H-019

CNN を用いた群衆パッチ学習に基づく人数推定の高精度化

Accuracy Improvement of People Counting Based on Crowd-Patch Learning Using Convolutional Neural Network

池田 浩雄† 大網 亮磨† 宮野博義†
Hiroo Ikeda Ryoma Oami Hiroyoshi Miyano

1. はじめに

近年、街頭や公共施設のような人が集まる場所に、数多くの防犯カメラが設置されている。セキュリティやマーケティングの分野では、既設カメラによる混雑状況での異常性の把握や群衆の流れ解析に対する強いニーズがある。これらの実現には、人数や密度の変化を捉えられる人数推定の技術が要求される。

そこで、1 枚画像から人数が推定できる、Convolutional Neural Network（以下、CNN）[1]を用いた合成群衆パッチ学習に基づく人数推定（以下、前手法） [2]を提案した。これにより、従来の背景差分や追跡ベースの手法[3][4]の問題を解決し、人物同士の重なりに強く、人物以外の前景に影響されない、フレームレートに非依存な人数推定を実現した。

しかし、多様な環境に適用するにつれ、2 つの問題が明らかになってきた。1 つは、照明やノイズの影響によって背景を群衆と誤推定し得ることである。推定精度に継続的に影響を与え、特に少人数ではその影響が高まる。もう 1 つは、人の見えが小さくなるような解像度の低下によって、推定精度が低下することである。

そこで、合成群衆パッチ学習に基づく人数推定の改良手法を提案する。1 つは、群衆パッチ内の輝度変動の大きさに応じて輝度の正規化法を変更することである。もう 1 つは、低解像度を含む多重解像度画像を用いた群衆パッチ学習である。これにより、多様な環境でのロバスト性を高め、推定精度を改善する。

2. 群衆パッチ学習に基づく人数推定

前手法では、シミュレーションにて自動生成された、合成群衆パッチと人数の教師ラベルを、CNN によって回帰学習し、人数を推定する。そして、頭部サイズを基に設定された推定窓毎に処理を行い、結果を統合して画像内の人数を推定する。

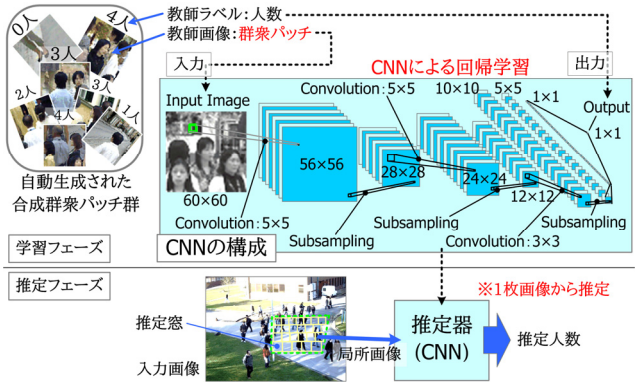

図 1 群衆パッチ学習に基づく人数推定

自動生成された合成群衆パッチは、俯角が浅いカメラ画角も考慮した、人物同士の重なりを含む画像である。学習サンプルを合成することで、網羅的な群衆状態の画像を大量に集められないという、群衆特有の問題を解決している。

3. 問題の詳細分析

本節では、精度低下の原因となる問題を詳細に分析する。

†NEC 情報・メディアプロセッシング研究所

3.1 背景の誤推定

照明の影響を吸収する為、群衆パッチ内の輝度 x_i をパッチ内の平均輝度 m と標準偏差 σ を用い、式(1)で正規化する。小さな σ で値が増幅され、ノイズが強調されやすい傾向にある。

$$x_i' = \frac{x_i - m}{\sigma} \quad (1)$$

そこで、群衆パッチ内の輝度変動に着目し、誤推定した背景の頻度を標準偏差毎に集計した。すると、図 2 のようになる。

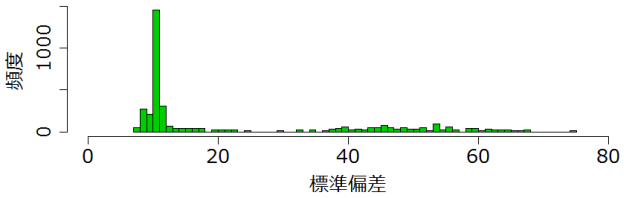

図 2 誤推定した背景に対する標準偏差の分布

標準偏差が一定以下の濃淡が少ない背景で誤推定が起こりやすいことが分かる。正規化によってノイズが過剰に強調され、想定外の画像パターンになっていると考えられる。解決方法として、濃淡が少ない背景の大量学習が考えられるが、符号化の影響等によるノイズの場合、符号化条件によって特性が大きく変わり得るため、汎用性の観点から学習するのは望ましくない。

3.2 低解像度による精度低下

群衆パッチの解像度と推定精度の関係に着目し、解像度に対する推定精度を評価する。評価結果を図 3 に示す。推定精度は正解人数と推定人数との平均二乗誤差（以下、MSE）で表す。

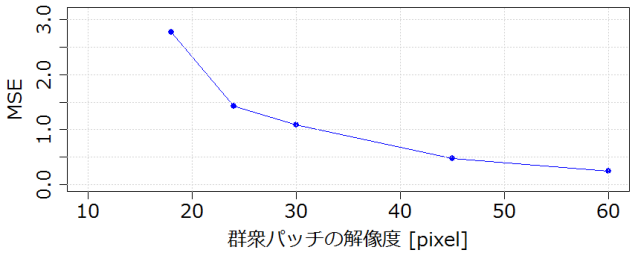

図 3 前手法（学習 12 万）の解像度に対する推定精度

群衆パッチ内で 1 人を区別できる MSE 0.25 以下の推定精度は 60×60pixel の解像度のみで達成され、それ以下の低解像度では未達成である。防犯カメラ画像では、頻繁に低解像度の人物も現れることから低解像度での精度改善が必須である。

4. 提案手法

4.1 輝度正規化法の改良

背景の誤推定改善の為、群衆パッチ内の輝度分散の大きさに応じて輝度の正規化法を変更することを提案する。提案手法は、一定以下の標準偏差で起こる過剰なノイズ強調を抑える為、

式(2)のように一定以下の標準偏差を固定化して正規化する。ここで、σ_0 は固定化の為の定数である。

$$x_i' = \frac{x_i - m}{f(\sigma)} \qquad f(\sigma) = \max(\sigma, \sigma_0) \tag{2}$$

これにより、濃淡が少ない背景は過剰なノイズ強調が抑えられ、輝度分散が大きい濃淡が明確な背景や群衆画像は、通常通り、照明の影響を吸収する正規化が行われる。

4.2 多重解像度画像を用いた群衆パッチ学習

低解像度による精度低下を改善する為に、低解像度を含む多重解像度画像の学習と、さらに低解像度画像のみ学習した推定器との統合を提案する。多重解像度画像の学習では、60×60pixel の現群衆パッチと、現群衆パッチを低解像度に縮小し、元のサイズに戻して生成された多重解像度のパッチを同時に学習する。統合に用いる低解像度画像のみの学習では、現群衆パッチを低解像度に縮小してそのまま学習する。本稿では、多重解像を 30×30pixel のみとした。学習には CNN の回帰学習を用い、CNN の構成には、多重解像度画像の学習に図 1 を、低解像度画像のみの学習に図 4 を用いる。

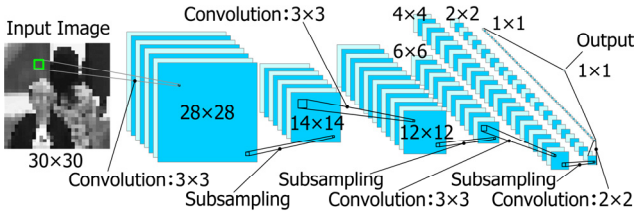

図 4 低解像度画像のみの学習に用いる CNN の構成

推定器の統合では、式(3)によって、各推定器の推定人数 y_i を重み付き平均する。重み w_i は入力の推定窓のサイズに合わせて切り替える。n は推定器の数である。実験では、推定窓のサイズに対して 0-30pixel、30-60pixel、60-∞pixel という 3 つの区間を用意し、3 つの重みセットを学習サンプルから決めた。

$$\bar{y} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \tag{3}$$

これにより、異なる性質の推定器の統合からも、精度改善を図っている。なお、低解像度のみ学習した推定器だけでは、高解像度の情報が欠け、高解像度の画像で精度が低下した。

5. 実験

提案手法の効果を評価する。学習に用いる合成群衆パッチは 50 万枚とする。

5.1 輝度正規化法の改良による効果

100 人程度の混雑映像に人数推定を行い、輝度正規化法の改良前後で推定精度を比較評価する。

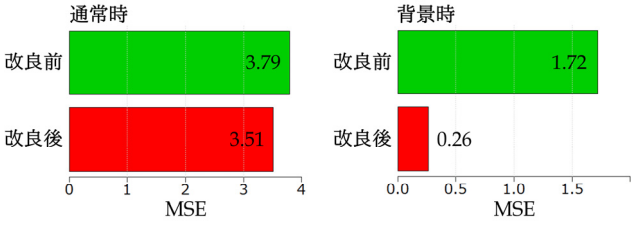

図 5 改良前後における推定精度の比較

評価の指標は画像上の人数に対する MSE で、通常時と背景時(0 人時)で評価する。図 5 に評価結果を示す。

通常時の推定精度を維持しながら、背景時の推定精度が改善され、提案手法が背景の誤推定に有効であることが分かる。

5.2 群衆パッチ学習の多重解像度化による効果

解像度別の評価 DB に対して人数推定を行い、提案法の導入前後で推定精度を比較評価する。図 6 に評価結果を示す。

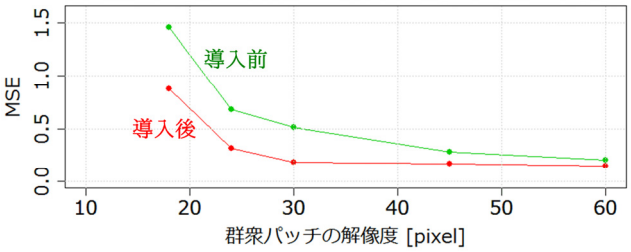

図 6 導入前後における解像度別の推定精度の比較

各解像度で推定精度が改善されている。特に、30 pixel の低解像度でも高解像度と同等の精度を実現しており、低解像度での精度改善に有効であるといえる。

5.3 前手法との推定精度の比較

低解像度を含む評価 DB に対して、前手法と推定精度を比較評価する。前手法は、学習 12 万(前回)と学習 50 万の 2 つを評価する。図 7 に評価結果を示す。

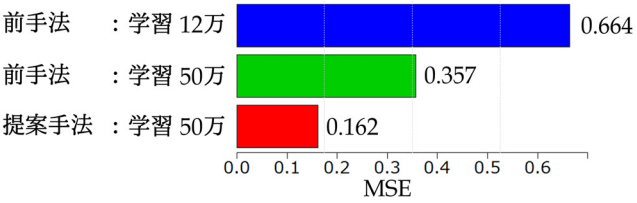

図 7 前手法との推定精度の比較

提案手法の MSE は学習 12 万の前手法から約 1/4 に低減される。これにより、前手法に比べ、より低解像度及び多様な背景に適用範囲を拡大できる。

6. おわりに

本稿では、CNN を用いた合成群衆パッチ学習に基づく人数推定の改良手法を提案した。提案手法は、群衆パッチ内の輝度変動の大きさに応じて輝度の正規化法を変更することで、背景の誤推定の低減を実現した。また、低解像度を含む多重解像度画像を用いた群衆パッチ学習によって、低解像度による精度低下を改善した。これにより、低解像度や背景という条件で多様な環境でのロバスト性を高めた。今後は、さらなる低解像度への対応や、多くの実環境による精度評価を行っていく。

参考文献

[1] LeCun,Y.,et al, “Gradient-Based Learning Applied to Document Recognition”, Proceedings of the IEEE 86.11, pp.2278-2324,1998.
[2] 池田浩雄,大網亮磨, “群衆パッチ学習に基づく人数推定”, 第 12 回情報科学技術フォーラム(FIT2013), 第 3 分冊, pp129-130, 2013.9.
[3] 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
[4] Lui,X.,et al., “Detecting and Counting People in Surveillance Applications”, Advanced Video and Signal Based Surveillance, IEEE Conf., pp.306-311, 2005