方向の辺) wid_step 、高さ (y 方向) hei_step とすると、ステップ S P 1 6 及びステップ S P 1 8 において、それぞれ x 方向に wid_step 、又は y 方向に hei_step 移動することにより、次の分割領域に進むことができる。

【 0 1 7 1 】

こうして、テンプレートマッチング部 9 2 において、顔画像であると判断されたウインドウ画像 (すなわちスコア画像) について、当該スコア画像の矩形領域の 4 隅部分を除去した後、濃淡補正及び続くコントラスト強調の平滑化を行い、更に 1 本のパターンベクトルに変換する。

40

【 0 1 7 2 】

そして当該パターンベクトルについて、元のスコア画像内の顔検出を行って顔データ又は非顔データを判断し、顔データが存在するスコア画像の画像領域の位置 (座標) 及びその大きさ (画素数) と、当該スコア画像の切出し元となるスケール画像のフレーム画像に対する縮小率とをリスト化したリストデータを生成する。

【 0 1 7 3 】

このように各スケール画像毎にそれぞれ全てのスコア画像についてリストデータを生成した後、当該各リストデータに含まれるスコア画像同士を比較して、相互に重なり合う領

50

域を除去した顔決定データを求めるにより、元のフレーム画像から顔画像を検出することができる。

【0174】

(3-3) 具体例の効果

上述した図13に示すマッチング結果画像R2から顔候補から、単に閾値以上である点を全て顔候補として抽出するのに比較して、テンプレートマッチング部92にて顔候補として抽出される点を低減し、これにより、後段の処理における計算量を飛躍的に削減することができる。

【0175】

本具体例においては、ウィンドウ画像について、平均的な顔画像を表すテンプレートを用いてマッチングをとって大まかに顔画像であるか否かを判断する際に、テンプレートマッチング結果画像を所定のサイズに仕切り、相関値の最大値を顔候補として抽出して明らかに顔画像でないウィンドウ画像を除去することにより、本来顔である領域を見逃すことなく、後段の顔検出処理に要する演算量及び時間を減少させることができ、かくしてリアルタイム性を格段と向上した顔検出装置及びこれを搭載したロボット装置を実現することができる。

【0176】

また、相関値が最大となる点と共にその周囲においても顔検索範囲とすることにより、顔検出精度を向上することができる。更に、所定の閾値以上の肌色占有率又は顔の色占有率を有する場合のみ、顔検索範囲として設定することにより、顔検出精度を保ちつつ顔候補を減らして後段の演算量を減らすことができる。更にまた、テンプレートのサイズを適宜切り替えることにより、更に演算量を減らすことができる。

【0177】

以上説明したように、顔検出部110においては、例えば所定の閾値以上の肌色占有率を有する領域を顔候補とするなど、所定の条件を満たすものを顔であると判断する。そして、本実施の形態においては、このような顔検出タスクの追加機能として、上述したような非顔判定部を設け、改めて顔でないものを抽出して排除することで、演算量を増大させることなく顔検出の誤検出を低減することができる。

【0178】

また、ロボット装置1は、顔領域の他、入力画像から障害物を検出したり、自己位置同定のためにランドマークを検出したり、入力画像から動体を検出してトラッキングしたりと、入力画像から所定の条件を満たす種々の物体を検出する処理を行う。その場合においても、本実施の形態における非顔判定部と同様、所定の条件を満たすものを検出対象ではないとして排除する誤検出防止手段としての非対象物検出手段を設けることにより、誤検出を低減することができる。

【図面の簡単な説明】

【0179】

【図1】本発明の実施の形態における顔検出装置における顔検出処理を機能別に示すブロック図である

【図2】同顔検出装置に設けられる非顔判定部における非顔判定のための第2の条件を説明するための図であって、顔候補として検出されている顔領域内の画素P(i,j)を示す図である。

【図3】本発明の実施の形態におけるロボット装置の外観構成を示す斜視図である。

【図4】同ロボット装置の回路構成を示すブロック図である。

【図5】同ロボット装置のソフトウェア構成を示すブロック図である。

【図6】同ロボット装置のソフトウェア構成におけるミドル・ウェア・レイヤの構成を示すブロック図である。

【図7】同ロボット装置のソフトウェア構成におけるアプリケーション・レイヤの構成を示すブロック図である。

【図8】同アプリケーション・レイヤの行動モデルライブラリの構成を示すブロック図で

10

20

30

40

50

ある。

【図9】同口ボット装置の行動決定のための情報となる有限確率オートマトンを説明するためには、使用した図である。

【図10】有限確率オートマトンの各ノードに用意された状態遷移表を示す図である。

【図11】同口ボット装置の非顔検出部を除いた顔検出タスク機能を示すブロック図である。

【図12】(a)及び(b)は、それぞれ入力画像(ウィンドウ画像)及びテンプレートを示す模式図である。

【図13】入力画像(ウィンドウ画像)とテンプレートとから求めた相関値の集合であるマッチング結果画像を示す図である。

【図14】本発明の実施の形態における顔検出装置のテンプレートマッチング部12における処理工程を示すフローチャートである。

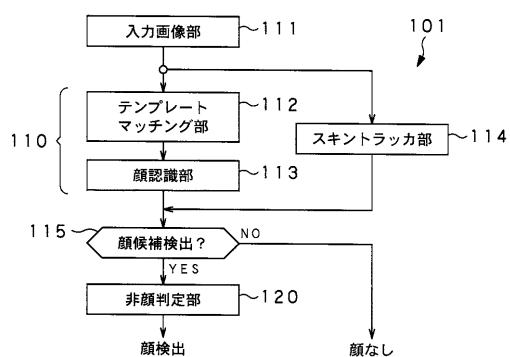
【符号の説明】

【0180】

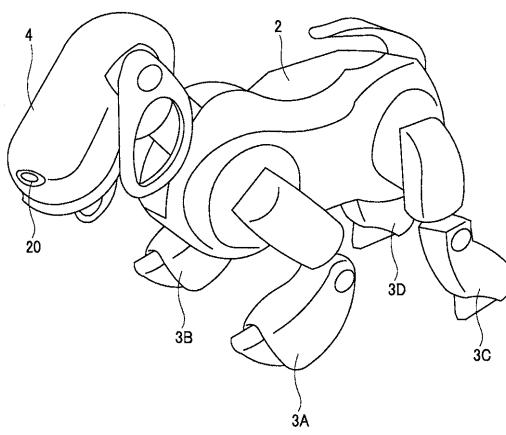
1 口ボット装置、 14 信号処理部、 11 D R A M、 20 C C D カメラ、
 90 入力画像スケール変換部、 91 ウィンドウ切出部、 92, 112 テンプレートマッチング部、 93 スケール変換及び切出部、 94 前処理部、 95 パターン識別部、 101 顔検出装置、 111 入力画像部、 113 顔認識部、 114 スキントラッカ部、 115 顔検出判定部、 120 非顔判定部

10

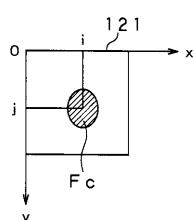
【図1】



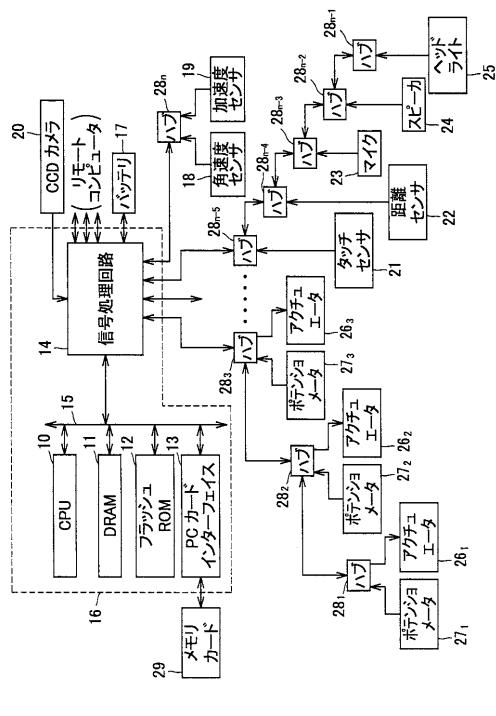
【図3】



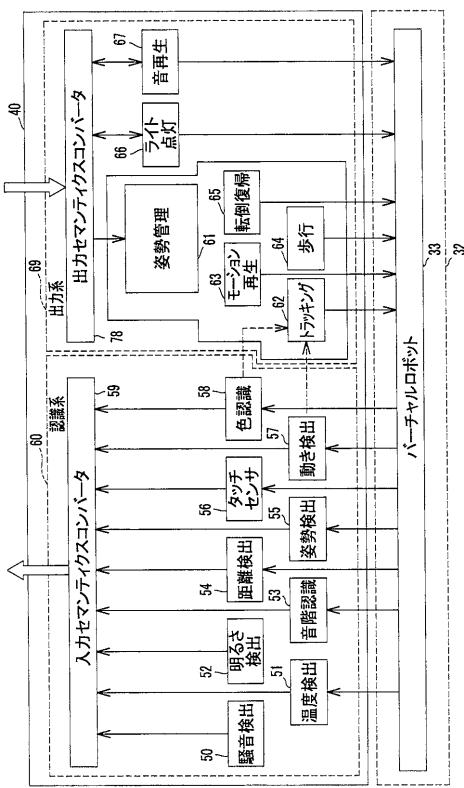
【図2】



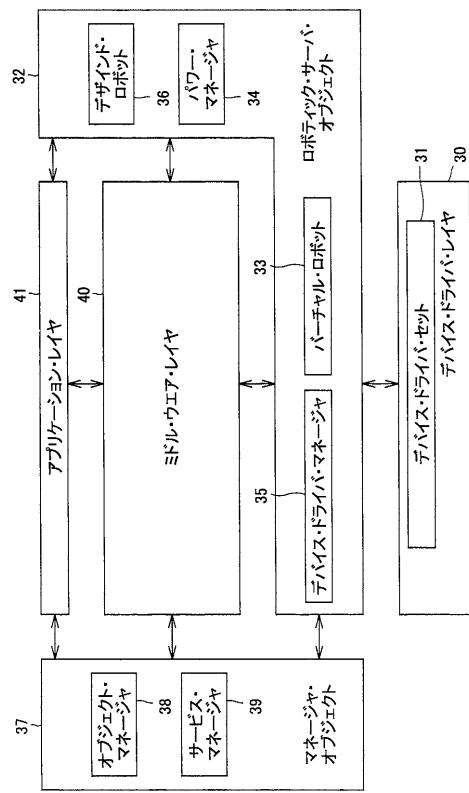
【図4】



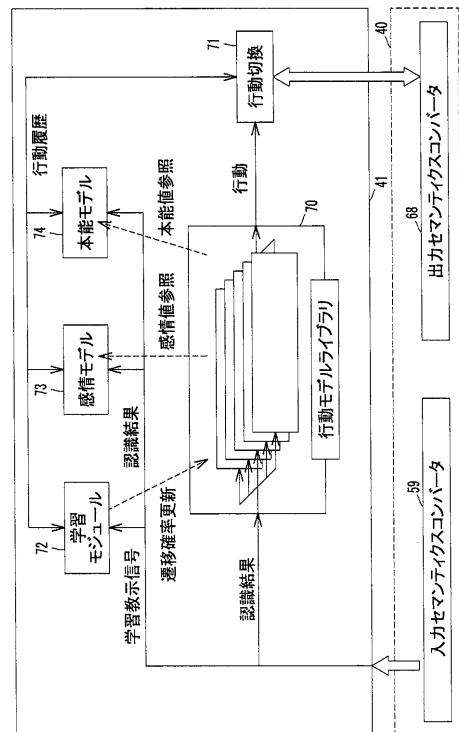
【図6】



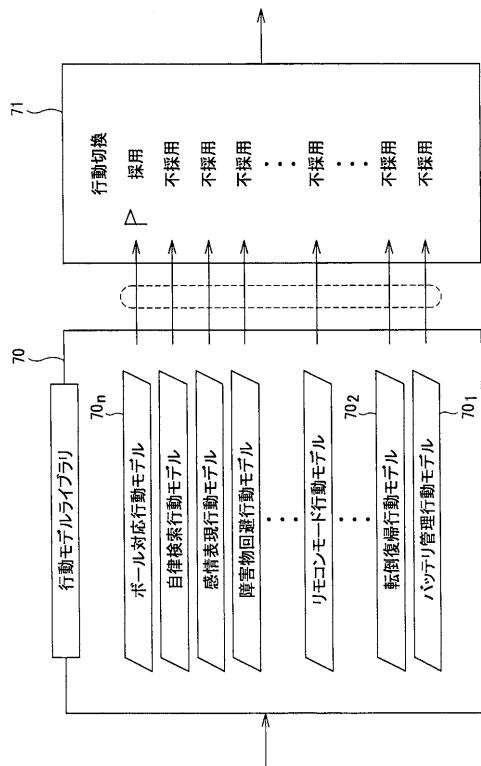
【図5】



【図7】



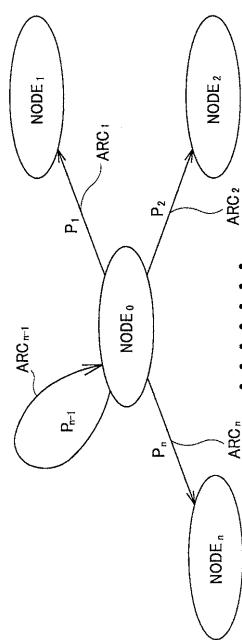
【 四 8 】



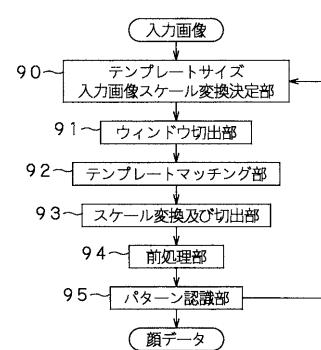
【 図 1 0 】

		入力/イベント名		データ名		デコード範囲		他のノードへの遷移確率				
								A	B	C	D	n
								node 120	node 120	node 1000	node 600	
								ACTION 1	ACTION 2	MOVE BACK	ACTION 4	
node 100												
遷移先ノード												
出力行動												
1	BALL	SIZE		0.1000		30%						
2	PAT											
3	HIT											
4	MOTION											
5	OBSTACLE	DISTANCE		0.100								
6	Joy			50.100								
7	SURPRISE			50.100								
8	SADNESS			50.100								

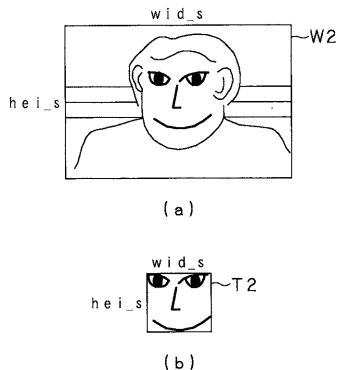
【 四 9 】



【 図 1 1 】



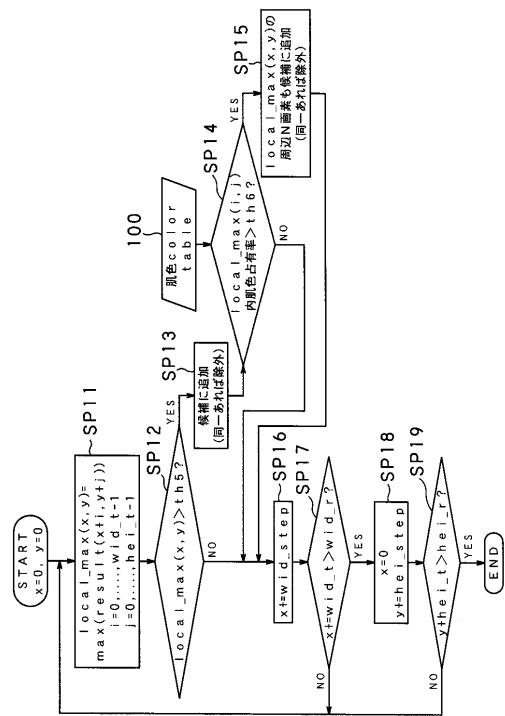
【 図 1 2 】



【図13】



【図14】



フロントページの続き

F ターム(参考) 5B057 AA05 BA02 CA01 CA08 CA12 CA16 CB01 CB08 CB12 CB16
CC01 CE16 DA07 DA12 DB02 DB06 DB09 DC04 DC25
5L096 AA02 BA05 BA18 CA02 CA14 FA15 JA03 JA09