

群衆パッチ学習に基づく人数推定 Number of People Estimation Based on Crowd-Patch Learning

池田 浩雄[†]
Hiroo Ikeda

大網 亮磨[†]
Ryoma Oami

1. はじめに

セキュリティやマーケティングの分野では、街頭や公共施設のような混雑状況下で、異常混雑の検知や群衆の流れの解析などのニーズが高まっている。特に、活用範囲が広い人の数/密度を推定する人数推定の技術が注目されている。

人数推定として、頭部を検出し集計することが考えられるが、監視カメラのような俯角が浅い場合には他人によって頭部の一部が隠され人数が推定できなくなる。俯角が浅い状況下を含むカメラ配置に対応できる従来技術として背景差分ベースの手法が提案されている [1]。背景差分によって前景領域を抽出し、前景領域と人数との関係を示す式に基づいて人数を推定する。この手法は、人物領域の背景への溶け込み、人物同士の重なり、人物以外の前景領域などの影響を受けやすい問題がある。また、従来技術として追跡ベースの手法が提案されている [2]。人物領域抽出の結果をフレーム間に対応づけて人物を追跡し特定のラインを超えた動線や動線そのものを集計して人数を推定する。この手法は、追跡への依存度が高く、フレームレートが下がると精度が低下する。

本稿では、合成群衆パッチを用いた Convolutional Neural Network (以下、CNN) 学習に基づく人数推定手法を提案する。提案手法では、人物同士の重なりを含めた群衆の状態を複数の人物画像を合成した局所画像 (群衆パッチ) で表し、群衆パッチを CNN によって回帰学習することで、1 枚画像から人数を推定する。

2. 群衆パッチ学習に基づく人数推定

群衆パッチ学習の特有な問題を示してから、その解決も含め群衆パッチ学習に基づく人数推定について述べる。

2.1. 群衆パッチ学習の問題

学習を行うには、多様なバリエーションの学習サンプルが必要である。学習サンプルにおいて群衆の状態を扱うには、特有な 2 つの問題がある。

1. 複数の人物が含まれる群衆状態の学習サンプルを大量に収集することが難しい
2. 人数/密度、人物同士の重なり方 (人物の配置) 等において、偏りのない群衆状態の学習サンプルを収集することが難しい

そこで、予め用意した人物画像から偏りのない様々な群衆状態の画像を合成し、学習サンプルである群衆パッチを大量に自動生成することで解決する。また、群衆状態を能動的に制御していることから、人数等の教師ラベルを後から付与する必要もなく、学習に要するコストを大幅に削減できる。

2.2. 合成群衆パッチの自動生成

合成群衆パッチの自動生成の構成を図 1 に示す。まず、予め背景画像、人物画像、人物領域画像に対応する人物領域画像の DB

を用意しておく。次に、背景画像 DB から背景画像を選択し、選択した背景画像から群衆パッチの背景を抽出する。次に、背景として扱う人物を背景に合成する。具体的には、人物画像 DB から人物画像を選択し、群衆状態を構成する頭部が、極端に基準サイズと異なる、もしくは、群衆パッチ内に納まらないように制御して背景に合成していく。さらに、群衆として扱う人物を合成する。人物画像 DB から人物画像を選択し、頭部が、基準サイズと同程度でパッチ内に納まるよう制御して背景に合成し、群衆パッチを生成する。そして、群衆パッチの生成と共に人数である教師ラベルを取得する。また、各画像選択はデータの偏りをなくす為、ランダムとした。

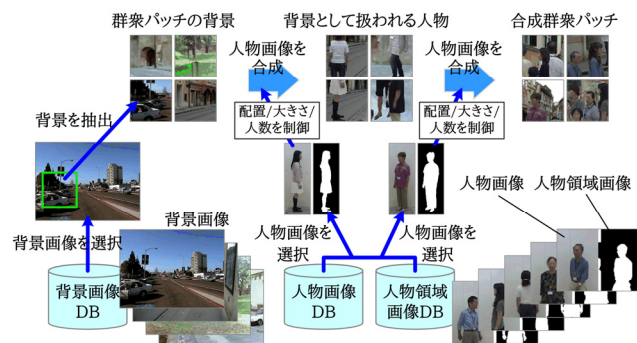


図 1 合成群衆パッチの自動生成

自動生成された合成群衆パッチの例を図 2 に示す。重なり方も再現されており、実画像と違和感がない。



図 2 自動生成された合成群衆パッチの例

2.3. 群衆パッチによる CNN 学習

自動生成された合成群衆パッチと教師ラベルである人数を CNN [3] によって回帰学習し、人数を推定する。本稿では、群衆状態という複雑なパターンを持つ画像を学習しなくてはならないことから、特徴抽出を考えず画像から直接学習が可能な CNN を用いる。

本稿で用いる CNN は、図 3 に示すように 5 階層で構成する。入力 is 自動生成された合成群衆パッチで 60×60 の輝度画像である。1 層目と 2 層目と 4 層目に対して 5×5 のフィルタを、3

[†] 日本電気(株) 情報・メディアプロセッシング研究所

層目に対して3×3のフィルタを用い、Convolution + Subsamplingの処理を行う。そして、最後の層で1ノードに統合する。

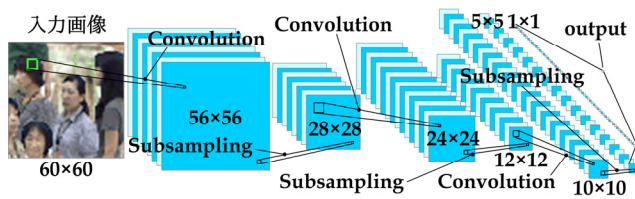


図3 CNNの構成

2.4. 人数推定

設定された探索窓に対し、CNNによって学習された識別器を用いて人数を推定する。推定を行う際の探索窓は、カメラの位置・姿勢を示すカメラパラメータに基づいて設定する。カメラパラメータから画像中の頭部サイズを求め、群衆パッチサイズにおける頭部の基準サイズを、求めた頭部サイズに合わせ探索窓を設定する。図4に探索窓の例を示す。探索窓毎に人数を推定し、統合することで画像中の特定の範囲の人数を推定する。

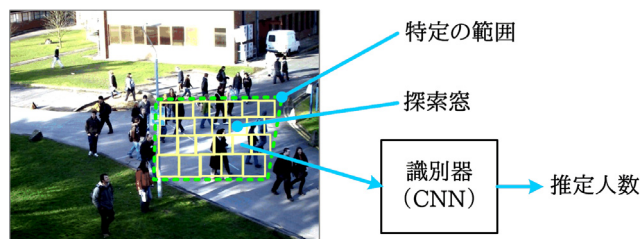


図4 人数推定を行う探索窓の例

3. 実験

3.1. 合成画像による群衆パッチ学習の効果

学習サンプルを、実画像から手動で集めて学習した場合と、合成画像を自動生成して学習した場合の精度を比較する。学習サンプルは互いに1200枚用意する。精度の指標として、正解人数に対する推定値の平均二乗誤差を用いる。図5に比較結果を示す。

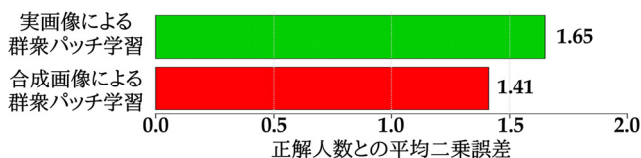


図5 実画像/合成画像による群衆パッチ学習の精度比較

評価の結果、平均二乗誤差で、実画像1.65、合成画像1.41（提案法）とほぼ同精度である。提案法は、合成にて学習サンプルを自由に増やせることから実画像を手動で集めて学習する場合より優位であるといえる。また、合成画像の精度が若干高いのは、手動で集めた実画像と異なり、ランダムな合成によって偏りなく網羅的に多様なデータを取得できたことに起因している。この面でも提案法が優位である。

3.2. 従来法との精度比較

従来法である背景差分ベースの手法と提案法の精度を比較する。学習サンプルは、互いに12万枚用意する。従来法は、12万枚の人物領域画像を基に、探索窓のサイズで正規化して面積率と人数の関係を回帰学習した。精度の指標として、上記に示す平均二乗誤差を用いる。図6に比較結果を示す。



図6 手法別の精度比較

評価の結果、平均二乗誤差で、従来法1.90に対して提案法0.25と約1/8に低減されることを確認した。従来法の精度低下の原因は、シルエットに起因する人物同士の重なり、頭部以外の部位の誤推定、などである。

さらに、両手法を照明変動がある映像で評価し比較した。推定結果を図7を示す。また、正解の人数も合わせて示す。

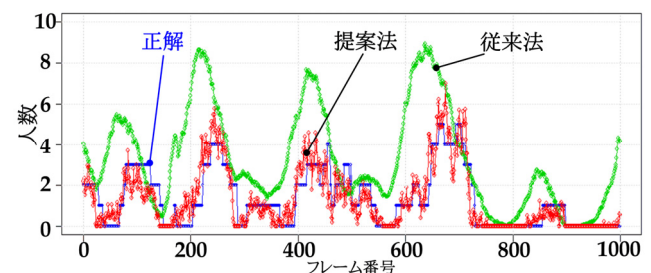


図7 手法別の映像における人数推定結果

全フレームにわたり、提案法が従来法に比べ正しく推定できていることが分かる。従来法は頭部以外の部位、影、直接的な照明変動等の影響による誤推定がある為、正解人数が0人の所でも誤って人数を推定するケースが多い。

4. おわりに

本稿では、合成群衆パッチのCNN学習に基づく人数推定手法を提案した。提案手法は、CNNによって学習された識別器により1枚画像からの人数推定を実現し、人物同士の重なりや人物以外の前景領域、等に影響されないことを示した。また、学習サンプルが集められない群衆パッチ学習の特有な問題を人物画像の合成によって解決した。今後は、より多くの環境に適用し、詳細な性能評価を行っていく。

参考文献

- [1] Ma,R.,et al., "On Pixel Count Based Crowd Density Estimation for Visual Surveillance", Cybernetics and Intelligent Systems, 2004 IEEE Conf., vol.1, pp. 170-173, 2004
- [2] Lui,X.,et al., "Detecting and Counting People in Surveillance Applications", Advanced Video and Signal Based Surveillance, IEEE Conf., pp.306-311, 2005
- [3] LeCun,Y.,et al., "Gradient-Based Learning Applied to Document Recognition", Proceedings of the IEEE 86.11, pp.2278-2324,1998.