

HOG 特徴量を用いた動画からの人検出の GPU による高速化

GPU acceleration of HOG algorithm for Human Detection from Videos

藤本 典幸, 大門 えり
Noriyuki Fujimoto and Eri Daimon

大阪府立大学 大学院理学系研究科 情報数理科学専攻 (〒 599-8531 大阪府堺市中区学園町 1-1)

Key Words: CUDA, human detection, video processing

一般物体認識とは、制約のない実世界シーンの画像に対して、画像を領域分割し、その各領域に対して物体カテゴリのラベルを付けることである [1]。

一般物体認識のための画像特徴量として、Histograms of Oriented Gradients(HOG) [2] が提案されている。HOG 特徴量は、回転やスケール変化に不変ではないが、局所的な幾何学的変化と明度変化に不変であり、そのため、大まかな物体形状を表現することが可能である。特に Dalal らは、HOG 特徴量を算出し、SVM (Support Vector Machine) を用いて人検出する方法を提案している [2]。HOG は高精度な人検出が可能であることが報告されているが、計算量が大きいという欠点がある。

そこで我々は、Dalal らの人検出アルゴリズムを CUDA を用いて高速化し、動画からのリアルタイム人検出の実現を目指した研究を行っている。本発表では、我々が実装した HOG の CUDA 実装について述べる。

2.93GHz Intel Core i3 530 CPU の 1 コア (SSE 命令は用いていない) と GeForce GTX 285 GPU を用いて評価を行った結果を図 1 と図 2 に示す。なお、OS は 32bit Windows XP Professional, コンパイラは Visual Studio 2005 Professional (コンパイルオプションは Release モードのデフォルト設定) を用いた。使用した CUDA のバージョンは 3.0, グラフィクスドライバのバージョンは 197.13 である。

図 1 は動画の解像度を様々に変化させた場合の動画 1 フレームの処理時間を CPU と GPU のそれぞれについて測定した結果である。この図より、GPU は CPU に比べて 100 倍程度高速に人検出ができることがわかる。

図 2 は携帯電話、ワンセグ、iPod、PSP、DVD の動画をリアルタイム処理できるための動画 1 フレームの処理時間の上限 (要求水準) と、それぞれの動画の 1 フレームを提案手法で処理するのにかかる時間を測定した結果である。この図より、携帯電話やワンセグ程度の動画であれば、現在の実装でも人検出をリアルタイム処理できることがわかる。しかし、DVD などの高解像度の動画をリアルタイム処理するには、さらなる改善が必要であることがわかる。

参考文献

- [1] 藤吉弘巨: 一般物体認識のための局所特徴量 (SIFT と HOG), 第 13 回映像メディア処理シンポジウム資料, pp.9-16, (2008)
- [2] N. Dalal and B. Triggs: Histograms of Oriented Gradients for Human Detection, proc.of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp.886-893 (2005)

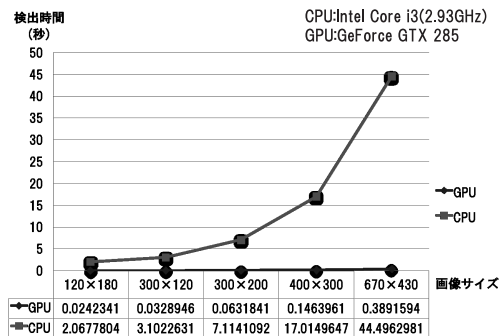


図 1: CPU に対する速度向上

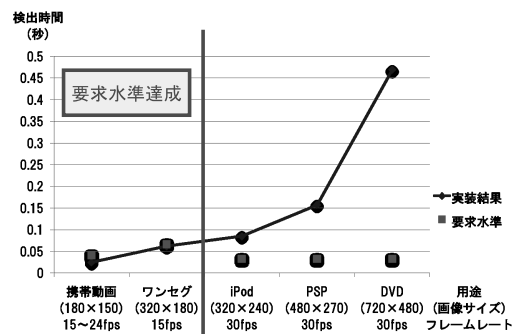


図 2: 1 フレームの人検出にかかる時間