

Smart Window Transform とエッジベース識別器に基づく人検出

日立製作所 日立研究所 李媛, 伊藤誠也, 三好雅則, 中部大学 藤吉弘巨, 北海道大学 金子俊一

Human Detection using Smart Window Transform and Edge-based Classifier

Hitachi Ltd., Hitachi Research Laboratory Yuan Li, Masaya Itoh, Masanori Miyoshi
Chubu University Hironobu Fujiyoshi, Hokkaido University Shun'ichi Kaneko

Almost human detection systems are based on classification of rectangular sub-windows, which are enumerated from 2D image. Therefore, the detection rate of such systems is affected by image distortion and titled orientation of the human in images. To overcome this problem, a new method called Smart Window Transform is proposed. This method creates a Smart Window for an observed object in 3D space, which is paralleled to the observed camera image plan, then, transforms Smart Window into 2D space, using camera intrinsic and extrinsic parameters and finally projects to a rectangular sub-window, specified by a Classifier. As experiments results, human detection rates using Smart Window Transform and Edge-based Classifier were successfully improved compared with the one without Smart Window Transform.

1. はじめに

安全・安心な社会作りのため、異常行動検知、人物追跡などの知的な監視システムが求められている。このような監視システムを実現するには、監視対象である人をロバストかつ高精度に検出する人検出技術が重要である。

近年では、人検出技術として画像全体に矩形領域を走査し、各画像位置に人が含まれるか否かを識別する手法が多く提案されている。例えば、HOG²⁾などの局所特徴量と統計的な学習法の組み合わせを用いて、画像をラスタスキャンすることにより高精度に人を検出する手法が知られている。しかし、これらの学習サンプルに含まれる人は直立した姿勢のものであり、走査した矩形領域内の人が、学習サンプルと異なる姿勢を取る場合、認識性能が低下するという問題がある。また、カメラパラメータによっては、画像の回転やレンズ歪みの影響から、人の形状に歪みを生じるため、認識性能が低下するという問題もある。これらの問題を解決するために、学習サンプルに歪みを考慮したデータを含める方法が考えられる。しかし、この方法は多くのサンプルを要するため、現実的ではない。また、仮に多くのサンプルを用意したとしても、学習サンプルに多様な形状が混在するため、認識性能を十分に上げることができないためである。

そこで、本研究は従来の問題点を改善するために、画像上に存在する人の領域を、カメラパラメータに基づき識別器が容易な形状で表現する Smart Window Transform を提案する。提案手法を人検出手法に適用することにより、画像上の人々の状態(歪み、傾き)に応じた、高精度な人検出ができることを評価実験により確認した。

2. Smart Window Transform

本章では、Smart Window Transform(以下 SWT)の定義及び特徴について述べる。SWT は 3次元空間上に定義した人が存在する空間を、カメラパラメータに応じて画像空間上で観測される領域で表現し、識別器に適合する形状(例えば、矩形)に変換する手法である。図1を用いて、SWTの処理手順について説明する。ここでは、カメラモデルはピンホールカメラを利用する。

まず、3次元空間上に Smart Window(以降 SW)を設定する手順について述べる。人の存在する空間は、3次元空間上の位

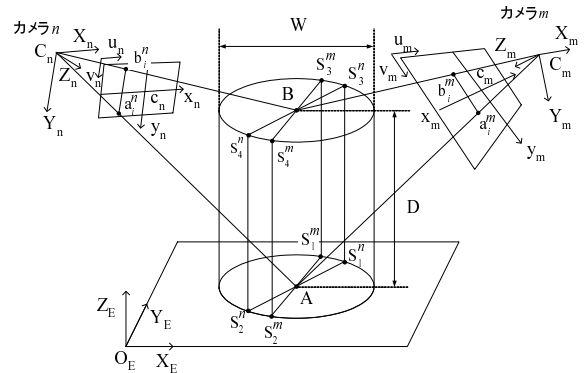


Fig. 1 Smart Window Transform の説明図

置 A を底面の中心とする円柱(直径 W , 高さ D)で定義する。3次元空間上にカメラ m が存在するとき、カメラ m と正対し、空間 i の中心軸を通る切断面を、カメラ m の SW として $\{S_j^m | j = 1, 2, 3, 4\}$ と表現する。 S_j^m 上の 1 点である S_1^m を、3次元空間上の位置 A から、カメラ m に定義した X_m 軸方向に $\frac{W}{2}$ 移動した点として式(1)に定義する。

$$S_1^m = M_m A + \begin{bmatrix} \frac{1}{2}W & 0 & 0 \end{bmatrix}^T \quad (1)$$

ここで、 M_m は 3次元空間 $O_E-X_E Y_E Z_E$ とカメラ座標系 $C_m-X_m Y_m Z_m$ との変換行列である。その他の頂点についても、式(1)と同様に表し、また、カメラ n の場合も $\{S_j^n | j = 1, 2, 3, 4\}$ として表現することができる。次に、3次元空間で定義したカメラ m の SW $\{S_j^m\}$ を、カメラ画像上に座標変換する。カメラ画像上での SW を $\{s_j^m | j = 1, 2, 3, 4\}$ とすれば、 S_j^m と s_j^m の間に式(2)の関係が成り立つ。

$$k s_j^m = H_m P_m S_j^m, \quad j = 1, 2, 3, 4 \quad (2)$$

ここで、 k はスカラー、 P_m は、カメラ座標系 $C_m-X_m Y_m Z_m$ と投影平面 $c_m-x_m y_m$ 、 H_m は、投影平面 $c_m-x_m y_m$ と画像座標系 $o_m-u_m v_m$ の変換行列である。なお、これらの変換行列は、Tsai の手法¹⁾を用いたカメラキャリブレーションにより求める。

最後に、カメラ画像上の SW を、識別に適した形に正規化する方法について述べる。まず、SW $\{s_j^m\}$ の頂点座標 (u_j^m, v_j^m) と、特徴抽出のために予め設定した矩形画像の頂点座標 (u_j, v_j)

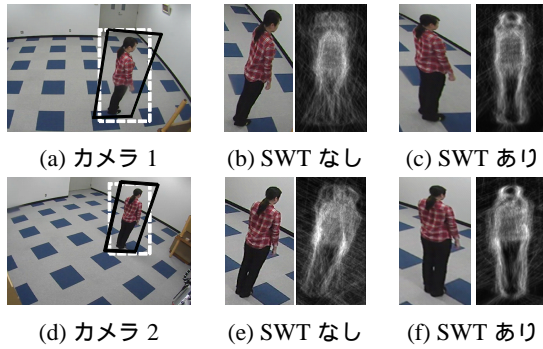


Fig. 2 設置条件の異なるカメラ画像の例



Fig. 3 評価データ

の4組の対応点から、透視変換行列を求める。式(3)に (u_j^m, v_j^m) と (u_j, v_j) の関係式を示す。

$$u_j = \frac{c_{00}u_j^m + c_{01}v_j^m + c_{02}}{c_{20}u_j^m + c_{21}v_j^m + c_{22}}, j = 1, 2, 3, 4$$

$$v_j = \frac{c_{10}u_j^m + c_{11}v_j^m + c_{12}}{c_{20}u_j^m + c_{21}v_j^m + c_{22}}, j = 1, 2, 3, 4 \quad (3)$$

ここで得られた4組の関係式より、透視変換行列の各要素 $\{c_{pq}|p, q = 1, 2, 3\}$ を求め、カメラ画像上のSW $\{s_j^m\}$ を特徴抽出及び識別器に適した矩形に透視変換する。

以上の処理を本研究ではSWTと呼ぶ。

図2に示した設置条件の異なるカメラ画像を比較し、SWTの特徴について述べる。

- 図2(a)(d)に示した実線は、3次元空間上で算出したSWを、各カメラのパラメータに応じてカメラ画像上に描画したものである。位置が異なるカメラであっても、カメラ画像上でのSWは、カメラと正対する平面に変換されていることが確認できる。
- 図2(b)(c)及び(e)(f)は、SWTを適用しない場合(実線)及び適用した場合(点線)の抽出画像と、様々な位置における抽出画像100枚の平均勾配を示す。SWTを適用しない従来の方法(図2(b)(e))では、カメラとの位置関係により人物形状に歪みや傾きが生じるため、平均勾配画像が不明瞭である。一方、SWTを適用した場合(図2(c)(f))は、平均勾配画像が人物に近い形状を示し、画像の特徴を識別に適した形で取得可能なことが確認できる。

3. SWTとエッジベース識別器に基づく人検出

本章では、SWTとエッジベース識別器による人検出手法を提案し、実画像を用いた評価実験を通して、提案手法の有効性について述べる。

3.1 エッジベース識別器

本研究では、SWTにより人の輪郭情報を安定に取得可能であるため、エッジベース識別器による人検出手法を用いる。ここでは、三井らが提案したJointHOG特徴を用いた人検出手法³⁾をエッジベース識別器として採用する。この手法は、HOG特徴では表現できない対象的な輪郭情報を取得可能なため、高精度に人検出が可能である。しかし、第1章で述べたように、カメラ設置状況によっては人の輪郭形状にも歪みが生じるため、検出率が低下する問題がある。そこで、本研究で提案するSWTを適用することにより、様々なカメラ設置状況に対して安定な人検出手法を提案する。

3.2 評価実験

提案手法の有効性を確認するため、従来の人検出手法(JointHOG)と本手法(SWT+JointHOG)を図3に示すカメラパラメータの異なる評価データを用いて比較する。まず、評価データから、従来のラスタスキャン法とSWTを用いて、人を含む画像(Positive)と含まない画像(Negative)を目視により抽出する。次に、作成した画像に対して、画像内に人が存在するか否かを文献3)の手法により識別する。評価指標は、一般に使われる検出率=Recallと誤検出率=1-Precisionを用いる。

Table 1 人検出結果の比較

Set	Positive (枚)	Negative (枚)	従来手法 (SWTなし)(%)		本手法 (SWTあり)(%)	
			検出率	誤検出率	検出率	誤検出率
1	60	54	88.3	22.1	96.7	32.6
2	56	54	50.0	26.3	96.4	31.6
3	106	160	84.9	2.2	91.5	0.0
4	94	160	80.9	8.4	92.6	1.1

表1は人検出の結果比較であり、識別器の閾値は全て同様に設定した。まず、Set1とSet2を用いてカメラの回転による性能を比較する。従来手法は、カメラの回転に伴う人の形状変化により、検出率が88.3%から50%まで大幅に低下したが、本手法は95%以上の高い検出率を示している。一方で、本手法の誤検出率が従来手法と比べて高く、これは、SWTにより背景に存在する物体に対しても直立した輪郭が取得されるためと考えられる。次に、カメラの位置が異なるSet3とSet4を比較すると、本手法では検出率、誤検出率ともに性能が改善したことがわかる。これは、カメラ位置の違いをSWTにより吸収した効果であり、主に画像の端付近に存在する人物の検出性能が向上した。

4. まとめ

本稿では、カメラの設置状況によって人検出性能が低下する問題に対して、カメラパラメータに基づき、識別に適切な形状に変換するSmart Window Transformを提案し、実画像による評価によりその有効性を示した。今後は、本手法を行動認識や人物追跡などのアプリケーションに応用するため、実環境での実験を行う予定である。

参考文献

- 1) R.Y.Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV camera and lenses", IEEE Journal of Robotics and Automation, Vol.RA-3, No.4, pp. 323-344, 1987.
- 2) Navneet Dalal and Bill Triggs, "Histograms of Oriented Gradients for Human Detection", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. II, pp. 886-893, 2005.
- 3) 三井相和, 山内悠嗣, 藤吉弘巨, "Joint特徴量を用いた2段階 Boostingによる物体検出", 電子情報通信学会論文誌, Vol. J92-D, No. 9, pp. 1591-1601, 2009.