

輝度勾配方向によるテンプレートマッチングを用いた 高速な人物探索手法に基づく人物検出

Human Detection Based on Fast Scan Method Using Template Matching with Intensity Gradients

学 有江 誠 (中央大/JST CREST) 生形 徹 (中央大/JST CREST)
○ 柴田 雅聡 (中央大) モロ アレサンドロ (中央大/JST CREST)
正 寺林 賢司 (中央大/JST CREST) 正 梅田 和昇 (中央大/JST CREST)

Makoto ARIE, Chuo University / CREST, JST, arie@sensor.mech.chuo-u.ac.jp

Toru UBUKATA, Chuo University / CREST, JST, ubukata@sensor.mech.chuo-u.ac.jp

Masatoshi SHIBATA, Chuo University, shibata@sensor.mech.chuo-u.ac.jp

Alessandro MORO, Chuo University / CREST, JST, alessandro.moro@stud.units.it

Kenji TERABAYASHI, Chuo University / CREST, JST, terabayashi@mech.chuo-u.ac.jp

Kazunori UMEDA, Chuo University / CREST, JST, umeda@mech.chuo-u.ac.jp

In this paper, we proposed a fast and stable human detection based on “subtraction stereo” which can measure distance information of foreground regions. Scanning an input image by detection window is controlled in their window sizes and number using the distance information obtained from subtraction stereo. This control can skip a large number of detection windows and leads to reduce the computational time and false detection for fast and stable human detection. Additionally, we propose template matching using intensity gradients and aSC descriptor which is local features with upper human body models. Experimental results show that the proposal is faster and less false detection than method described in the reference [1].

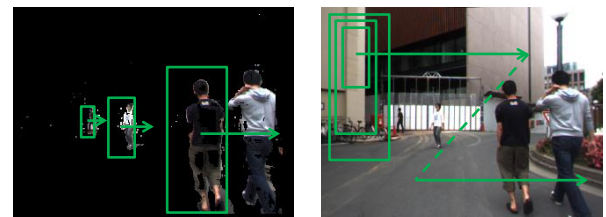
Key Words: Stereo Camera, Human Detection, Intensity gradients, aSC descriptors

1. 序論

近年、監視カメラからの不審者検出や人流計測のために、カメラ画像から自動的に人物をリアルタイムで検出することが期待されている。しかし、人物の様々な外観やオクルージョンの問題から人物検出は非常に困難である。人物検出精度の高い手法の1つとして、Dalalら[1]によって提案されたHOG (Histograms of Oriented Gradients)特徴量に基づく人物検出がある。しかし、この人物検出手法では、(1) 特徴量の算出、(2) 入力画像全体に検出ウィンドウを走査、(3) 異なるサイズの検出ウィンドウを複数回走査の過程において多くの計算時間を必要としている。(2)の検出ウィンドウの画像全体への走査は、計算コストの問題に加え、検出ウィンドウの数が増えることで人物に形状の似た背景領域での誤検出が増加し、検出精度の低下をまねく。処理速度と検出精度を向上させるためには、背景領域を走査する検出ウィンドウを削減する必要がある。また、(3)の大きさを変えた検出ウィンドウの複数回走査は、画像中における人物の大きさが分からないために必要となる過程で、遠近距離の人物が同時に存在する監視カメラ用途などでは様々な大きさを試みる必要があり、処理速度の低下を引き起こす。検出ウィンドウの大きさをあらかじめ知ることができれば、計算量の削減が期待できる。

そこで本論文では、前景領域の距離情報を少ない計算量で取得することのできる差分ステレオ[2]を用いて、検出ウィンドウの効率的な走査手法を提案する。差分ステレオにより得られる距離情報に基づき、画像中の前景領域に適した検出ウィンドウのサイズを決定する。

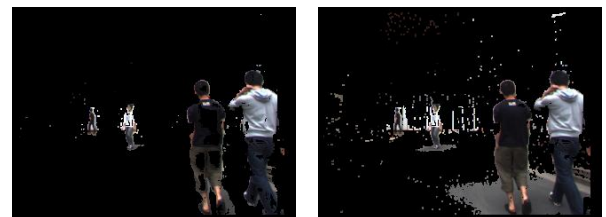
またHOG特徴量は、人物の大まかな形状を表現することが可能である一方、処理時間がかかるという問題がある。混雑環境下で検出ウィンドウ数が多い際にHOG特徴量を算出すると、リアルタイムでの人物検出が困難である。そのため、本論文では、人物の大まかな形状を表現しつつ算出時間の少ない輝度勾配方向を用いたテンプレートマッチングを提案する。



(a) Proposed

(b) Previous

Fig.1 Scan method of detection window



(a) With shadow detection

(b) Without shadow detection

Fig.2 Background subtraction with shadow detection

また、カメラから人物までの距離が近い場合や人物同士のオクルージョンにより、人物の下半身が画像内に映らず、テンプレートマッチングによる人物検出精度が低下する問題がある。そこで本論文では、局所領域の人物の形状を2値で表すaSC (approximated Shape Context)特徴量[3]による人物特有の上半身形状を用いた人物検出手法を提案する。

2. 差分ステレオを用いた検出ウィンドウの走査

本論文では、差分ステレオを用いた高速で誤検出の少ない検出ウィンドウの走査手法を提案する。人物検出を行うとき入力画像の前景領域にのみ検出ウィンドウを走査する。また、前景領域の距離情報に基づいて、検出ウィンドウの大きさを適切に決めることにより、ウィンドウサイズを変えた複数回

の走査をせずに、検出精度の向上を目指す(図 1(a)). 従来手法 [1]と違い人物画像全体に対して複数サイズの検出ウィンドウを走査(図 1(b))しないため、計算時間と誤検出の削減ができる。また、図 2(a) に示すような人物や物体の影を除去することで人物検出の精度は更に向上する。前景領域の抽出手法と影検出手法、ウィンドウの走査手法を以下に示す。

2.1 差分ステレオ

差分ステレオ[2]は、背景差分法によって入力画像から前景領域を抽出し、左右カメラの前景領域をマッチングすることで距離情報を得る。この手法は、対応点探索を左右カメラの前景領域に限定することができ、計算量の削減とともに、マッチングのロバスト化を行える。また、前景領域の実際の高さと幅を計測することができるため、検出ウィンドウの大きさを適切に決定することが期待できる。

2.2 影検出

図 2(b) のように影の影響で前景領域に人物ではない領域が含まれる。それに伴って、人物までの距離計測に誤差が生じる可能性がある。そこで、前景から人物や物体の影を除去するために影検出[4]を行なう。

2.3 検出ウィンドウサイズの動的変化

検出ウィンドウの走査領域を前景に絞り、差分ステレオより得られる距離情報からウィンドウのサイズを動的に変化させる手法を提案する。図 1(a)のように背景差分から得られる前景領域に動作領域を絞ることで、輝度勾配の算出時間と誤検出が増加する問題を解決することができると思われる。

また、画像内の距離による人物の大きさ、位置に対応するため、異なるサイズの検出ウィンドウを複数回走査し、計算時間が増加する問題が生じる。この問題に対して、画像内の前景領域を走査する検出ウィンドウのサイズを距離やカメラの仰角の情報から動的に求め解決する。まず、走査する検出ウィンドウのサイズを前景領域の距離情報から求める。次に、擬似透視投影を仮定し、検出ウィンドウの高さを補正する。弱透視投影では、距離情報のみで検出ウィンドウの高さ補正を行うが、擬似透視投影では、カメラの高さと仰角、画像内の人物の位置を考慮して検出ウィンドウの高さ補正を行っている。以上の理由から、擬似透視投影を仮定する。

監視カメラのように高所からカメラの光軸を向下きに設置した場合、画面内での人物の位置によって、近くでは小さくなり、遠くでは大きくなるというように見え方が異なる。本手法では、人物の高さと幅の比例定数 k_h, k_w と差分ステレオより得られる前景領域の距離情報を用いて、擬似透視投影を仮定した式(1), (2)より検出ウィンドウのサイズの補正を行う。

$$height = \frac{k_h}{W_Y} (\cos \theta - y \sin \theta) \quad (1)$$

$$width = \frac{k_w}{C_Z} \quad (2)$$

ここで、 W_Y は世界座標系でのカメラから前景領域までの距離、 C_Z はカメラ座標系でのカメラから前景領域までの距離、 θ はカメラの仰角、 y は焦点距離を1に正規化した縦方向の画像座標である。以上の手法を用いることで、図.1(a)のように画面内の前景領域に対して検出ウィンドウを密に走査し、走査回数を1回にすることができる。

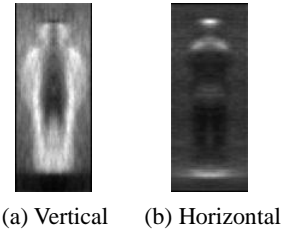


Fig.3 Intensity gradient

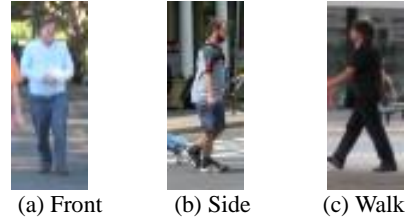


Fig.4 Positive sample for each template

3. テンプレートマッチングを用いた人物検出

本論文でのテンプレートマッチングで用いるテンプレートは、人物検出に有効な輪郭形状を表現するものであり、輝度勾配の縦方向のみに着目することで、従来手法の HOG 特徴量に対して少ない計算量で算出可能な点の特徴である。更なる計算時間の削減のため提案手法では、2次元のテンプレート画像での1次元方向で輝度勾配強度を累積した特徴量を用いる。これは、入力画像とテンプレートの類似度計算の際に、従来手法のように各ピクセルからではなく、テンプレート画像での縦軸と横軸それぞれの輝度勾配強度を累積した特徴量である。これ以降、この特徴量は累積輝度勾配特徴量とする。

3.1 輝度勾配方向を用いたテンプレート

入力画像との類似度計算に用いるテンプレートは、人物のサンプル画像の輝度勾配方向と強度から構築し、人物の大まかな輪郭形状を表現する。従来手法の HOG 特徴量は、9方向の輝度勾配を用いているが、本論文では縦方向の輝度勾配のみからテンプレートを構築する。以下の式(3), (4)より輝度勾配の方向と強度を算出する。

$$orientation(i, j) = \tan^{-1} (I_j(i, j) / I_i(i, j)) \quad (3)$$

$$magnitude(i, j) = \sqrt{I_i^2(i, j) + I_j^2(i, j)} \quad (4)$$

ここで、 $I_i(i, j)$ は横方向の輝度勾配、 $I_j(i, j)$ は縦方向の輝度勾配であり、それぞれ注目画素の左右、上下のピクセルで差分を行い算出したものである。本論文では、輝度勾配方向の $45^\circ \sim 135^\circ$ である縦方向のみを用いる。この縦方向での勾配強度を累積することで図 3(a) のようなテンプレート画像を構築する。図 3(a) のテンプレートは、図 3(b) の横方向の輝度勾配のみで構築された画像と比べて、人物の大まかな輪郭が表現されているのが分かる。テンプレートマッチングを用いて人物検出を行う際には、人物の輪郭が表現されている図 3(a) のみを用いる。

本論文では、人物の歩行状態を3種類に分けている。3種類に分けている理由は、人物の典型的な輪郭形状が3種類だからである。具体的には、図.4に示す(a) 人物の正面または背面、(b) 人物の側面、(c) 歩行の3種類である。これらの歩行状態毎の人物サンプル画像を1000枚ずつ用いて3枚のテンプレートを構築する。

3.2 テンプレートマッチング

検出ウィンドウ毎に算出した縦方向の輝度勾配情報と、前節にて算出したテンプレート画像の類似度を計算することで、テンプレートマッチングによる人物検出を行う。一般にテンプレートマッチングは、テンプレート画像と入力画像内の各ピクセルから得られる類似度を用いて検出するが、各ピクセルではマッチングの計算時間がかかる。また、同一の画像ではなく大まかな形状の輝度勾配を入力画像から探索するため、従来手法では類似度が小さくなる可能性がある。そこで、本論文のテンプレートマッチングは、テンプレートと入力画像内の各ピクセルから類似度計算を行わず、テンプレートから得られる累積輝度勾配特徴量を用いて類似度計算を行う。累積輝度勾配特徴量は、図.5 に示すテンプレート画像の縦軸と横軸を bin とした 2 つの特徴量である。累積輝度勾配特徴量の横軸は、テンプレート画像の縦、横軸での座標を示している。また、縦軸はそれぞれの座標での輝度勾配強度を累積し、0～1 に正規化したものである。

構築したテンプレートと対象画像からの類似度算出には、正規化相互相関を用いる。正規化相互相関算出には、テンプレートから作成した勾配強度を累積した特徴量を用いる。正規化相互相関は以下の式(5), (6), (7)より算出する。

$$R_{ZNCC}(i) = \frac{\sum_{i=0}^{M-1} ((I(i) - \bar{I})(T(i) - \bar{T}))}{\sqrt{\sum_{i=0}^{M-1} (I(i) - \bar{I})^2 \times \sum_{i=0}^{M-1} (T(i) - \bar{T})^2}} \quad (6)$$

$$R_{ZNCC}(j) = \frac{\sum_{j=0}^{N-1} ((I(j) - \bar{I})(T(j) - \bar{T}))}{\sqrt{\sum_{j=0}^{N-1} (I(j) - \bar{I})^2 \times \sum_{j=0}^{N-1} (T(j) - \bar{T})^2}} \quad (7)$$

$$R_{ZNCC} = (R_{ZNCC}(i) + R_{ZNCC}(j))/2 \quad (8)$$

ここで、 I は入力画像の勾配強度を累積した特徴量、 T はテンプレート画像の勾配強度を累積した特徴量、 M は、 i 軸での bin の総数、 N は j 軸での bin の総数、 i は i 軸の bin である、 j は j 軸の bin である。

人物検出は R_{ZNCC} の値に閾値処理を行う。 R_{ZNCC} の値は -1.0～1.0 の間の値をとり、1.0 に近づけばテンプレートに類似していることを示している。本研究では、経験的に人物判定の閾値を 0.3 としている。人物の歩行状態別の 3 つのテンプレートのうち、1 つでも閾値以上であれば人物と判定する。

4. aSC 特徴量を用いた人物検出

人物の下半身が画像内に映らない場合や人物同士のオクルージョンへ対応するために、局所領域から大まかな形状を表すことができる特徴量を用いた人物検出について検討する。

3 章で述べたテンプレートマッチングを用いた人物検出では、画像内に人物全体が映っている場合のみ人物検出が可能である。これは、監視カメラのように高所からカメラの光軸を下向きに設置した場合、画面内での人物の位置に近いほど下半身が映らず、人物検出の精度が低下するからである。そこで、人物特有の上半身形状から局所特徴量を算出し、テンプレートマッチングでは行うことで困難なシーンでの人物検出精度の向上を目指す。

本論文では、上半身形状を表す特徴量として aSC 特徴量を用いる。aSC 特徴量[3]は人物全体ではなく、局所領域で大まかな形状を表すことができる。また、次元数も少なく 2 値画像から容易に算出できるため、テンプレートマッチングを用

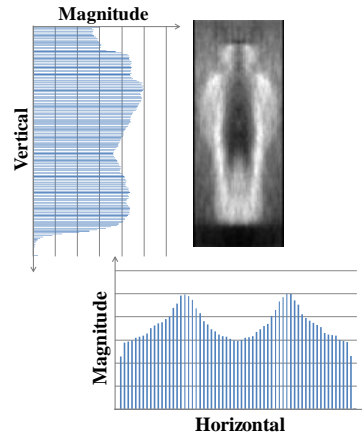


Fig.5 Histogram obtained from intensity gradients of a template image



(a) Color image (b) Binary image (c) aSC descriptor

Fig.6 Outline of the aSC descriptors

いた人物検出手法と組み合わせが可能である。

テンプレートマッチングでの正規化相互相関の値が人物判定の閾値以下の場合、aSC 特徴量を用いた人物検出と組み合わせ、人物検出性能の高い人物検出手法を提案する。

4.1 aSC 特徴量

aSC 特徴量[3]とは、図.5 のように人物の画像に対して前景と背景を分けた 2 値画像から算出する特徴量であり、3x3 のセルから構成される。具体的には、それぞれのセル内で前景と背景でのピクセル数の大小で 2 値化を行い、パターン化したものである。人物領域周辺から算出する特徴量で、人物のサンプル画像を多く必要とするが、学習が容易に行えるという特徴がある。

4.2 Codebook 構築

人物の上半身形状を表現するために、本論文では局所特徴量である aSC 特徴量を用いている。1 つの aSC 特徴量では、人物の上半身を表現できないため、aSC 特徴量の集合より人物の上半身形状全体を表現する。この aSC 特徴量の集合を Codebook とする。

3.1 節の図.3(b) のように、輝度勾配方向での横方向の輝度値は、人物の肩や頭等の上半身領域に出やすいという特徴がある。そこで、人物のサンプル画像から輝度勾配の横方向での勾配強度を算出し、勾配強度が高い領域から aSC 特徴量を算出する。人物の上半身の肩や頭周辺の領域から算出された aSC 特徴量のみで Codebook を構築することができる。この Codebook 内の aSC 特徴量を用いて検出ウィンドウ内の aSC 特徴量が人物か否かを判定する。提案手法では、Codebook 構築に用いる aSC 特徴量を 15 種類としている。

検出ウィンドウ内から 15 種類の aSC 特徴量を算出し、Codebook のものと半数以上一致するならば検出ウィンドウ内の前景領域を人物と判定する。検出ウィンドウ内からの aSC 特徴量算出は、Codebook 構築時と同様で輝度勾配の横方向で

の勾配強度が高い領域から行なう。理由としては、検出ウィンドウ内から領域を固定して aSC 特徴量の算出を行うと、前景と背景の境界である人物領域周辺から算出されないことがあるためである。

4.3 テンプレートマッチングと aSC 特徴量の組み合わせ

テンプレートマッチングを用いた人物検出の検出精度向上のため、テンプレートマッチングと aSC 特徴量を組み合わせた人物検出手法を提案する。テンプレートマッチングでは、正規化相互相関の閾値が 0.3 以上の検出ウィンドウを人物と判定しているが、0~0.3 の前景領域も人物である可能性がある。そのため、これら検出ウィンドウに対して、aSC 特徴量を用いた人物検出を行い、テンプレートマッチングでの人物検出で未検出扱いになった前景領域に対して再び人物検出を行う。

5. 実験

提案手法の有効性を検証するために人物検出精度評価実験を行った。3章で述べたテンプレートマッチング、4章で述べた aSC 特徴量、両手法の組み合わせでの人物検出精度を比較した。全ての手法で差分ステレオによる検出ウィンドウの削減とウィンドウサイズの動的変化手法を用いる。人物検出精度を比較する3つの手法に、HOG 特徴量に基づく人物検出[1]を比較対象として加える。

5.1 実験条件

本実験では、1000 フレーム分で人物検出精度評価を行った。従来手法では、検出ウィンドウを画像全体に対して密に走査する。実験では検出ウィンドウのサイズを 30×60 [pixel] とした。提案手法は、2章で述べた差分ステレオを用いたウィンドウの走査手法を用いた。

人物検出結果の検証には、以下の4つの評価指標を用いた。

- True Positive rate (T.P.): 真値の人数のうち、システムが人物を検出した割合
- False Positive rate (F.P.): システムが人物でないものを人物として検出する割合
- Precision: システムが検出したものが人物である割合
- Processing Time (P.T.): 画像1枚にかかる処理時間

実験に使用したステレオカメラは Bumblebee2 (Point Grey Research, カラー, f=3.8mm, VGA, 48fps), 処理には CPU が Intel Core2 Duo (3.06GHz), メモリ 6GB RAM を搭載したノート PC を使用した。

テンプレート構築には、NICTA[5]のデータセットを用いた。また、aSC 特徴量を用いた Codebook 構築には2値化した人物の画像を用いるため、独自のデータセットを用いた。この2値化画像は、100枚用いた。

5.2 実験結果

人物検出精度の結果を表1に示す。表1より差分ステレオによるウィンドウの走査手法を用いた人物検出では、画像全体でのウィンドウの走査手法を用いた人物検出よりも P.T.の性能が向上しているのが分かる。また、テンプレートマッチングでの検出手法とテンプレートマッチングと aSC 特徴量を組み合わせた検出手法では、F.P.の性能が良いことが分かる。これは、人物の形状を細かく表現する HOG 特徴量ではなく、輝度勾配方向や aSC 特徴量においても、提案する走査手法を用いることで人物検出には関係のない背景情報を除去できたからである。さらに、HOG 特徴量は、特徴量の次元数が多く

Table 1 Performance comparison between the proposal and reference [1]

	T.P. (%)	F.P. (%)	Precision (%)	P.T. (ms)
Reference [1] (30×60)	79.4	9.6	85.9	158.3
Template matching (Proposed scan)	72.3	8.2	89.8	8.2
aSC descriptors (Proposed scan)	81.4	12.8	86.4	11.8
Template + aSC (Full scan)	82.9	15.0	84.7	64.8
Template + aSC (Proposed scan)	80.5	7.9	91.1	16.2



(a) Template Matching (b) aSC descriptors (c) Template + aSC

処理時間が膨大になるが、本論文で用いる輝度勾配や aSC 特徴量は容易に算出でき処理時間も少ないため、P.T.の性能が向上した。

テンプレートマッチングでは図7のように画面下端での下半身がほとんど映っていない人物に対しての T.P.の性能が低下した。これは、人物全体が映っているサンプル画像から勾配強度を累積した特徴量を構築しているからである。そこで、上半身の形状を特徴量とした aSC 特徴量と組み合わせることで、カメラの設置条件に関係なく T.P.の性能が向上することが表1, 図7(c)より分かる。

6. 結論

本論文では輝度勾配方向を用いた、少ない処理時間で人物の大まかな形状を表すテンプレートマッチングによる人物検出手法を提案した。また、テンプレートマッチングと局所領域の人物形状を表現できる aSC 特徴量を組み合わせた、高速度かつ高精度な人物検出手法を実現した。

今後の展望としては、テンプレートの枚数を増加させ、更なる人物検出精度の向上が挙げられる。

文献

- [1] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection," *CVPR, Vol. 1*, pp.886-893, June 2005.
- [2] K. Umeda, et al, "Subtraction Stereo - A Stereo Camera System That Focuses On Moving Regions -," *Proc. Of SPIE-&T Electronic Imaging, Vol. 7239 Three-Dimensional Imaging Metrology*, 723908, 2009.
- [3] C. Beleznaï, "Fast Human Detection in Crowded Scene by Contour Integration and Local Shape Estimation," *CVPR*, pp.2246-2253, June, 2009.
- [4] A. Moro, et al, "Auto-adaptive threshold and shadow detection approaches for pedestrian detection," *In Proc. AWSVCI*, pp.9-12, 2009.
- [5] http://www.nicta.com.au/research/projects/AutoMap/computer_vision_datasets.