

HOG 特徴量探索と時空間 MRF の連携による人物トラッキング

上條 俊介[†] 藤村 嘉一[†]

東京大学生産技術研究所[†]

1. まえがき

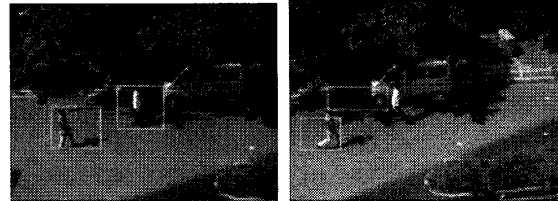
近年、駅構内や空港における監視目的から、不審者の自動追跡の要求が高まっている。また、歩行者は移動するため、追跡を広域的に行う技術が求められている。広域的な自動人物追跡には、カメラのパンニング技術によるカメラ内広域人物追跡も一つの重要な要素技術である。

著者らが開発してきた時空間 MRF[1]は、追跡対象・外部環境に依存せずに、オクルージョン問題を解決し、オクルージョンにロバストな移動体追跡を可能としている。しかし、空間 MRF モデルによる人物追跡では、カメラパンニング時において、時間経過に伴い背景オブジェクトに対して追跡人物オブジェクトの形状が同化してしまい、長時間の安定した人物追跡を行うことが困難となる問題が生じる。そこで、本稿では広域カメラネットワーク内のパンニングカメラ画像においても、安定した人物追跡を可能にするために、時空間 MRF による追跡に HOG 特徴量による歩行者オブジェクトの探索を組み合わせることにより、カメラパンニング中の安定した人物追跡手法を構築することを目的とする。

2. 時空間 MRF のパンニング画像への適用

時空間 MRF モデル[1]は、時空間画像の時間軸方向の相関関係に着目し、MRF モデルを時空間モデルとして拡張したものである。図 1 は、時空間 MRF モデルによるオブジェクトのトラッキングをパンニングカメラの画像に適用したものである。図 1 の(a)および(b)は、そのトラッキング結果を時系列にあらわしたものである。なお、画像は 10 フレーム/秒で取得されている。

カメラが固定されている状態では、画像シーケンスから静止背景画像が生成され、動いている人物等のオブジェクトのみが背景差分としてトラッキング対象となる。しかし図 3 のようにカメラがパンニングしている間には、パンニングしている背景画像も一つのオブジェクトとして、時空間 MRF によるトラッキングの対象となる。ここで、パンニング背景のオブジェクトは、



(a) frame 381 (b) frame 393

図 1 時空間 MRF モデルのパンニング画像への適用結果

前景の人物等のオブジェクトとまったく同等に扱われ、パンニング背景と人物との間で、時空間 MRF モデルによる領域分割が行われる。このように、時空間 MRF モデルは、固定カメラ画像のみならずパンニングカメラ画像にも適用可能である。しかし、処理を続けるうち、歩行者とパンニング背景との間の領域分割の境界が徐々に曖昧になりずれが生じる場合がある (図 3(b))。

3. HOG 特徴量探索による人物の検出

歩行者を含む画像領域は時空間 MRF により分割可能である。しかし、前述のようにその分割領域は歩行者の正確な領域から外れる可能性がある。この問題を解決するために、時空間 MRF により分割された歩行者であると思われる領域を ROI (region of interest) として、HOG 特徴量を用いたパターン認識を行うことで、正確な歩行者領域を抽出することを試みる。

HOG 特徴量は、画像におけるエッジ方向の空間分布を表すものである。本稿では、様々な背景画像を含む歩行者画像データ (MIT と INRIA のデータ) を SVM のトレーニングデータとして用いる。それぞれのデータは 64×128 [pixels] に成形される。これらの画像において、 8×8 [pixels] のセル単位で局所の HOG 特徴量が抽出される。各ピクセルにおける勾配方向情報は 64 スケールに量子化され、1 セルあたり 9 次元の bin にヒストグラムとしてプロットされる (9 次元ベクトルを形成する)。また、勾配強度は画像の照度や前景と背景とのコントラストにより変化するので、この影響を軽減するために、各ベクトルは正規化される。本稿では、この正規化の単位として 4 個の 8×8 [pixels] のセルを含む 16×16 [pixels] のブロックから得られる 36 次元のベクトルを定義する。この正規化ブロックをラスタスキャン順に 8

Pedestrian Tracking with Cooperation between ST-MRF and HOG

Shunsuke Kamijo[†] Kaichi Fujimura[†]

[†]Institute of Industrial Science, the University of Tokyo

[pixels] ずらして 7×15 ブロックのベクトル列を構成する。最終的には、学習画像に関して 105 個の正規化ベクトルをつなげた 3780 次元のベクトルが得られる。

時空間 MRF モデルから得られる歩行者のトラッキング結果は、各オブジェクトの識別子を与えられたブロック単位の分布を表すオブジェクトマップとして示される。時空間 MRF モデルにおけるブロックは 8×8 [pixels] で構成され、各ブロックには、画像上におけるその位置を占有する各オブジェクトの識別子が貼られる。これらの同一の識別子を持つ分布の境界相関を評価することで、それらのオブジェクトの矩形領域が得られる。最終的には、この矩形領域が ROI として選定され、ROI は 128 [pixels] の高さにスケールされる。なお、幅は 64 [pixels] である必要はない。スケール後の ROI は HOG/SVM により歩行者の領域であるかどうか検査される。本稿では、高さが 128, 120, 112 [pixels] である 3 種類の異なるスケールの学習データが用意され、それぞれのスケールに関して SVM による学習が行われる。これにより検出された領域は元のスケールに変換され、オブジェクトマップ上に反映される。

HOG/SVM により歩行者の領域が検出されない場合は、時空間 MRF モデルによるトラッキング結果が、そのフレームにおける最終的な歩行者検出結果として用いられる。このような問題は、パンニング背景画像が複雑なテクスチャのノイズを含む場合や、照度不足、オクルージョン環境下においてに生じやすい。

4. 実験結果

図 2 は、パンニングカメラにおけるトラッキング結果の例を示している。前述の図 1 が時空間 MRF モデルのみのトラッキング結果であるのに対し、図 2 は HOG による特徴量探索との協調アルゴリズムによる結果を示している。本実験の結果、時空間 MRF モデルのみの場合、トラッキング継続時間が平均 22 フレームであったのに対して、HOG/SVM との協調によれば、40~60 フレームの間トラッキングの継続に成功した。撮影上の制約から、撮像された歩行者の平均のパンニング時間が 51 フレームであったことから、40~60 フレームのラッキングの継続時間は、ほぼトラッキングの失敗がないことを意味している。なお、図 2 は、パンニング画像中でオクルージョンが発生している場合のトラッキング結果を示している。オクルージョンが発生した場合には、オクルージョンしている人物とオクル

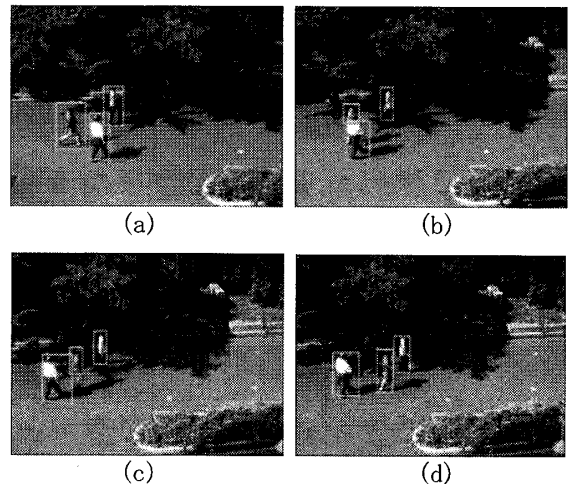


図 2 パンニングカメラでのトラッキング結果 (オクルージョン環境下)

ージョンされている人物の両者を包含する形で HOG 特徴量探索が行われ、この情報をもとにパンニング背景と前景の歩行者との領域境界が正しく訂正される。その上で、抽出された輪郭線内部においては時空間 MRF モデルにより、両歩行者の間の領域分割が行われる。実験全体における 74 サンプルのオクルージョン環境下でのパンニング画像を用いて実験を行ったところ、85.1%の成功率が得られた。

5. むすび

本稿では、時空間 MRF モデルによる人物追跡に、HOG 特徴量探索によるオブジェクト形状の回復を補完的に組み合わせることで、広域カメラネットワーク内のパンニングカメラ画像においても、安定した人物追跡を可能にする手法を提案した。実験の結果、本手法ではカメラのパンニング中にオクルージョンが発生している環境下においても、85%の人物追跡に成功した。また、従来の ST-MRF モデルによる手法で平均 22 フレームであったトラッキング継続時間は、40~60 フレームに向上したことを示した。今後の課題としては、今後の課題としては、アルゴリズムの更なる精度の改善が挙げられる。

謝辞

本成果は NICT 民間基盤技術研究促進制度に基づくものである。

参考文献

[1] S.Kamijo, Y.Matsushita, K.Ikeuchi, M.Sakauchi, "Occlusion Robust Vehicle Tracking utilizing Spatio-Temporal Markov Random Field Model", 7th World Congress on ITS, Torino, Nov. 2000.