

## JointHOG を用いたプライバシー保護のための任意姿勢の人物領域検出手法

武田洋<sup>†</sup> 北澤仁志<sup>†</sup>東京農工大学<sup>†</sup>

## 1 はじめに

近年、防犯カメラの増加に伴い、プライバシー保護を目的とした人物領域マスキングの研究が進められている。Google などでは、顔にのみマスキングを施していることが多いが、プライバシー保護には不十分である。一方、車や草木の揺れに対してもマスキングを施すと、監視カメラとしての価値が損なわれてしまう。従って、向きや姿勢にかかわらず人物領域のみを検出しマスキングを施すことが望ましい。

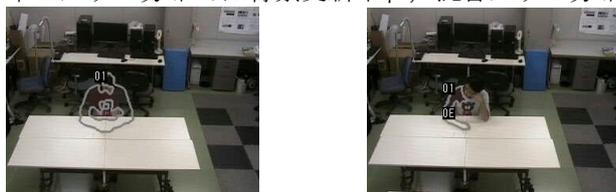
現在、人物領域検出には JointHOG[1] と Real AdaBoost, Support Vector Machine (SVM) [2] などを用いた手法が提案されているが、何れも立位の人物の検出であり、席に座っている(座位)人物や横になっている(臥位)人物など、任意姿勢の人物領域を高精度に検出する手法はまだ確立されていない。

そこで、本研究では、立位のみでなく、座位や臥位など複数のテンプレートを用い、どのような姿勢でも高精度に人物領域を検出し、マスキングを施す手法を提案する。

## 2 固定カメラにおけるプライバシー保護の問題点

本研究では、プライバシー保護は「人物は特定できないが、どのような行動をとっているかは特定できる」ことで満たせるものとした。監視カメラにおいて、行動まで隠蔽してしまうと映像中から、不審な行動とそうでないものの区別ができなくなってしまうからである。

また、本研究では、前景領域検出には単一ガウス分布近似と混合ガウス分布近似に基づく背景差分法を用いたが、この方法では、動画像中で停止した人物が時間経過とともに背景に吸収され、マスキングが外れる(図 1)。また、動画像中で静止していた人物が動き出した跡(ゴースト)を抽出し続けるといった問題点がある。そこで、抽出された領域に対して、単一ガウス分布では背景更新率  $\rho$ 、混合ガウス分布



(a) 抽出直後

(b) 数分後

図 1: 背景差分のみでの抽出

では学習率  $\alpha$  を切り替え、人物と判定された領域の更新率を下げ、逆に人物以外と判定された領域の更新率は上げることで、上記の問題解決を試みる。

## 3 JointHOG を用いた任意姿勢の人物検出

HOG[1]は検出された領域をセルに分割し、各セルにおいて、勾配強度と勾配方向からヒストグラムを作成することで特徴量とする。JointHOG は、複数の HOG をセル毎に組み合わせ、1 段階目の学習により、識別に有効な特徴の組み合わせを選択する。

本研究では HOG を算出する際、画像サイズは  $64 \times 64$  ピクセル、1 セル  $8 \times 8$  ピクセルとし、ブロックは  $3 \times 3$  セルとした。正規化は中心ブロックのみに行い、外周 1 ブロックは使用しないものとした。また、JointHOG の次元数はセルの組み合わせ数となるので、 ${}_{36}C_2 = 630$  とした。

## 4 姿勢テンプレート

学習用データセットと評価用データセットを用意する。データセットは、立位、座位、臥位、跳位(ジャンプしている人物)の 4 つの姿勢の人物と、人物以外の 5 つを用意した。

座位、臥位、跳位については、固定カメラで屋内を撮影し、人物が中心になるよう  $64 \times 64$  ピクセルで切り出したものを用いた。立位については、文献[3]よりダウンロードし、同一サイズに修正した。学習用データセットとして、各姿勢 1000 枚、人物以外を 5635 枚、評価用データセットとして、各姿勢 1000 枚、人物以外を 2695 枚用意した。図 2 にデータセットの一部を示す。



図 2: データセット

## 5 実験結果

## 5.1 学習方法

特徴量には JointHOG (630 次元)を用い、1 段階目の識別において Real AdaBoost により 20 次元または、50 次元を選択した。2 段階目の識別器には

Various Pose Human Area Detection for Privacy Protection Using JointHOG and Machine Learning

Yo Takeda, Hitoshi Kitazawa

Tokyo University of Agriculture and Technology

SVM(cost:1000, gamma:0.1) 及び Real AdaBoost (学習回数:1000 回, 閾値:0.0) を使用した. 精度評価には accuracy, 人物を人物以外とした False Negative(FN), 人物以外を人物とした False Positive(FP)を用いた.

5.2 姿勢テンプレートの組み合わせ

識別結果を表 1, 2 に示す. 表の左端は検査画像中の Positive とした姿勢を指す. また, 複数の姿勢を混ぜた場合, RealAdaBoost, SVM とともに同じような結果となったので, 複数姿勢を混ぜた場合は Real AdaBoost のみ示す. 表を見ると, 1 つの姿勢のみを Positive とした場合は 95%程度の精度で識別できているが, 複数の姿勢を混ぜると, 精度が低下していた. これは, 複数の姿勢を混ぜることで, 識別に有効な共通した特徴量の選択が困難になったためであると考えられる. そこで, 次項の検出実験では単一の姿勢で構築した識別器を並列に用い, いずれかで検出されれば人物とした.

表 1: Real AdaBoost での実験結果

1段階目 学習回数	n120							n150								
	FP						FN	accuracy	FP						FN	accuracy
	立位	座位	臥位	跳位	人物以外	合計			立位	座位	臥位	跳位	人物以外	合計		
立位	0	1	0	0	19	20	130	97.76%	0	0	0	12	12	186	97.04%	
座位	0	2	23	60	85	268	94.73%	0	1	1	41	43	321	94.56%		
臥位	0	0	0	0	0	1	99.99%	0	0	0	0	0	0	0	100.00%	
跳位	0	1	0	2	3	347	94.77%	0	0	0	0	0	0	485	92.76%	
立位&座位			3	71	125	199	443	90.41%		0	20	70	90	457	91.83%	
立位&臥位		5		1	33	39	176	96.79%		2		2	24	28	197	96.64%
立位&跳位			17	3	44	64	425	92.70%		3	0		28	31	549	91.34%
座位&臥位	0				96	126	222	334	91.70%	0		47	86	133	319	93.25%
座位&跳位	0		9		158	167	409	91.40%	0	3		93	96	527	90.69%	
臥位&跳位	0	19			15	34	283	95.27%	0	9		5	14	317	95.06%	
立位&座位&臥位					128	186	314	420	89.04%			79	137	216	445	90.13%
立位&座位&跳位				9	218	227	568	88.13%			3		147	150	642	88.17%
立位&臥位&跳位					60	105	380	92.76%		19			39	58	461	92.25%
座位&臥位&跳位	1				209	210	368	91.37%	0				145	145	424	91.50%
全姿勢					255	255	460	89.32%					201	201	497	89.57%

表 2: SVM での実験結果

1段階目 学習回数	n120							n150								
	FP						FN	accuracy	FP						FN	accuracy
	立位	座位	臥位	跳位	人物以外	合計			立位	座位	臥位	跳位	人物以外	合計		
立位	0	0	0	0	9	9	175	97.25%	0	0	0	7	7	279	95.73%	
座位	0	0	2	49	51	374	93.65%	0	0	0	20	20	479	92.55%		
臥位	0	0	0	0	0	8	99.88%	0	0	0	0	0	10	10	99.85%	
跳位	0	0	0	4	4	414	93.76%	0	0	0	0	0	479	92.85%		

5.3 任意姿勢の人物領域検出実験

背景差分により抽出された移動物体に対して, 5.2 より構築した識別器を用いて, 固定カメラで撮影した動画像 (解像度:320×240) 中から人物領域検出実験を行う. 実験シーンには, 歩いてきて席に着き, 数分後席から離れてフレーム外へ出ていく 602 フレームのデータを用いた.

JointHOG の算出は, 背景差分により検出された領域の長辺を一辺とした正方形に整形し, 64×64 ピクセルにリサイズし算出した. また, 1 段階目の学習回数は 20 回のみで行った. 精度評価は人物領

域を誤って人物以外と判定した割合, Unmasking Person Rate(UPR)を用いた (式 1) .

$$UPR = \frac{FN}{TP + FN} \times 100[\%] \quad (1)$$

5.4 実験結果

表 3, 図 3 に単一ガウス分布での結果を, 図 4 に混合ガウス分布での結果の一部を示す. 単一ガウス分布を用いた場合は, 更新率を切り替えることで, 人物領域にマスキングし, 影などの誤抽出を削除できた. Real AdaBoost では, 図 3 下段に示すような抽出漏れがあったが, SVM では検出できた. 一方, 混合ガウス分布を用いた場合は, 影などの誤抽出は少なかったが, 人物領域がすぐに背景に吸収され, 検出されなくなった. したがって, このような状況下では単一ガウス分布による背景差分のほうが適していると考えられる.

表 3: 単一ガウス分布

	TP				FN	UPR
	立位	座位	臥位	跳位		
Real Adaboost	539	566	198	423	32	5.32%
SVM	602	602	211	601	0	0.00%



図 3: 単一ガウス分布 (ρ=0.0005/0.000005)



図 4: 混合ガウス分布 (α=0.1/0.000001)

6 まとめと今後の課題

JointHOG と機械学習を適用することで, 座位や臥位など任意姿勢の人物について精度よく識別・検出できることが分かった.

今後の課題としては, 今回の実験で検出漏れがあったような姿勢テンプレートを追加することによる精度向上と, ゴーストと人物, 人物同士が重なっているような状況での検出手法を模索していく.

参考文献

[1]三井相和, 他 “Joint 特徴量を用いた 2 段階...”, IEICE D, Vol. J92-D No. 9 pp. 1591-1601  
 [2]山内悠嗣, 他 “画像からの統計...”, 信学論, 2013, D, 9.  
 [3] [http://www.vision.cs.chubu.ac.jp/joint\\_hog/](http://www.vision.cs.chubu.ac.jp/joint_hog/)