

柔軟な動画像認識システムのための動きフィールド特徴ベクトル表現

早川 仁

半導体技術の進歩により、マイクロプロセッサには数億ものトランジスタが搭載され、1秒間に何十億回もの演算が行えるにもかかわらず、人間であれば瞬時に行える認識処理を実現することは非常に難しい。これに対し、認識処理において基本となる演算が、マイクロプロセッサの行う数値演算ではなく、入力に似た過去の記憶を思い起こす連想に基づいていると考え、この連想処理を高速に実現するようなハードウェアを作成し、それによって柔軟な認識を目指すシステムが提案されている。このようなシステムで重要になってくるのは、当然のことながら連想処理の高速ハードウェアである。そしてそれと同じくらい重要なものとして、どのような形で情報を表現し、連想処理を行うかがある。本研究では、このような背景のもと、動画像を対象とした連想に基づいた高度で柔軟な認識システムの構築を目指し、動画像の特徴ベクトル表現およびそれを処理する連想プロセッサアーキテクチャについて提案した。

従来の動画像認識では画像から特徴を取り出し、それを Hidden Markov Model(HMM)のような動画像に限定されないより一般的な手法で認識を行っていた。このような手法では、ジェスチャーの認識といった比較的単純な用途に対して必要以上に計算コストがかかってしまう。そこで、動きフィールドを基本として、シーケンス全体を1つのベクトルとして表現する手法を提案した。これにより、動画像認識を単純なベクトルのテンプレートマッチングによって実現できる。また、従来の連想プロセッサでは連想処理に特化していたため、ベクトルの加工や連想処理の階層的な実行等の認識システムを作るうえで必要となる処理を行うためには、複数の連想プロセッサおよび汎用プロセッサを用意し、その間でデータを転送する必要があった。そこで本研究では、連想処理に必要な演算器を再利用して算術演算のような、より一般的な処理やプログラム制御を行える仕組みを備えた、連想プロセッサアーキテクチャを提案した。

まず、動画像の認識では、まず、その第1段階として動画像から動きを抽出する部分が必要になってくる。この部分については、動きの内容に依存せずに一般的に使うことができるように、各ピクセルにおいて動きの方向と大きさを保持している動きフィールドを採用した。動きフィールドの代表的な手法には勾配に基づくものとブロックマッチングに基づくものがある。前者は計算コストが少

ない代わりに精度に問題があり、後者は精度がよい代わりに計算コストが高い。また、このどちらの手法においても輝度の変化に弱く、また、動きを検出するフレーム間隔を動きの内容に応じて適切に設定する必要がある。これらの問題を解決すべく、本研究では、方向性エッジの変化に着目したブロックマッチングに基づく動きフィールドの作成アルゴリズムを提案した。本手法では、グレイスケール画像ではなく2値化された方向性エッジを使うことで輝度変化に強く、そのエッジの変化した部分すなわち動いた部分のみを抽出したマップを作成することで動きの誤検出を減らすと同時にフレーム間隔の自動調整を行う。さらに、ブロックマッチングを、射影した1次元のヒストグラムのマッチングとして行うことで計算量を削減する。このようにして求めた動きフィールドの精度を評価するため動きフィールドからベクトル表現に変換し、そのベクトルのシーケンスで簡単なジェスチャーを、時系列データの認識で一般的なHMMを用いて認識させる実験を行った。そして、提案手法によって動きフィールドの作成が、勾配法の代表的な手法であるノーマルオプティカルフローと比べて高い精度で行えることを示した。

続いて、動きフィールドを元にシーケンス全体を単一ベクトルとして表現する方法を提案した。本手法を用いることで、動画認識をHMMのような複雑な認識手段ではなく、単純なテンプレートマッチングで実現することができるようになる。提案手法では、まず、動きフィールドから4方向の成分マップを作成し、それを動きの方向と垂直な軸に対して射影を行うことで、動きの方向が異なる4つの部分ヒストグラムを作成し、それらを連結してProjected Directional Motion Histogram (PDMH)を作成する。すなわち、動きフィールド1枚からPDMHと呼ばれるヒストグラムを1本作成する。そして、元のシーケンスに対応するPDMHのシーケンスに対して、空間軸および時間軸へ射影し8個の部分ベクトルを作成し、それらを連結することでシーケンスの特徴をあらわすベクトルの表現とする。これによりシーケンス全体から1本のベクトル表現が作られることになる。これを動きフィールドシーケンスベクトルと呼ぶ。この特徴ベクトルでは、時間および空間での動きの分布がベクトルの要素番号として反映されており、そのマッチングにおいてベクトルに対してシフト等の処理をすることで、空間的および時間的な位置のずれを補正することができる。また、PDMHシーケンスからベクトルを作成する過程で、射影するときの部分ベクトルの要素に入れる範囲を変更することで空間的、時間的な位置だけでなく大きさに対する補正も同時に行う補正法を提案した。これらの補正法は動きフィールドシーケンスベクトルを作成した後に、ベクトル要素を加工することで実現可能である。したがって、計算コストの高い各フレームでの処理は不要で、マッチングの段階でベクトルに対して処理を行うことで、たとえサンプルの動きの大きさや位置にずれがあったり、時間的な大きさや位置すなわち期間や時刻にずれがあったりしても、それらを補正したマッチングを行うことができる。単純なジェスチャー認識に対して本手法を適用し、単一ベクトル表現でもHMM以上の性能を出せることを示すとともに、時空間の位置および大きさのずれに対する補正法によりさら性能を高めるこ

とができることを示した。

ところで、動画像の認識システムにおいては、動きの認識結果、静止画の認識結果など複数の認識結果を統合してより高次での認識及び判断を行う必要がある。このような複数の認識処理を行い、そこからさらに階層的に認識処理を行うような高次の認識を行うことを考えた場合、従来の連想プロセッサはチップ全体で 1 つの連想処理のみが行えるため、それぞれの連想処理に連想プロセッサが必要となり、さらに、連想以外の処理に対して汎用的なプロセッサまで必要となってしまう、その間のデータ転送等で効率が悪くなってしまう。そこで、連想処理だけでなく汎用処理も実行できる Multiple Instruction-stream Multiple Data-stream(MIMD)エレメントを複数持った MIMD 型の連想プロセッサを提案した。これを用いることで、連想処理の必要の度合いに応じて柔軟に連想処理を行う MIMD エレメントの数を調整でき、CPU のような汎用処理も実行できるので効率のよい柔軟な連想処理が可能となる。MIMD エレメントは、連想処理に必要な演算器を分割し、連想処理には専用のパイプライン構成を用いることで従来の連想専用プロセッサの効率を維持したまま、乗算などの一般的な算術処理も実現できるようにした。また、レジスタに工夫を加えることで、繰り返しなどの構造を持ったプログラムを実行できるようにし、内積など並列されている要素間の演算も可能にした。これらによりベクトルに対する加工を並列処理できるようになり、また、それらの処理の制御を連想プロセッサ内部で実現できるようになった。一方、命令に使用されるメモリを抑制するため、命令をグループに分けて必要に応じて切り替えることで短い命令長を実現した命令セットを考案し、それを実装した。その結果 8 ビットの命令長で 70 種類以上の命令を実装できることを確認した。そして、このアーキテクチャの実現可能性を示すとともに、顔画像認識を例に従来の連想プロセッサと汎用プロセッサの組み合わせに対して優位であることを示した。

このように、本研究では、柔軟な動画像認識ハードウェアシステムを目指し、動画像の特徴ベクトル表現アルゴリズム、及び、ベクトルを柔軟に処理する連想プロセッサアーキテクチャについて提案し、それらの有用性を示した。そして、動画像認識システムにおいて、ハードウェアと緊密に連携する柔軟な認識システムがひとつの選択として十分に実現可能であることを示した。