論文の内容の要旨

論文題名　　画像情報を用いた人物動作認識システムの研究

氏名　　大和淳司

本論文では動画像を用いて人間の動作を認識するシステムの実現を目指した。そのための認識手法、前提としての動画像中の移動物体の追跡、抽出を行う手法、さらに応用として動画像データベース検索への適用について論じた。これらの技術は動画像処理の重要な要素であり、人間と自然なインタラクションを行う機械の構築に重要な役割を果たすものである。

第1章では、画像認識に関する研究の背景をまとめるとともに、本論文の基本方針として、(1)人間の動作を対象とした、(2)動作の「識別」を、(3)サンプルからの学習を用いて行うことを述べた。以下、この方針に沿って動作認識システムのアルゴリズムや応用例について各章で述べる。

第2章では、動画像認識の研究および隠れマルコフモデル(HMM)の画像認識への適用について述べた。
従来中心であったモデルベースの動作認識手法は、幾何学的なモデルフィッティングを用いることが多い。このため時としてマッチングのミスが大きな間違いにつながる。特に照明条件など環境条件に対して頑健でないことがある。そこでモデルベースの手法ではなく、画像の特徴量を用いたボトムアップな手法をとり、事例からの学習によって認識系を構築する基本方針を示した。この方針に適した手法として、音声認識で広く用いられているHMMを適用することとした。従来あった画像関連分野の応用例としては、文字認識、閉曲線図形の識別、顔画像の分割、顔画像による個人識別、などがあったがいずれも時間的な変化を伴わない対象への適用であった。著者らによる動画像への適用以降、動作認識へのHMMの適用は広く用いられる手法となり、手話の認識、顔表情の認識など多くの適用事例がある。

第3章では、HMMをどのようにして動作認識に適用したか、その基本的な手法について述べ、テニスの実画像を用いた実験について説明した。
認識対象の動画像は、まず前処理によって人物領域の抽出と正規化を行う。抽出は、高周波成分を除いた後、背景画像との差分によって行った。この領域を重心によって位置正規化、平均半径によって大きさの正規化を行った。前処理後の動画像の各フレームから特徴ベクトルを抽出し、特徴ベクトル列を得る。特徴ベクトルとしてはメッシュ特徴を使用した。これは、画像を8画素×8画素のブロックに分割し、その内部の人物領域の比率を並べて特徴ベクトルとするものである。次にこの特徴ベクトル列をベクトル量子化によりシ

ンボル列に変換した。こうして動画像の画像フレーム列をもとにして得られたシンボル列をHMMで学習・認識する、というのが本手法の骨子である。

HMMによるシンボル列の学習は、HMMのパラメータ推定という形で行われる。HMMのパラメータは、主に状態遷移確率行列、シンボル出力確率行列と初期状態確率、である。学習過程では、Baum-Welchアルゴリズムでパラメータの推定を行い、与えられた学習パターンを発生しやすいパラメータをもつHMMを得る。認識対象の各カテゴリごとにHMMを用意しこれら全体が認識系となる。認識の過程では、認識対象のシンボル列をどのHMMが発生しやすいかを計算し、尤度最大のHMMのカテゴリが認識結果となる。この計算はforwardアルゴリズムによって行われる。以上の手法により、動画像中の動作画像を学習認識することができる。

この手法の有効性を確認するために、テニス動作の画像を用いた実験を行った。テニスの動画像は、適当なカテゴリ数がある、同じカテゴリの動作を複数の被験者で容易に行うことができる、あるカテゴリの動作が被験者によってほぼ同じであるが適度にバラツキが発生する、など適度な難しさをもつ課題である。実験では、6カテゴリのテニス動作を使用した。フォアボレー、バックボレー、フォアストローク、バックストローク、スマッシュ、サービスである。3人の被験者の動作を各カテゴリ10サンプル収集した。

実験の結果は以下の通りであった。学習対象と同じ被験者を認識対象とした場合には、90%以上の認識率を得た。学習対象と異なる被験者を認識対象とした場合には、認識率は悪化し、60%台であった。しかし、学習に使用する被験者の人数を２人に増やして混合して使用した場合、認識率は70%～80%に向上した。

この結果から、提案手法によってHMMを使用して人物の動作認識を行うことができることが確認できた。また、誤認識の原因となったサンプルの中身を精査したところ、誤認識となったサンプルの多くが、VQの段階で他のカテゴリに属するコードワードに割り当てられていることが確認された。そこで、特にVQ段階の改良を中心として本手法の改良を行うことが必要であると考えられる。

第4章では、前章で述べた手法をもとに、より人物動作認識に特化した改良を施した。

まず、前章での実験のエラーの中身の検討から、ベクトル量子化段階でのミスが多いことに鑑み、この段階の改良を行った。すなわち各カテゴリごとに個別のベクトル量子化コードブックを持たせることである。通常は全体で共通の一つのコードブックでのベクトル量子化を行うが、各カテゴリごとに個別のコードブックを持てばこの段階でのミスは原理的にない。但し、異なるカテゴリのコードワードでもマッチする場合は距離が大きい場合もあるため、コードワードからの距離に応じたペナルティ関数を与える。これはコードブック全体のクラス平均の分散をもつガウシアン関数とした。さらに各カテゴリごとに均一のコードブックを作成するために、LBGアルゴリズムによるコードブックの自動作成を行った。

次に時間軸に関しても、通常人物動作が時間的に一方向に変化するものであることから、LRモデル（Left to Rightモデル）をHMMに導入した。具体的には、状態遷移確率のうち、同じ状態および直後3状態以外への遷移を禁止し、ある程度の伸縮のある一方向への流れになるように制限を加えた。これによって、推定するパラメータの数を大きく減少させ、モデルの安定化とパラメータの推定精度を向上させる効果が期待できる。

以上の改良を加えた上で実験を行い、以下の結果を得た。まず、全体として認識率は大きく向上し学習と認識対象が同じ被験者の場合で99%程度、異なる場合でも３人の動作を学習に使用すれば86.5%の認識率を得た。これは改良前の手法に比べて約9ポイントの向上であり、改良の効果が確認された。この程度の認識率では、ミスの許されないセキュリティ分野への適用は難しいが、統計的な結果を求められる調査の分野や人間がインタラクティブに使用する場合等は十分に実用的なレベルであると考えられる。次章ではそのような例の一つである、動画像データベースへの適用について検討する。

第５章では、HMMを用いた動作認識手法を動画像データベースの内容ベース検索に適用するための検討を行った。

検索対象として、動画像の標準圧縮方式であるMPEGでエンコードされたデータを扱う。このため、画像特徴として、DCT（離散コサイン変換）係数を使用することを検討し、メッシュ特徴と同等以上の結果が得られることを確認した後、DCT係数を特徴量として使用し、実際に動画像検索を行ったときどの程度の検索精度が得られるかを評価した。基本的な考え方としては、検索対象動作で学習を行ったHMMでデータベースをスキャンし、一定以上の尤度がある部分を該当部分として拾い出すことで検索を行う。この場合の検索はある動作が何であるか識別する問題に比べて絶対的な類似性の評価が求められる点でより難しい。検索実験を６カテゴリのテニス動作画像を用いて行った結果、Precision rate, Recall rate ともに80%程度の精度を得ることができた。これは、ユーザが動画像のサンプルを選んで類似の物をシステムが提示し、その中からさらに所望のものをユーザが選択する、というインタラクティブな検索を行う場合には十分に使用に耐える精度である。

また、前処理として必要な人物領域の抽出のため動き情報をMPEGデータのMC（動き補償ベクトル）から取り出して使用することを検討した。MCの大きさとDCT係数から取り出したテクスチャ情報との組み合

わせによって、安定して動領域を抽出できる見通しを得た。

第6章では、これまで述べてきた動作認識を実際に使用する際に前提となる前処理として重要な、人物領域の抽出と追跡のためにステレオビジョンヘッドを用いた動物体の追跡法と奥行き情報を用いた物体の抽出について述べた。

ステレオビジョンヘッドはヒューマノイド用に製作されたものを使用し、制御系は、DSP（デジタル信号プロセッサ）のネットワークを階層的に構成したものを使用した。まず、タスクを階層的に分割し、それに対応した階層的ネットワークを順次構築して段階的な実装を行った。タスクは発達心理学の知見にもとづき、単眼視での注視、単眼視での追跡、両眼視での輻輳制御と距離情報による抽出の3つに分割した。それぞれに対応するDSPネットワークを実装し、良好な動作を確認した。ステレオビジョンヘッドは動く物体に注意を向け、スムーズに追跡し、追跡が失敗すると新たな動物体に再度注意を向ける。両眼視での注視物体の抽出結果を輻輳制御の計算領域にフィードバックすることで、輻輳制御の安定化向上も確認した。

以上、全体を通して本研究により、動画像を用いた人間の動作認識システムの各要素について検討を行った。本論文で提案されたHMMによる動作認識手法は広く用いられており今後実用的な応用例が多く出ることが期待される。