

早稲田大学大学院情報生産システム研究科

博士論文審査結果報告書

論 文 題 目

Research on Robust Local Feature
Extraction method for Human Detection

申 請 者

TANG, Shaopeng

情報生産システム工学専攻
マルチメディアシステム研究

2011年 2月

人物の検出問題はコンピュータビジョンや画像処理研究において、画像解析、画像理解のために人物という特定の情報を抽出するという1つの重要な技術分野となっている。インテリジェントカー、監視系システム、ロボットなど多くの応用が人物検出に関わっている。膨大な画像データから、高速に、高精度に人物検出処理を行うことは、人間の活動を理解するための第1歩でもある。画像上には人物は様々な姿、ポーズ、サイズで現われ、また照明や背景の違いを考慮して、人物検出問題を扱わねばならない。人物検出問題は古くからのコンピュータビジョンの課題であり、多くの研究がされてきた。従来から提案された方法として、テンプレートマッチング法、コードブック法があり、これらは高速な処理は行えるが、背景が複雑な場合や人物の様々なポーズに対応するには精度良く検出できないという欠点がある。一方、学習モデルに基づく方法は高速化には難があるが、固定されない様々なポーズにも対応でき高精度な検出が可能という長所がある。本論文は学習モデルに基づく方法で、高速かつ高精度を目指している。学習モデルに基づく方法は大量のデータからあらかじめ定義された特徴を学習メカニズムで抽出しておき、その特徴データを分類処理に使うことになる。検出処理においては、新しいデータに対して同様の特徴データを抽出し、分類処理を行うこととなる。現在まで、2005年にHOG(Histogram of Oriented Gradient)法や2008年にCOV(Covariance matrix)法が提案されている。HOG法とCOV法とも、ピクセルの勾配を計算することで、特徴抽出を行っている。HOG法は計算量は少ないが、検出精度に問題があり、COV法は高精度の検出は行えるが、計算量が膨大になるという欠点がある。本論文では学習モデルに基づく新しい手法を考案し、COV法と同程度以上の精度を持ち、計算量はHOG法と同程度であるというHOT(Histogram of Template)法を提案している。更に、拡張HOT法を提案し、より高精度な検出率を達成している。また様々な人物のポーズに対応するために、ポーズに不変な特徴を抽出する手法を提案している。この結果、計算量が少なく、高精度な検出することが可能となることを確かめている。

第1章(Introduction)は、画像から物体検出と人物検出する問題に対して、適用する応用と、現在までに提案された研究成果と課題について整理している。

第2章(Histogram of Template)は、提案するHOT法を説明している。分類処理のために、前もって定義するテンプレートと特徴を表現する計算式を提案している。HOG法で使われている各ピクセル毎の勾配だけではなく、色彩(RGB)と輝度を考慮したテクスチャー情報を特徴量として用いている。3x3ピクセルに対して、勾配とテクスチャー情報からなる連結した8種類の3ピクセルをテンプレートとして用いている。HOG法では1ピクセルだけを計算するのと比較し、連結した3ピクセルを計算することに特徴がある。INRIAが公開しているデータセットに対して、分類手法としてはKSVM(kernel Support Vector Machine)を共通に用いて実験した。評価は、DET(Detection Error

Tradeoff)曲線を用いて行った。約 4 百万個のウィンドー（部分画面）に対して実験したところ、 10^{-3} FPPW (False Positive Per Window) を達成するのに、HOG 法では 94%、HOT 法は 99%の正しい検出率(Detection Rate)を確認している。一方、COV 法は可変のサブウィンドウを設定でき、高い検出率が得られることが期待できたが、98%の検出率であり、HOT 法よりも少し悪く、かつ多くの計算量を必要とすることが分かった。またウィンドウサイズの影響を知るために、ウィンドウサイズをランダムに設定し、実験したところ、 10^{-3} FPPW に対して、検出率が HOG 法と COV 法ともに 80%であったが、HOT 法は 91%であり、大きな改良が見られている。テンプレート数を 12 種類に増加させ、評価したところ、96%に更に改良できている。様々な輝度に対して実験したところ、HOT 法はほとんど輝度の影響を受けないことも確かめられた。HOT 法は輝度の正規化の処理が必要なく、浮動小数点処理を行う必要はなく、整数処理で良いことが確かめられ、計算量とハードウェア量の観点からもメリットがあることが判明している。本成果は今までにない新手法を生み出したことで学術的な貢献は大きく、高く評価できる。

第 3 章 (Extension of HOT Feature) では HOT 法のテンプレートをピクセルレベルからブロックレベルへ拡張した。1 ブロックは 2×2 ピクセル又は 3×3 ピクセルからなり、ブロックレベルへ拡張することで、大局的な特徴抽出が可能となり、検出率が向上すると期待される。画像分野で一般的に使われている INRIA データセットに対して実験した。 10^{-3} FPPW に対して、LSVM(Linear SVM)を用いたところ、オリジナルな HOT 法では 94%、ブロックレベルへ拡張した方法では 97.5%であった。ブロックレベルへ拡張した HOG 法は 94%、COV 法は 97.5%であった。COV 法は HOT 法より計算量を必要とすることから、提案している拡張 HOT 法の有効性が確かめられた。

次に、静止画像から動画の人物検出問題を扱うために、人物の動きのオプティカルフロー量と HOT 法による特徴量を同時に考慮した特徴量を用いることを提案している。CAS(Chinese Academy Science)データセットに対して実験したところ、OP 法 (Optical Flow,2004)、KLT 法(Karhunen Loeve Transform, 2007)、IMHCD 法 (Intra Motion Histogram Central Difference,2006) と比べたところ、 10^{-2} FPPW において、拡張 HOT 法は 96%、OP 法は 60%、KLT 法は 80%、IMHCD 法は 92.5%の検出率が得られ、その有効性が確かめられている。

第 4 章 (Acceleration Method) では、GPU(Graphic Processor Unit)を使用して、計算時間の短縮を図ることを試みている。人物の特徴抽出の計算では、多くの並列演算があり、多数の並列演算器をもつ GPU を用いることにより、特徴抽出処理と分類処理を大幅に計算時間短縮できる可能性がある。特徴抽出の処理において、テンプレートの各要素の状態を計算際に、 16×16 個のピクセルの計算は独立にできることに着目し、256 個のプロセッサコアを用いて並列に実行することとした。更に、Histogram 計算においても、32 の個の並列化

が可能となることが分り、32 プロセッサコアを用いることとした。GPU としては、GTX285 (NVIDIA 社製で 240 コアからなる) を用い、ソフトウェア環境は CUDA フレームワークを用いた。640x480 ピクセルの画像に対して行った実験では、CPU(Intel Quad Q9550)の1コアを利用した場合は 63.875 秒かかったものが、GPU を用いることで、48.924 ミリ秒であり、約 1300 倍の速度向上が図られた。1 コア自身のクロック周波数は GPU(GTX285)では 1.45GHz であり、CPU(Q9550)では 2.83GHz であり、1/2 倍程度のスピードしかないが、計算アルゴリズムの工夫で並列計算を活用したことと、メモリバスの速度が 10~20 倍速いことの原因で、1000 倍以上の速度向上が図られた。この結果、リアルタイムで人物検出が可能となることを示しており、非常に高く評価できる。

第 5 章 (Pose-Invariant Human Feature) では、様々なポーズを取る人物に対しても、人物検出が可能なることを示している。最初に、人がどのようなポーズを取っているのかを推定する。この際に、Loss function を導入し、推定されたポーズと実際のポーズ間の差分をとり、学習モデル化手法により、差分が最小化されるように最適化を行い、正確なポーズを推定している。人間の体を 6 要素に分け、各要素は 2 次元の点の接続関係図で表現される。各要素の各々の特徴量を計算し、全体の人物の特徴量を得ることができる。この特徴量をもとに、人物検出を行うため、様々なポーズに対応可能な人物検出が可能となっている。従来提案されている SLBD 法 (Structure Learning Based Detector, 2007) と比較し、 10^{-3} FPPW では検出率はほぼ同じであるが、 10^{-4} FPPW では HOG 法で 92.5%、SLBD 法では 91% の検出率が得られ、提案手法の有効性を示している。

第 6 章 (Conclusion) では本論文のまとめと今後の課題について記述している。

人物検出という従来から取り組まれてきた課題であるが、HOT 法、拡張 HOT 法という新手法を考案し、検出率の向上を図り、計算量を大幅に削減したことは学術的に高く評価できる。また、高速化を図るために、GPU 上に実装して、リアルタイムに検出できる性能までを実現したことは、実用的にも有効であることを示したことを高く評価する。

よって本論文は博士 (工学) の学位論文として価値あるものと認める。

2011年 1月 14日

審査員

主査	早稲田大学教授	工学博士 (早稲田大学)	後藤 敏
	早稲田大学教授	工学博士 (大阪大学)	吉村 猛
	早稲田大学教授	工学博士 (早稲田大学)	松山 泰男
		Ph.D (スタンフォード大学)	
	北九州市立大学准教授	博士 (工学) (慶応大学)	奥田 正浩