

2台のカメラを用いた複数人物を対象とした 単体追跡システムの構築

齋藤, 哲平 / SAITO, Teppei

(発行年 / Year)

2011-03-24

(学位授与年月日 / Date of Granted)

2011-03-24

(学位名 / Degree Name)

修士(工学)

(学位授与機関 / Degree Grantor)

法政大学 (Hosei University)

2010年度 修士論文

2台のカメラを用いた複数人物を
対象とした単体追跡システムの構築

指導教授 赤松 茂 教授

法政大学工学研究科
システム工学専攻

09R6111

齋藤 哲平

概要

本研究では、2 台のカメラを用いた複数人物を対象とした自動追跡システムを構築した。2 台のカメラの内、1 台は複数人数を同時に検知・追跡を行い、もう 1 台でその複数人の中から指定した 1 人をより細かく追跡することにより複数人物を対象とした単体追跡を実現した。主な手法として、照明や影に頑健な HOG(Histograms of Oriented Gradients)を特徴値とし、RealAdaBoost を用いて識別機を構成して追跡を行った。本研究で実験を行った結果、複数人物を対象とした単体追跡を可能にすることができた。

キーワード: 人物追跡, HOG , RealAdaBoost

Abstract

This paper proposes an algorithm for achieving automatic, stable, accurate, and real-time detection and tracking of moving objects, such as pedestrians in a scene, using two cameras for image acquisition. One camera takes charge of the tracking of the multiple people. And, the other camera takes charge of the tracking for the specified object in multiple people. To deal with the environment of brightness and color variation, HOG (Histograms of Oriented Gradients) is proposed as feature value of human tracking. In this approach, a RealAdaBoost class detects initial regions of objects. In this paper, the detection accuracy and the tracking performance by the proposed algorithm were investigated by experiments in pedestrian tracking tasks. The result of my experiment clearly shows that the tracking for the specified object in multiple people was able to be enabled.

Key Words: *Human tracking, HOG , RealAdaBoost*

目次

概要

Abstract

第 1 章	はじめに	・・・ 1
第 2 章	システム構成	
2.1	システム概要	・・・ 3
2.2	PTZ カメラと CCD カメラ	・・・ 4
2.3	画像処理ソフト	・・・ 7
第 3 章	移動物体検出方法	
3.1	背景差分法	・・・ 10
3.2	フレーム間差分法	・・・ 11
3.3	モフォロジー処理	・・・ 12
3.3.1	収縮・膨張処理	・・・ 12
3.3.2	モフォロジー処理	・・・ 12
3.3.3	本システムに用いたモフォロジー処理	・・・ 13
3.4	Histograms of Oriented Gradients	・・・ 14
3.5	RealAdaBoost	・・・ 16

第 4 章 移動物体追跡方法	
4.1 提案手法	・・・ 18
4.2 複数追跡手法	・・・ 19
4.3 単体追跡手法	・・・ 20
4.3.1 ナンバリング	・・・ 20
4.3.2 カメラ制御	・・・ 22
第 5 章 実験	
5.1 学習実験	・・・ 24
5.1.1 実験結果	・・・ 26
5.1.2 考察	・・・ 27
5.2 検出実験	・・・ 28
5.2.1 実験結果	・・・ 29
5.2.2 考察	・・・ 31
5.3 追跡実験	・・・ 32
5.3.1 実験結果	・・・ 33
5.3.2 考察	・・・ 36
第 6 章 むすび	・・・ 37
謝辞	・・・ 38
参考文献	・・・ 39

第1章 はじめに

現代社会において、映像監視システムの必要性はますます高まっている。2001年に発生したアメリカ同時多発テロ事件以降、現代社会の人々のセキュリティに対する関心が飛躍的に高まり、テロ事件に対して世界各国の空港をはじめとした公共交通機関においてセキュリティに対する必要性が飛躍的に高まった。これに伴い、銀行・百貨店・駐車場・コンビニエンスストアなど、我々の日常生活の近くでも映像監視システムの需要が高まってきている。これはセキュリティに対する需要が高まった他に、カメラやシステムなどが安価になったことにより、個人でも映像監視システムを導入することができるほどになったためである。映像監視システムの導入は、事件や犯罪の発見のみならず、このようなことを未然に防ぐという抑止の効果も期待されている。

防犯用途以外の映像監視システムの需要としては、工場の製造ライン監視、原子力発電所や火力発電所などの人が立ち入れない場所の異常監視、ダム、河川、火山などの状況の監視・記録、人物のジェスチャや姿勢の認識、人物が倒れるなどの異常行動の検知、顔認識の前処理などがある。これら従来のシステムの多くは、監視員がモニターの映像を見続けて異常行動などを監視する必要がある、コンビネーションカメラのカメラ制御などもモニターを見ながら監視員が手動で操作する必要があった。これは、異常行動などが無い場合などが多い映像監視システムにおいて、監視員への負担は大変大きいものであった。そこで近年では、人物追跡や異常行動検知などを、画像処理を用いて自動化しようということが試みられている[1]。異常があった場合に監視員に警告することにより、監視員は常に映像監視システムのモニターを見る必要がなくなり、監視員への負担やコストは大幅に軽減されることになる。

近年、Histograms of Oriented Gradients (以降より HOG と略) [2] [3]や、Scale Invariant Feature Transform (以降より SIFT と略) [4]といった、画像の輝度勾配をヒストグラム化する画像特徴抽出法が流行しており、物体認識アルゴリズムのデファクター・スタンダードの1つとなりつつある。しかし、HOGなどの手法は広範囲にわたる複数人物を対象とした追跡に優れ、人物の大きさによる追跡性能の変動を受けやすいため、個人をより細かく追跡するためには Zoom 機能や Mean-Shift[5]といった手法をとる必要がある。

そこで本研究では、装置の設置やコストの面を考慮し、2 台のコンビネーションカメラ、Pan・Tilt・Zoom カメラ（以降より PTZ カメラと略）と Charge・Coupled・Device カメラ（以降より CCD カメラと略）を用いた複数の移動物体を対象とした追跡システムの構築を目標にし、自動で安定かつ高性能・高精度、実時間でも動作が可能なシステムを目指した。

今回は、移動物体の検出・追跡においてオーソドックスな手法であるフレーム間差分法と SIFT 特徴量などに比べ、計算コストに優れ、照明や影、局所的な幾何学的変化に頑健な HOG 特徴量、そして単純な識別器を組み合わせることによって精度の高い識別機を構成する Boosting 法である Real AdaBoost を組み合わせた手法を用いた。

これらを用いて、複数人物を対象とした単体人物の追跡の実験を行い、提案するシステムの有効性について検討する。

第2章 システム構成

2.1 システム概要

本研究でのシステム構成は以下の通りである.

OS : Windows XP

CPU : Intel Core2 Duo 3.33GHz

メモリ : 4GB

画像処理ソフト : HALCON

画像入力ボード : Picport color (Letron vision) ×2

PTZ カメラ : Sony EVI-D100 ×1

CCD カメラ : Edmund Optics CV-S3200 ×1

本研究では装置及び設置のコストを考え, 2台のカメラ, PTZ カメラと CCD カメラを用いて人物自動追跡システムを構築した. またいろいろな手法を試し, 最適な手法を見つけるため, プログラミング時間の短縮を図るために画像処理ソフト HALCON を使用した.

2台のカメラについては2.2で, 画像処理ソフト HALCON については2.3で解説する.



図 2.1 PTZ カメラ



図 2.2 CCD カメラ



図 2.3 Picport color

2.2 PTZ カメラと CCD カメラ

今回、2台のカメラを用いた。PTZ カメラは図 2.1 に示した Sony EVI-D100、CCD カメラは図 2.2 に示した Edmund Optics CV-S3200 である。PTZ カメラは、高速パン・チルトと広角ズームレンズ内蔵のパン・チルト・ズーム一体型カラービデオカメラである。Sony が開発したコンシューマカムコーダなどを制御するプロトコル VISCA を用い、カメラを RS-232C 接続し、シリアル通信によりコンピュータからカメラを制御する。RS-232C を通じて、ホスト側から厳密に設定、制御することが可能である。基本的な性能は以下の通りである。

<基本性能>

- ・画素数：41 万画素
- ・パン・チルト機能：水平 $\pm 100^\circ$ ，最大速度 $300^\circ/\text{秒}$
垂直 $\pm 25^\circ$ ，最大速度 $125^\circ/\text{秒}$

また、RS-232C のパラメータは以下の通り。

<RS-232C のパラメータ>

- ・通信速度：9600bps
- ・データ長：8 ビット
- ・スタートビット：1 ビット
- ・ストップビット：1 ビット
- ・パリティなし

VISCA の通信形式とパケット構造について述べる。VISCA 通信の基本単位をパケットと呼ぶ（図 2.4）。パケットの最初のバイトはヘッダーと呼び、差し出し宛て先のアドレスが入っている。例えば、アドレス 0 のコントローラーからアドレス 1 の EVI-D100 へ送るパケットのヘッダーは 16 進数で 81H となる。アドレス 2 の EVI-D100 へ送るパケットは 82H となる。コマンドリスト表例（表 2.1）ではヘッダーが 8X となっているので、X の部分に EVI-D100 のアドレスを入れる。また、アドレス 1 の EVI-D100 からの応答パケットのヘッダーは 90H になる。アドレス 2 の EVI-D100 からのパケットは A0H となる。

一部の EVI-D100 設定用コマンドは一度にすべての機器に対して送ることができる（ブロードキャスト）。ブロードキャストの場合はヘッダーを 16 進数で 88H とする。ターミネーターは FFH でパケットの終わりを示す。

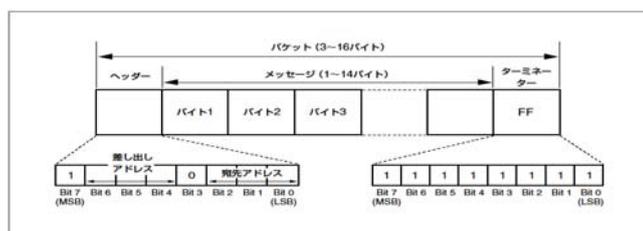


図 2.4 パケットの構造

表 2.1 コマンドリスト表の例(PanTilt)

Command Set	Command	Command Packet	Comments
<u>Pan-tiltDrive</u>	Up	8x 01 06 01 VV WW 03 01 FF	VV: Pan 速度設定 01 (低速) ~ 18 (高速)
	Down	8x 01 06 01 VV WW 03 02 FF	WW: Tilt 速度設定 01 (低速) ~ 14 (高速)
	Left	8x 01 06 01 VV WW 01 03 FF	<u>YYYY: Pan Position FA60 ~ 05A0 (center 0000)</u>
	Right	8x 01 06 01 VV WW 02 03 FF	<u>ZZZZ: Tilt Position FE98 ~ 0168 (center 0000)</u>
	UpLeft	8x 01 06 01 VV WW 01 01 FF	
	UpRight	8x 01 06 01 VV WW 02 01 FF	
	DownLeft	8x 01 06 01 VV WW 01 02 FF	
	DownRight	8x 01 06 01 VV WW 02 02 FF	
	Stop	8x 01 06 01 VV WW 03 03 FF	
	AbsolutePosition	8x 01 06 02 VV WW 0Y 0Y 0Y 0Y 0Z 0Z 0Z 0Z FF	
	RelativePosition	8x 01 06 03 VV WW 0Y 0Y 0Y 0Y 0Z 0Z 0Z 0Z FF	
	Home	8x 01 06 04 FF	
	Reset	8x 01 06 05 FF	

任意の位置にカメラを動かすために、

Absolute position : 8x 01 06 02 VV WW 0Y 0Y 0Y 0Y 0Z 0Z 0Z 0Z FF

絶対値駆動

YYYY:Pan Position FA60(-100 度) ~ 05A0 (100 度) (center 0000)

XXXX:Tilt Position FE98(-25 度) ~ 0168(25 度) (center 0000)

の[0Y 0Y 0Y 0Y][0Z 0Z 0Z 0Z]を任意の位置に決める指示を出せばよい。

Pan Position の FA60(-100 度) ~ 05A0 (100 度) は 16 進数で書かれている。カメラを 100 度動かしたい場合は、Absolute position の [0Y 0Y 0Y 0Y]の所に 05A0 を[00 05 10 00]のように入れてあげればよい。

水平方向の移動範囲±100 度を FA60 ~ 05A0 で表現しているので、動かしたい画素数が何度に相当するか逆算して、任意の位置にカメラを動かす。Position の絶対角度は表 2.2 のようになる。

表 2.2 絶対角度表

位置	-100度	-95度	-90度	-85度	-80度	-75度	-70度	-65度	-60度	-55度
16進数	FA60	FAA8	FAF0	FB38	FB80	FBC8	FC10	FC58	FCA0	FCE8
数値	64096	64168	64240	64312	64384	64456	64528	64600	64672	64744
位置	-50度	-45度	-40度	-35度	-30度	-25度	-20度	-15度	-10度	-5度
16進数	FD30	FD78	FDC0	FE08	FE50	FE98	FEE0	FF28	FF70	FFB8
数値	64816	64888	64960	65032	65104	65176	65248	65320	65392	65464

位置	+5度	+10度	+15度	+20度	+25度	+30度	+35度	+40度	+45度	+50度
16進数	0048	0090	00D8	0120	0168	01B0	01F8	0240	0288	02D0
数値	72	144	216	288	360	432	504	576	648	720
位置	+55度	+60度	+65度	+70度	+75度	+80度	+85度	+90度	+95度	+100度
16進数	0318	0360	03A8	03F0	0438	0480	04C8	0510	0558	05A0
数値	792	864	936	1008	1080	1152	1224	1296	1368	1440

表 2.2 を用いることで、任意の方向へカメラを向けることができる。本研究では、Pan Position に FE50(-30度) ~ 01B0 (30度), Tilt Position に FEE0(-20度) ~ 0120 (20度) の範囲で使用する。

表 2.3 コマンドリスト表の例(Zoom)

<u>CAM Zoom</u>	Stop	8x 01 04 07 00 FF	ズームコントロール
	Tele (Standard)	8x 01 04 07 02 FF	
	Wide (Standard)	8x 01 04 07 03 FF	
	Tele (Variable)	8x 01 04 07 2p FF	p: 速度設定 0 (低速) ~ 7 (高速), 8ステップ
	Wide (Variable)	8x 01 04 07 3p FF	
	Direct	8x 01 04 47 0p 0q 0r 0s FF	<u>pqrs: Zoom Position</u> <u>光学: 0000 (wide) ~ 4000 (tele)</u> <u>デジタル: 4000 (1倍) ~ 7000 (4倍)</u>
	D-Zoom On	8x 01 04 06 02 FF	デジタルズーム オン/オフ
D-Zoom Off	8x 01 04 06 03 FF		

また、Zoom 機能を使用する場合も同様に、
Direct: 8x 01 04 47 0p 0q 0r 0s FF
の[0p 0q 0r 0s]に任意の拡大率に決める指示を出せばよい。

2.3 画像処理ソフト

本研究では画像処理ソフトに HALCON を使用した。

HALCON とは、1200 を越える高機能画像処理ライブラリと、それをベースにマシンビジョンシステムを極めて効率的に構築できるラピッドプロトタイピングツール HDevelop で構成されている。マシンビジョンシステムとは、従来目視により行っていた各種検査業務を、コンピュータやカメラなどのハードウェアと画像処理ソフトウェアをベースにした画像処理システムで置き換えたシステムである。その適用範囲として、FA 分野での検査（外観検査、寸法計測など）、ロボットビジョン（位置決め制御）、画像認識、医療・材料画像解析、文字認識・確認、バーコードリーダ、3 次元計測、リモートセンシング、モニタリングから移動物体解析まで幅広い分野があげられる。図 2.5 に IC リード検査を例にマシンビジョンの構成例を示す。

この HDevelop を利用するだけでも、必要な画像処理ライブラリに対応する画像処理用オペレータ（スクリプト言語）をメニューから選択して並べるだけで、数 100 行のコードを要していた複雑なマシンビジョン（検査）用プログラムもたったの数行で構築できる。Windows NT または UNIX (X-Window) 対応の対話型開発ツール。HALCON はソフトウェアツールであり、任意の画像ボードとインターフェースがとれる。すぐにオンラインでの画像入力に基づくマシンビジョンシステムが開発できる。なお、画像入力ボードなしでも、ディスク内にあらかじめ保存した画像ファイル（TIFF や BMP フォーマット）から画像を読み込んで HALCON を利用することも可能である。

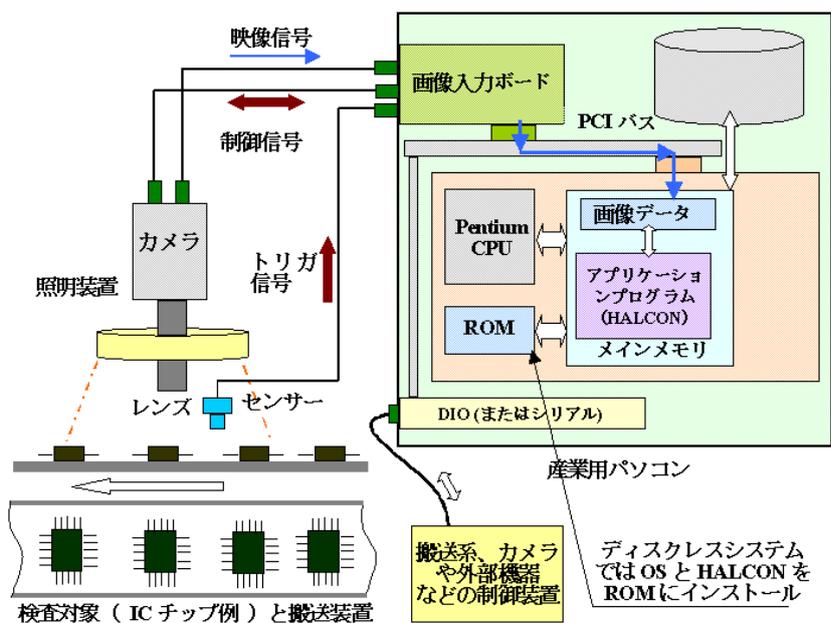


図 2.5 マシンビジョンシステム構成例 (IC リード検査)

以下に、HALCON の開発画面、モジュール構成を示す。

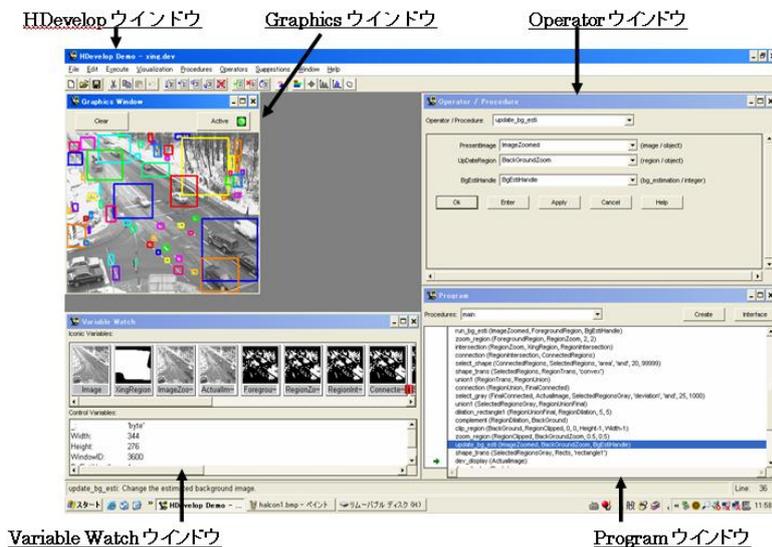


図 2.6 HALCON 開発画面

表 2.4 HALCON のモジュール構成

モジュール名	内容
HDevelop	対話型開発環境および 1200 以上の画像処理・解析用オペレータ
C-Interface	ユーザ独自の関数組込み用インターフェース：C 言語で開発したユーザ関数を Hdevelop のオペレータとして登録するためのツール
Frame Grabber Interface	ユーザ独自の画像入力ボード組込み用インターフェース：HDevelop で任意の画像入力ボードから画像入力を行うための、画像入力ボードオペレータ登録用ツール
Run-time Libraries (Image Processing Library)	C++, C 言語および COM インターフェース (Visual Basic) 用画像処理・解析ライブラリ：HDevelop から自動生成した C++または Visual Basic コードや、ユーザ作成プログラムコードから直接利用可能

第3章 移動物体検出方法

移動物体を検出する方法として様々な方法があるが，本研究では人物検出ではオーソドックスな手法と言える背景差分法，フレーム間差分法に加えて，照明の影響を受けにくく局所的な幾何学変化に頑健な特徴を持つ Histograms of Oriented Gradients 特徴量や，単純な識別器を組み合わせることによって精度の高い識別機を構成する Boosting 法の一つである Real AdaBoost などを用いた．以下にその手法について記述する．

3.1 背景差分法

背景差分法とは、あらかじめ移動物体の存在しない背景画像を取得しておき、入力画像と背景画像との差分処理結果に応じた 2 値化を行なうことによって移動物体領域を抽出する手法である。個々の物体全体をかたまりとして抽出することができるので物体抽出性は高いと言える。しかしながらその反面、物体の中で背景濃度に近い部分が欠落する、照明（日光）の変化に弱い等の問題がある。また、移動物体の存在しない背景画像を取得するのが難しい状況や背景自体が変化する状況においては背景画像を画像処理によって生成する必要がある。この背景画像を精度良く生成できれば背景差分の精度も向上すると考えられる。背景差分法のアルゴリズムを(3.1.1)に、背景差分法による検出の様子を図 3.1 に示す。

$$s(m, n) = |f(m, n) - g(m, n)| \quad (3.1.1)$$

ここで、 m, n は画像座標位置、 $f(m, n)$ は入力画像(図 3.1(b))、 $g(m, n)$ は背景画像(図 3.1(a))、 $s(m, n)$ は背景差分画像(図 3.1(d))である。

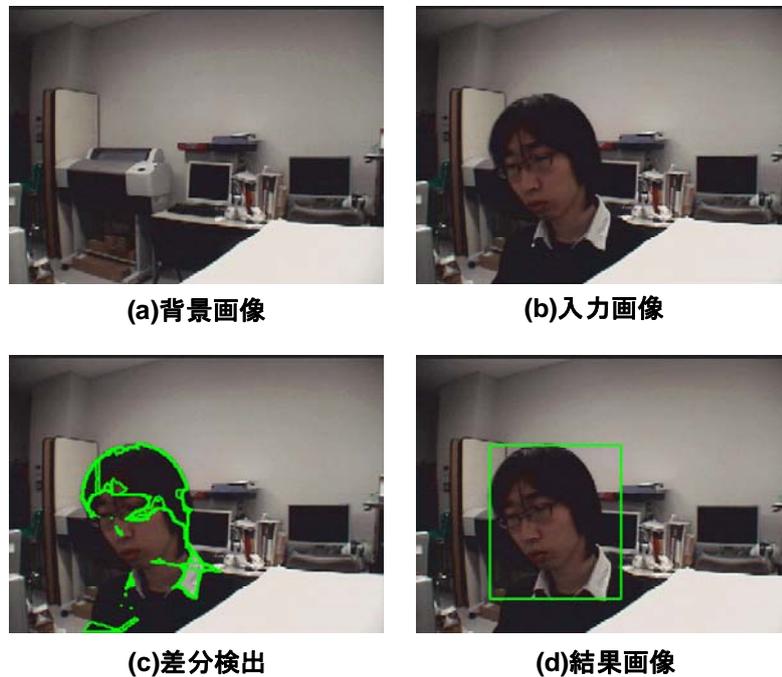


図 3.1 背景差分法での検出

3.2 フレーム間差分法

フレーム間差分法とは、時刻(t)の入力画像と時刻($t-n$)の入力画像との差分を取ることであり、処理すべき対象を得る手法である（一般的に $n=1$ ）。人物のようなエッジ部分の少ない物体が動いた場合、処理結果としてはそのエッジ部分が検出される。フレーム間差分法は対象に動きがない場合、差分が表れないなど不安定な部分があるが、常に背景モデルを更新している形になるので、照明の変化など環境の変化は無視出来る。つまり、環境の変化に強いといえる。フレーム間差分法のアルゴリズムを(3.2.1)に、フレーム間差分法による検出の様子を図 3.2 に示す。

$$s_t(m, n) = |f_t(m, n) - f_{t-1}(m, n)| \quad (3.2.1)$$

ここで、 m, n は画像座標位置、 $f_t(m, n)$ は時刻 t における入力画像(図 3.2(a))、 $f_{t-1}(m, n)$ は時刻 $t-1$ における入力画像(図 3.2(b))、 $s_t(m, n)$ はフレーム間差分画像(図 3.2(d))である。

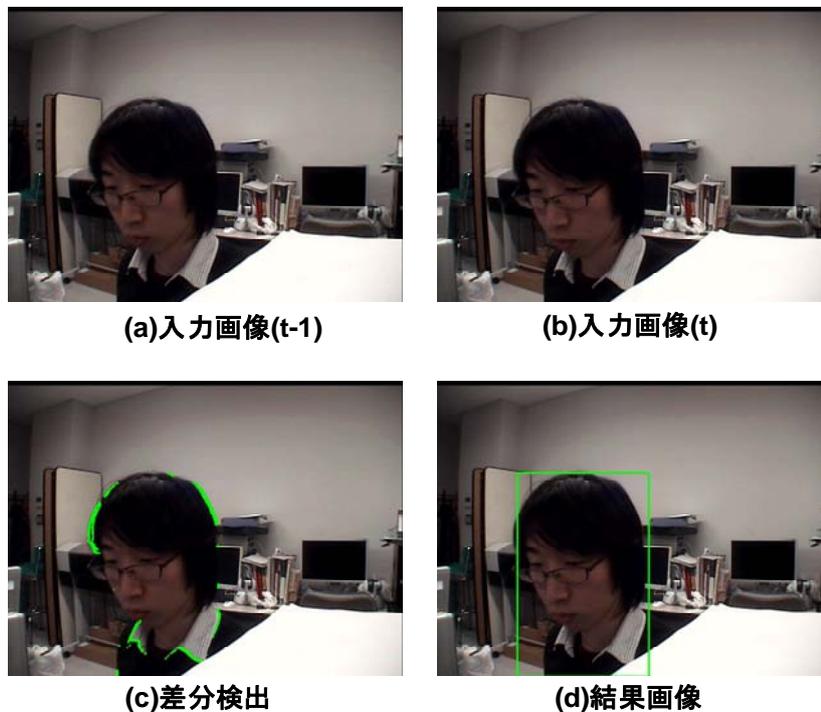


図 3.2 フレーム間差分法での検出

3.3 モフォロジー処理

3.3.1 収縮・膨張処理

図形を 1 画素分だけ細める, あるいは太める処理をそれぞれ図形の収縮, 膨張という. 収縮は侵食とも呼ばれている. また, 膨張は拡散, 拡張, 伝播などとも呼ばれている. 図形の収縮は, 図形の境界点 (近傍に 0-画素を持つ 1-画素) の値を 0 にする操作であり, 図形の膨張は背景の境界点 (近傍に 1-画素を持つ 0-画素) の値を 1 にする操作である.

3.3.2 モフォロジー処理

一般に n 回の収縮と n 回の膨張を組み合わせた処理を図形融合と呼び, 2 値画像の平滑化やノイズ除去, 小さい穴や凹凸の除去などに用いられる.

収縮の次に膨張を行う処理をオープニング処理と言い, 細かいノイズや不要な突起部分を除去することができる.

一方膨張の次に収縮を行う処理をクロージング処理といい, 図形の穴を埋めたり, 切れている箇所をつないだりすることができる.

このように, 収縮と膨張を組み合わせることにより, 2 値画像中の孤立点除去, 2 値画像の平滑化, ノイズの除去などが可能になる.

また, 本来 1 つの図形であったものが, 亀裂状のノイズのために 2 つに分かれた図形として入力された場合などにおいて, それを修正して元の 1 つの図形に戻すために利用できる. 画像のノイズ成分は, 正確な図形の解析を妨げる可能性が高いので, そのための前処理としてモフォロジー処理はよく用いられる.

モフォロジー処理は一般的には 2 値画像に対して行われる処理であるが, 濃淡画像にも応用が可能である.

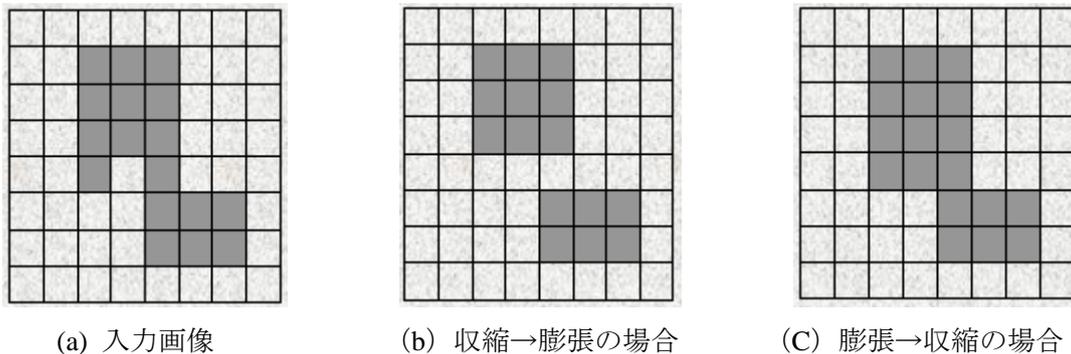


図 3.3 処理順序で結果が異なるモフォロジー処理の一例

3.3.3 本システムに用いたモフォロジー処理

フレーム間差分法によって検出された領域が、切れている箇所や穴がある場合がある（図 3.4）. そのため、穴を埋めたり、切れている箇所をつないだりするクロージング処理を施す. 半径 12.5 の円形構造要素より小さい領域内の穴を閉じるようにする（図 3.5）.

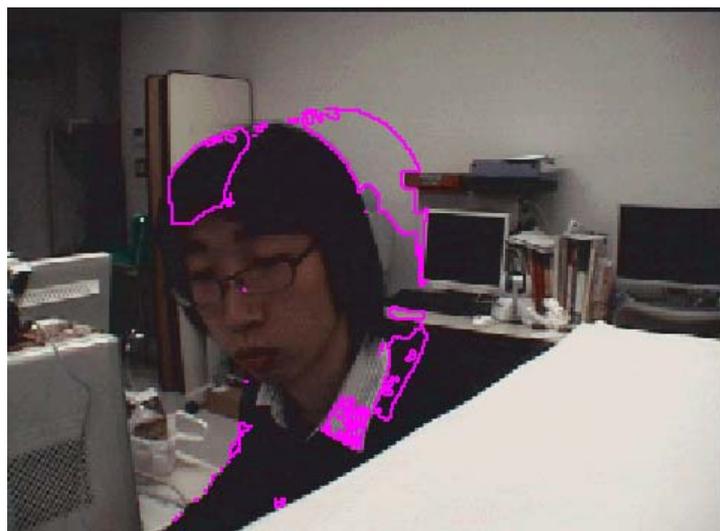


図 3.4 提案手法によって検出された領域

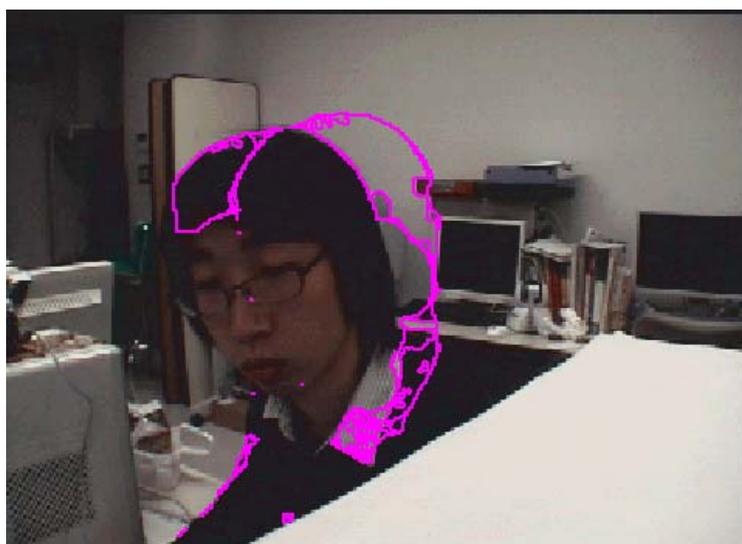


図 3.5 モフォロジー処理を施し，領域を結合した画像

3.4 Histograms of Oriented Gradients

本研究では, Dalal らにより提案されている Histograms of Oriented Gradients 特徴量(以降, HOG と略す)[3]を用いる. HOG 特徴量とは, 局所領域における輝度の勾配強度と勾配方向をヒストグラム化した特徴量であり, 物体の形状を表すことが可能である. 近接画素の勾配を局所領域によってヒストグラム化するため, 照明の影響を受けにくく, 局所的な幾何学変化に頑健という特徴がある. 以下に簡単な HOG の算出方法を示す.

Step1 : 輝度の勾配強度と勾配方向の算出

Step2 : 輝度の勾配方向ヒストグラムの作成

Step3 : ブロックによる正規化

最初に, 全ての画素において, 各ピクセルの輝度から勾配強度と勾配方向を下記の式により算出する.

$$f_u(u, v) = I(u + 1, v) - I(u - 1, v) \quad (1)$$

$$f_v(u, v) = I(u, v + 1) - I(u, v - 1) \quad (2)$$

$$m(u, v) = \sqrt{f_u^2 + f_v^2} \quad (3)$$

$$\theta(u, v) = \arctan \frac{f_v}{f_u} \quad (4)$$

なお, ここで I は入力画像を, (u, v) は画素の位置を表すものとする.

次に, 5×5 画素大の, 互いに重ならないように敷き詰めた「セル」を図 3.6 のように配置する. 算出された算出された勾配方向は, $0 \sim 360^\circ$ となるが, $0 \sim 180^\circ$ 変換し, 勾配方向を 20° ずつに分割することにより 9 方向の勾配方向ヒストグラムを得る. それぞれの各画素の勾配方向に勾配強度を加算していくことで 9 次元の特徴ベクトルを生成する.

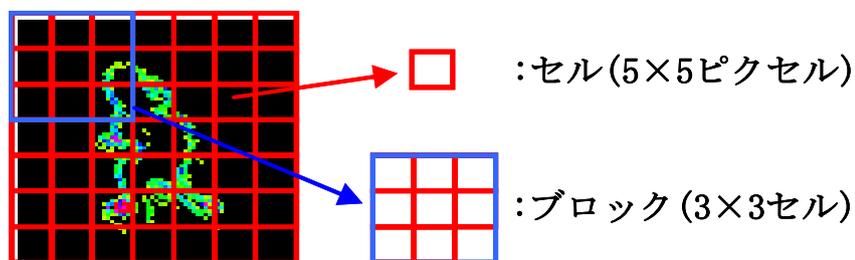


図 3.6 HOG の流れ

さらに3×3の大きさのセルを「ブロック」とし、1セルずつずらしながら多数のブロックの形成を行う。この時、セルの時とは異なり、ブロックは互いに重なっていくように生成する。それぞれのブロック内のヒストグラムを連結して正規化を行い、図 3.7 のように 81 次元のベクトルを抽出する。

本研究では45×25の画像を用いているため、 $7 \times 5 \times 81 = 1701$ 次元となる。

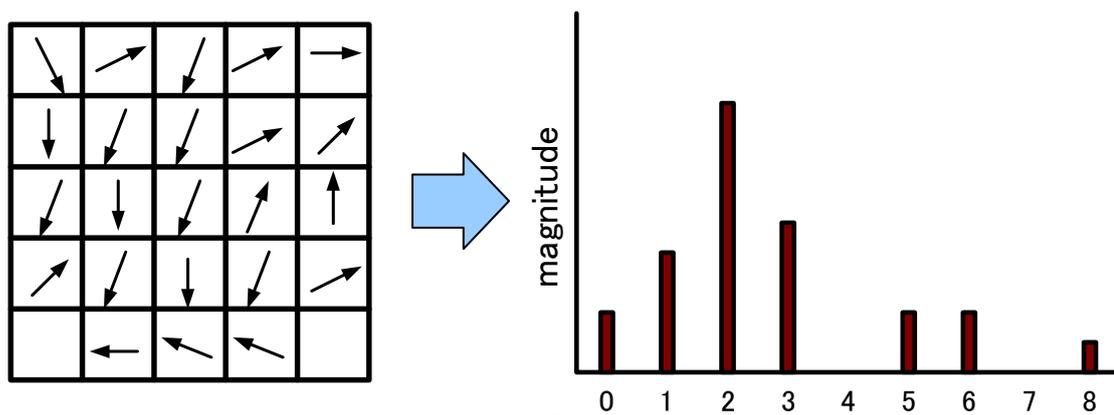


図 3.7 正規化の流れ

3.5 RealAdaBoost

AdaBoost とは、一つ一つはランダムに判定するよりはましというレベルの単純な識別器を組み合わせることによって精度の高い識別機を構成する Boosting 法である(図 3.8). 今回は AdaBoost を発展させた Real AdaBoost という手法を用いた.

今回, 提案する共起特徴は Real AdaBoost の複数の弱識別器から得られる出力を演算子によって結合した特徴量である. 通常の AdaBoost は各弱識別器が 0 か 1 の二値をとり, それらの重み付き和を強識別器としていたが, RealAdaBoost では弱識別器が離散的な連続値をとるため, 複数の特徴量を演算子により多様な捉え方で観測することが可能であり, 単一の特徴のみでは識別困難なパターンに対しても高精度な検出が可能となる.

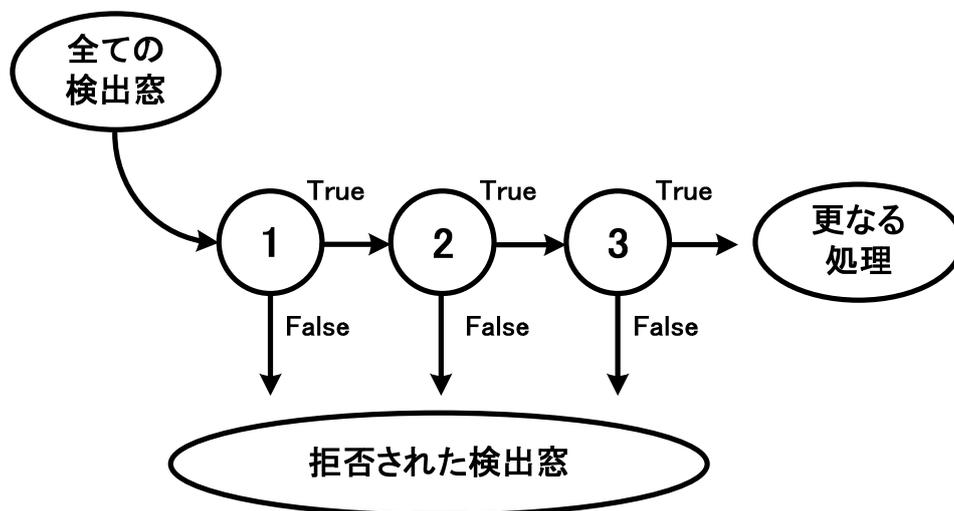


図 3.8 AdaBoost の流れ

共起特徴算出の方法について説明する. まず HOG 特徴量 x を入力とした Real AdaBoost の弱識別器の出力 $h(x)$ を算出するために, ポジティブクラスとネガティブクラスの確率密度分布 w_+, w_- を作成する. 確率密度分布 w_+ は, 1次元のヒストグラムにより表現され, 学習サンプルの重み D_i に基づいて次式より作成する.

$$w_+^j = \sum_{i: j \in J \wedge y_i = +1} D_i(i) \quad (5)$$

$$w_-^j = \sum_{i: j \in J \wedge y_i = -1} D_i(i) \quad (6)$$

学習サンプル i から特徴量を算出し、特徴量の値に対応するヒストグラムの BIN の番号 j へ学習サンプルの重み $D_i(i)$ を加算することで、確率密度分布 w_{\pm} を作成することができる。

次に作成した確率密度分 w_{\pm} を用いて、RealAdaBoost の弱識別器の出力 $h(x)$ を算出する。学習サンプルより得られる HOG 特徴量 x の値から、ヒストグラムの BIN の番号 j を算出し、それに対応した確率密度分布 w_{\pm} から弱識別器の出力 $h(x)$ を下記の式より算出する。

$$h(x) = \frac{1}{2} \ln \frac{W_+^j + \varepsilon}{W_-^j + \varepsilon} \quad (7)$$

ここでの ε とは除算不能を防ぐ係数である。

算出した複数の弱識別器の出力 $h(x)$ を用いて共起特徴 C を算出する。下記の式により N 個の弱識別器から得られる出力を演算子により結合した共起特徴 C を算出する。

$$C = \sum_{n=0}^N h_n(x) \quad (8)$$

共起特徴 C は弱識別器の出力の和であるため、複数の特徴量を総合的に捉える特徴量となる。

第4章 移動物体追跡方法

4.1 提案手法

本研究室では、PTZ カメラを用いた移動物体の自動追跡システムについて研究されてきた。近年では、追跡精度の向上を目的とした1台のPTZ カメラを用いたフレーム間差分及びオプティカルフローによる移動物体追跡[6]や、追跡対象を均等な大きさにする Zoom 機能に着目した追跡実験などを行ってきた。しかし、いずれも追跡範囲内に追跡対象が1人しかいない場合を想定しており、2人以上の対象がPTZ カメラに入った場合、誤報となることが非常に多かった。実際に監視カメラを設置した場合、追跡対象が単体であるとは限らず複数いることの方が普通であり、状況によっては複数の対象に対して追跡を行わなくてはならない。

そこで本研究では、複数の対象を追跡するために2台のカメラを使用する。複数の対象の検知・追跡を行う CCD カメラとその検知した複数の対象から指定した対象をより細かく追跡する PTZ カメラに分けて、複数の対象を追跡する手法を提案する。以下では、CCD カメラを用いる複数追跡手法と PTZ カメラを用いる単体追跡手法について記述する。



図 4.1 2台のカメラ

4.2 複数追跡手法

本項では、複数の対象を検知・追跡を行う CCD カメラについて記述する。

複数の対象を追跡する手法として、第 3 章の 3.4, 3.5 で記述した HOG 特徴量と RealAdaBoost を用いて強識別器を構築する。詳しくは第 5 章の 5.1 学習実験にて記述する。構築した強識別器を元に人物検出を行うが、計算コストの少ない HOG 特徴量ベースの強識別器を用いても画像全体に全探索をかけるのは膨大な計算時間がかかり、実時間での動作は困難である。

そこで本研究では、探索範囲を狭めるため前処理として第 3 章の 3.2 で記述したフレーム間差分法を使用した。以下に簡単な手順について記述する。

Step1 : フレーム間差分法を用い、前フレームとの差分領域の中心座標を取得する。

Step2 : 得られた座標を中心に最大 80×80 の領域を構築する。

(座標が画面の端などで 80×80 の領域を作れない場合、その領域分を狭める。)

Step3 : その領域に対して、HOG 特徴量ベースの強識別器を用いて探索をかける(図 4.2)。

Step4 : 人物領域(人の有無)を判別する。

これにより、探索領域は“差分領域×80×80”となり、処理コストを大幅に削減することができる。

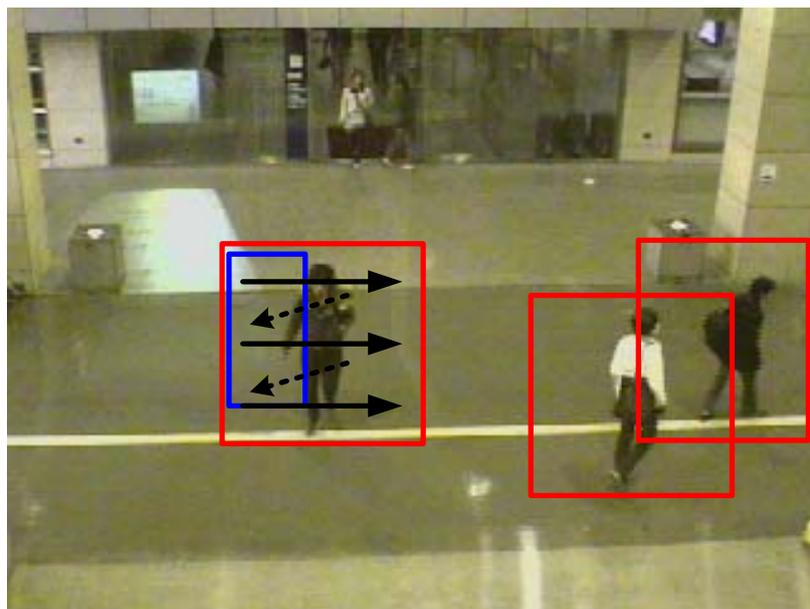


図 4.2 複数追跡手法

4.3 単体追跡手法

本項では、CCD カメラにて検知した複数の対象から指定した対象をより細かく追跡する PTZ カメラについて記述する。

4.3.1 ナンバリング

複数の対象の中から任意の人物を追跡する上で、常に同じ人物を追跡する必要がある。そこで CCD カメラにて取得した画像に対して、ナンバリングを行う。

- Step1 :** 初回時刻(t)に限り、検出した人物領域に対して、最大 3 人までナンバリングを行う(図 4.3)。この際、得られたそれぞれのナンバーごとの人物領域の中心座標を取得する。
- Step2 :** 時刻(t+1)にて、4.2 複数追跡手法の手順で人物領域の判別する。
- Step3 :** 時刻(t+1)の人物領域の中心座標と時刻(t)の人物領域の中心座標の距離をそれぞれ算出する(図 4.4)。
- Step4 :** それぞれのナンバーからの距離に対して、最も距離が短いものに同じナンバーを割り振る。



図 4.3 ナンバリングの様子



図 4.4 最短距離の算出

以上を繰り返すことにより，ナンバリングを更新する．

この手法はすれ違いで重なった場合に誤報となることがあるが，本研究では，建物の 2 階から斜め下を撮影しているため，最速で 3fps で撮影することができる上，人物領域が完全に重なることはほぼないためこの手法を利用する．

4.3.2 カメラ制御

本研究の目的の部分にあたる単体追跡について記述する.

カメラの制御は第2章の2.2PTZカメラとCCDカメラで記述した絶対角度を使用する. 前述で行ったナンバリングに対して, 追跡したい対象の番号を指定する. 指定された対象の番号の座標を用いて, PTZカメラを動かす距離・方向を求める. あらかじめ 320×240 の画像に対して図4.5のように絶対角度を指定しておく. 指定された対象の番号の座標の差の x , y 成分の値により, PTZカメラを動かす方向を求める.



図 4.5 カメラの制御

第5章 実験

本研究では，研究の有用性を示すため，3つの実験を行った．

- ・人物検出を行う上で，検出対象である“人”を追跡するために人物の強識別器を構築する『学習実験』．
- ・人物検出を行う上で，検出手法で用いる HOG 特徴量の有用性を示す『検出実験』．
- ・そして，本研究の目的である複数人物を対象とした単体追跡システムの性能を測る『追跡実験』．

以下に3つの実験について詳しく記述する．

5.1 学習実験

この実験では、人物検出を行う上で検出対象である“人”を追跡するために3章で記述した HOG 特徴量と Real AdaBoost を用いて、人物の強識別器を構築する。

共起特徴 C を算出し、強識別器を構築するため、学習のための学習用サンプル画像と評価するための評価用画像を収集し、判別を行った。今回、検出対象である人の画像をポジティブ画像(図 5.1)、それ以外の画像をネガティブ画像(図 5.2)とする。

学習用サンプル画像と評価用画像は今回実験を行う場所である法政大学小金井キャンパス西館 2 階から撮影した映像から 45×25 の画像を切り出すことにより作成した。ネガティブ画像に関しては実験環境の画像の他に、全く関係のない画像(デスクトップの壁紙など)から同様に 45×25 の画像を切り出して作成した。

ポジティブ学習用画像サンプルは 2100 枚、ネガティブ学習用サンプル画像は 4000 枚用意した。同様に、ポジティブ評価用画像は 1000 枚、ネガティブ評価用画像は 1000 枚用意した。これらの学習用サンプル画像から強識別器を構築し、評価用サンプルに対して判別を行う。実験結果では、RealAdaBoost により算出された人物の未検出率(Miss Rate)と誤検出率(False Positive Rate)のしきい値ごとの出現率の関係をグラフに出した。

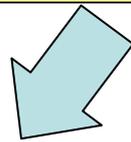


図 5.1 ポジティブ画像



図 5.2 ネガティブ画像

5.1.1 実験結果

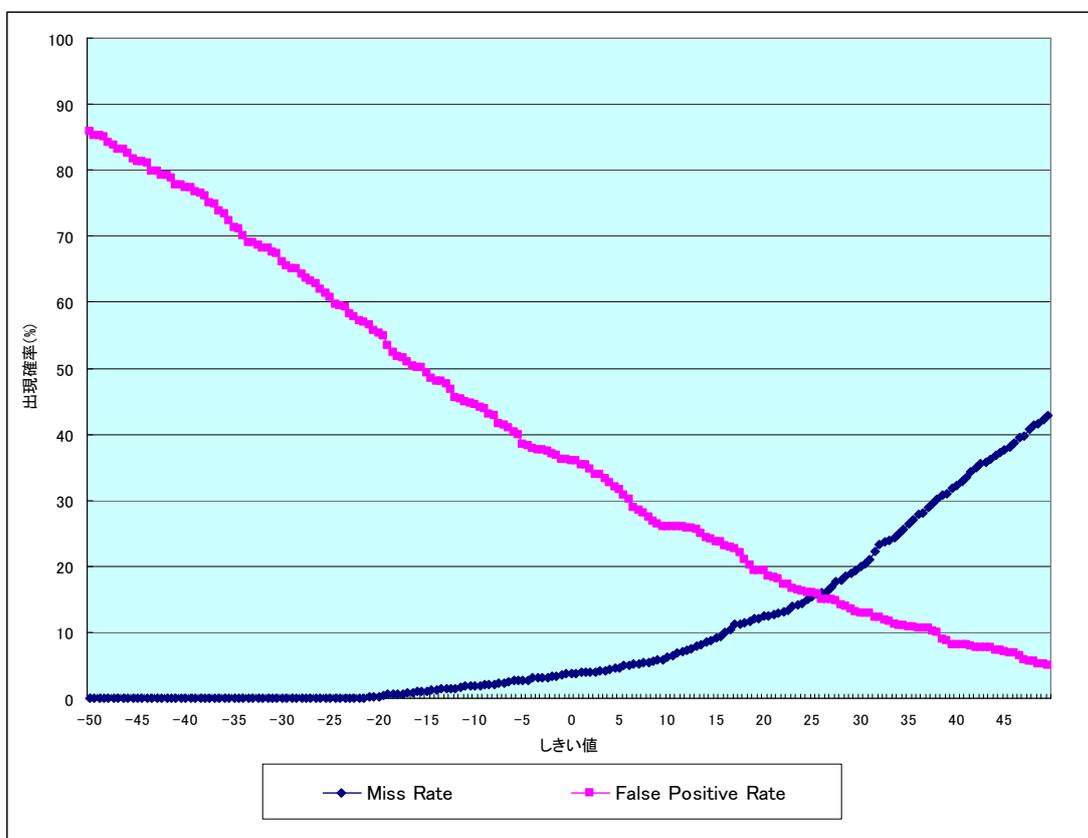


図 5.3 未検出率と誤検出率の関係

5.1.2 考察

実験の結果，RealAdaBoost によって算出したしきい値 25.5 の地点で人物の未検出率と誤検出率の最小出現率が 15%となった．これは少し高いように思えるが，以後の実験において目立つほどの誤報は見られなかった．理由として背景と一部同化しているポジティブ画像(図 5.4)や輝度勾配の大きいネガティブ画像の一部が誤報となったと考えられる．また良質な画像のみを選別した場合，最小出現率を 10%程度まで下げることができたが，逆に人型の判別の値が高くなりすぎて，車などにより身体の一部が少し隠れただけで人物だと判別してくれない場合が多くあった．本研究ではポジティブ画像を 2100 枚しか用意しなかったが，もっと様々なケースの画像を収集して学習にかけることで最小出現数を下げることが出来たのではないかと考えられる．

よって本研究では，以後の検出実験・追跡実験では，しきい値 25.5 の地点を人物の強識別器として利用する．

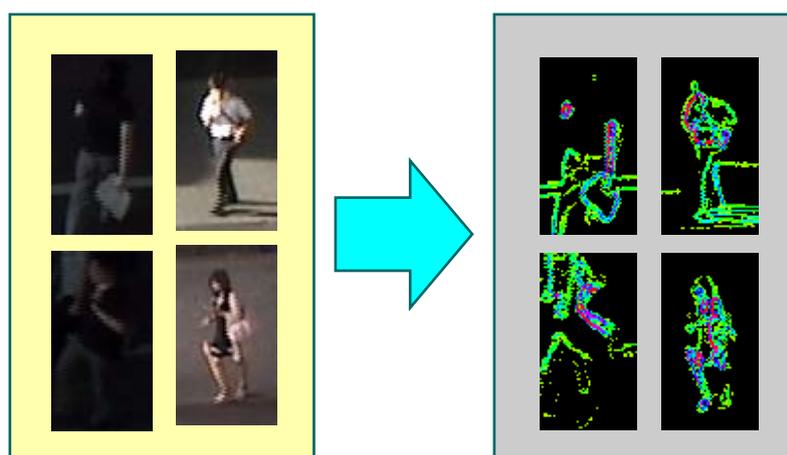


図 5.4 誤報となった人物画像の一部とその輝度勾配の視覚化

5.2 検出実験

この実験では、CCD カメラ 1 台を用いて先ほどの学習実験で構築した強識別器を元に人物の検出を行う。

日中と夜間での環境下にて実験を行い、照明や影、局所的な幾何学的変化に頑健な HOG 特徴量の有用性を示す。

『検出手法』として、

- ・フレーム間差分のみ
- ・フレーム間差分+HOG

『実験環境』(図 5.5)として、屋根と照明がある環境下で、

- ・日中
- ・夜間

上記の条件で実験を行う。実験結果では、それぞれの人物の検出数をフレーム数ごとに計測し、グラフにした。



図 5.5 実験環境

5.2.1 実験結果

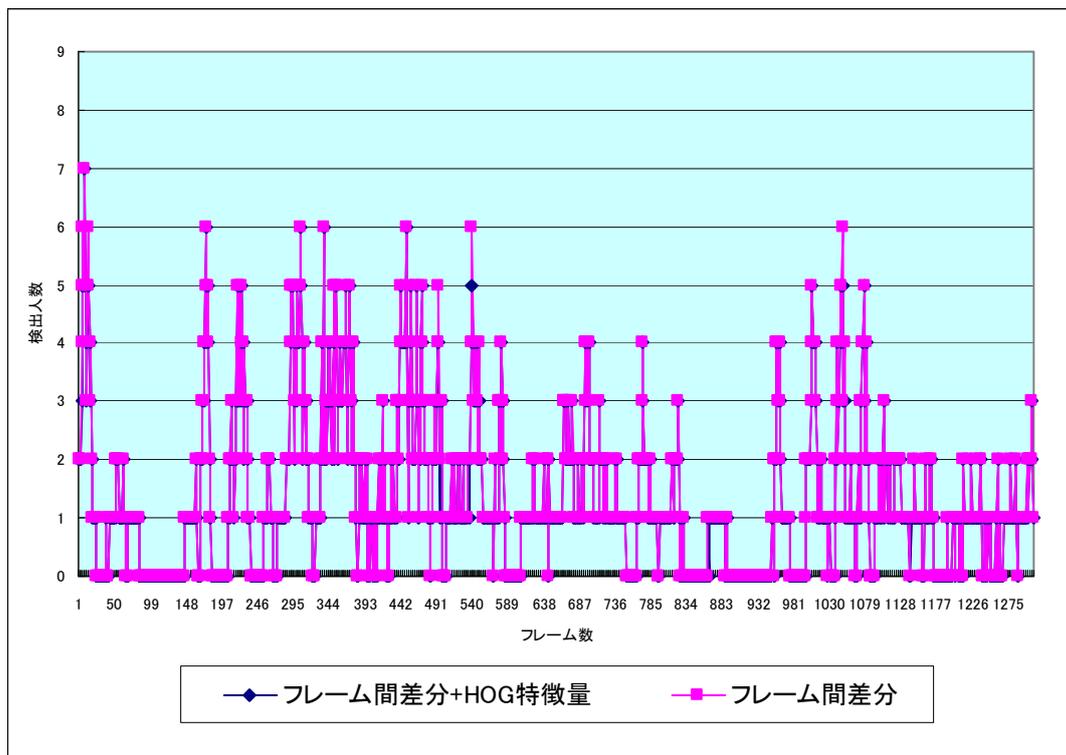


図 5.6 日中における撮影間の人物検出の経緯

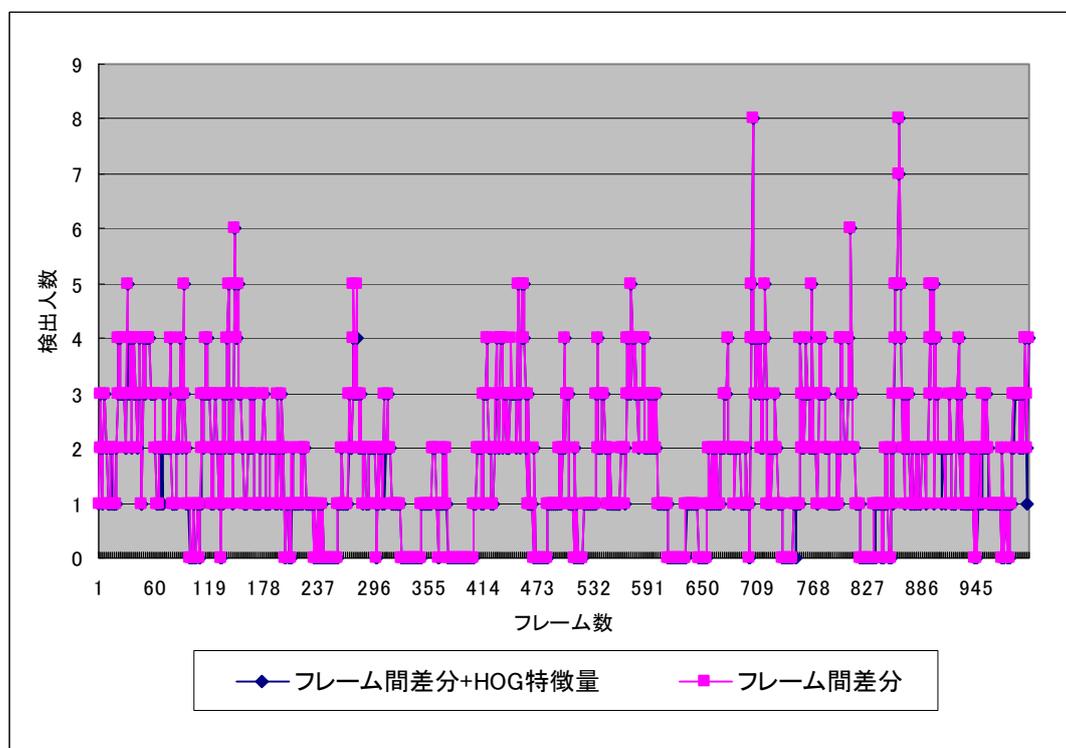


図 5.7 夜間における撮影間の人物検出の経緯

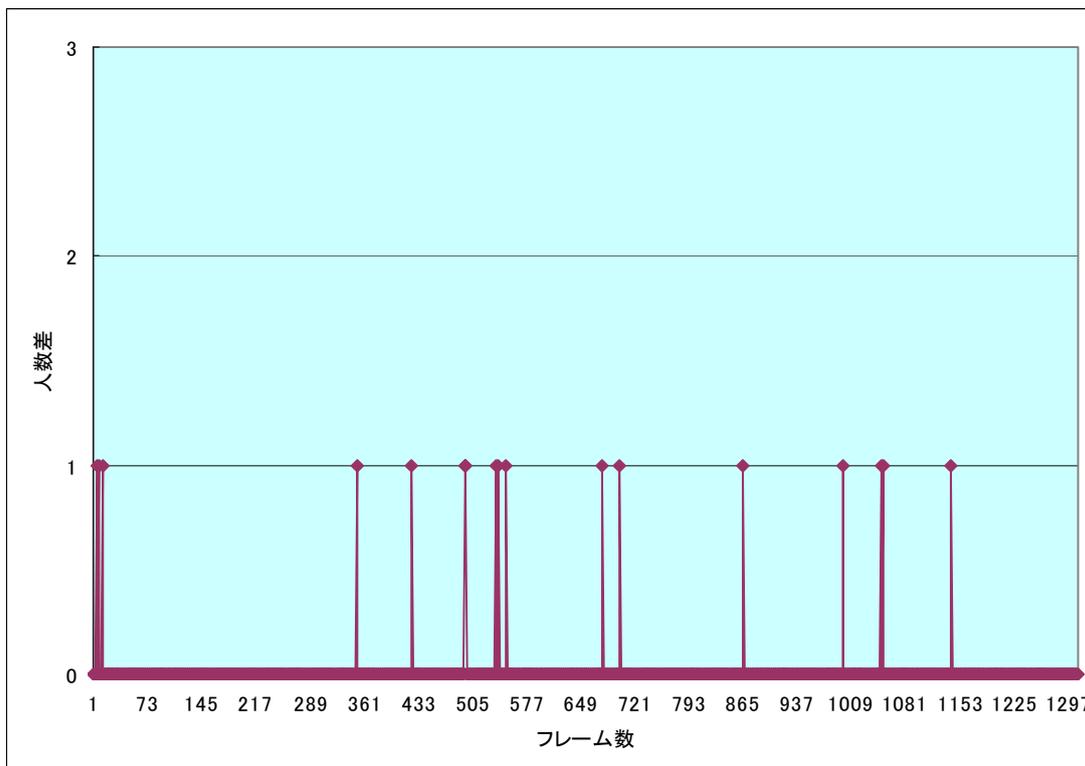


図 5.8 日中における撮影間の人物検出の差異

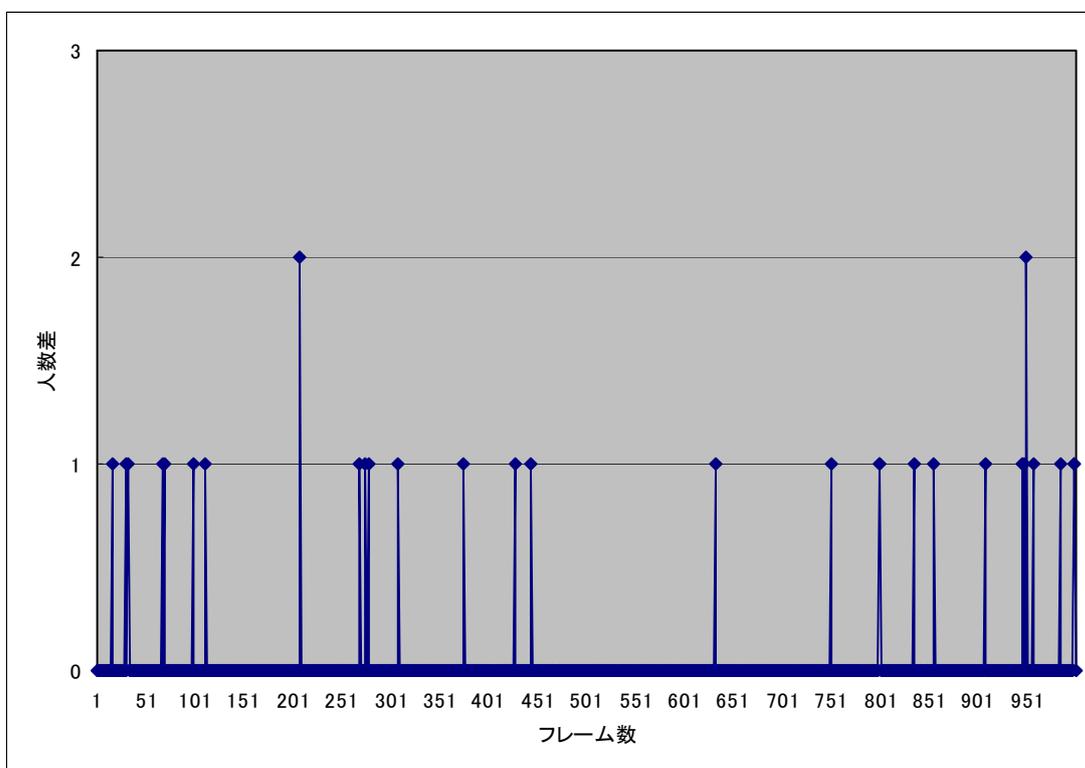


図 5.9 夜間における撮影間の人物検出の差異

5.2.2 考察

今回、日中と夜間でほぼ同条件(約 5 分程度、人物領域 1700 カウント程度)での人物検出の経緯をグラフにした(図 5.6, 図 5.7). 二つの図から見ても、“フレーム間差分のみ”と“フレーム間差分+HOG 特徴量”の差は分かりにくいので、その差を検出したものをグラフにした(図 5.8, 図 5.9). これはフレーム間差分のみでの人物以外に人物領域だと判別してしまったグラフでもある. この主な原因として、車の移動や風によるゴミなどの介入、照明によるノイズなどである. また日中に比べ、夜間の方に認識ミスが多いのは、フレーム間差分のみの場合、照明などによるノイズによる影響が主だと考えられる. また実験中、目視にてフレーム間差分+HOG 特徴量の人物領域判別を観測していたが、目立つ認識ミスは見られなかった. いずれも日中・夜間あわせて 60 回ものミスが出ており、HOG 特徴量ベースの強識別器の有用性が分かる.

5.3 追跡実験

この実験では、2台のカメラを用いて本研究の目的である複数人物を対象とした単体人物の追跡を行う。

CCDカメラの画面範囲内に人物が2人以上存在し、指定した対象が画面外に消えるまで指定対象のみを追跡できたかを判別するという評価実験を行う。

成功判定として、中心付近にいる任意の人物を指定してから CCD カメラに人物が2人以上の人物が3秒以上存在している状態で、その指定した対象が CCD カメラの画面外に消えるまで、PTZカメラで指定した対象を常に追跡できていた場合を成功とする。

失敗判定は、CCDカメラの画面外に消える前に、PTZカメラで指定した対象を見失った場合を失敗とする。

『追跡手法』として、

- ・フレーム間差分+HOG

『実験環境』として、屋根と照明がある環境下で、

- ・日中
- ・夜間

上記の条件で、日中に500回、夜間に500回、計1000回の実験を行い、撮影間の人数ごとの成功率と失敗率を計測した。

5.3.1 実験結果

表 5.1 複数人物を対象とした単体追跡の結果

	回数	割合(%)	総合回数
2人成功	478	85.82	557
2人失敗	79	14.18	
3人成功	231	71.30	324
3人失敗	93	28.70	
4人成功	28	38.89	72
4人失敗	44	61.11	
5人成功	6	19.35	31
5人失敗	25	80.65	
6人成功	1	9.09	11
6人失敗	10	90.91	
7人成功	0	0.00	5
7人失敗	5	100.00	
総合成功	744	74.40	1000
総合失敗	256	25.60	
総合計	1000		

表 5.2 日中と夜間別の追跡結果

	昼合計	昼割合(%)	夜合計	夜割合(%)
2人成功	213	88.38	265	83.86
2人失敗	28	11.62	51	16.14
3人成功	137	75.27	94	66.20
3人失敗	45	24.73	48	33.80
4人成功	22	40.74	6	33.33
4人失敗	32	59.26	12	66.67
5人成功	3	17.65	3	21.43
5人失敗	14	82.35	11	78.57
6人成功	1	25.00	0	0.00
6人失敗	3	75.00	7	100.00
7人成功	0	0.00	0	0.00
7人失敗	2	100.00	3	100.00
総合計	500		500	

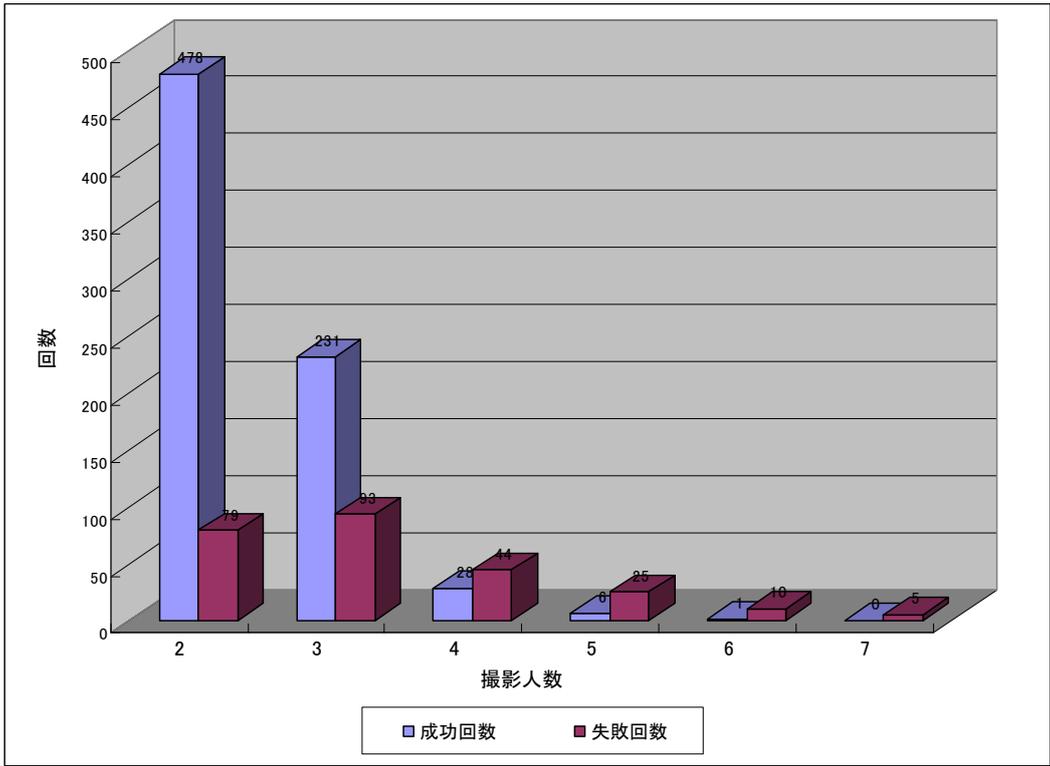


図 5.10 複数人物を対象とした単体追跡の成功数

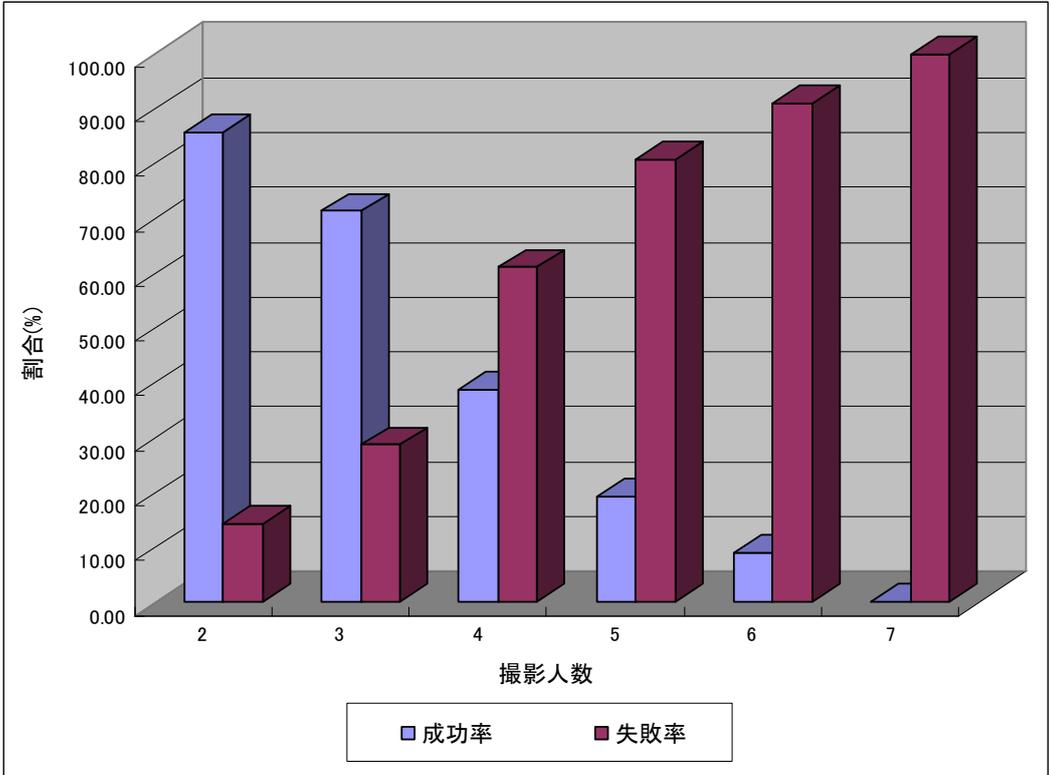


図 5.11 複数人物を対象とした単体追跡の成功確率

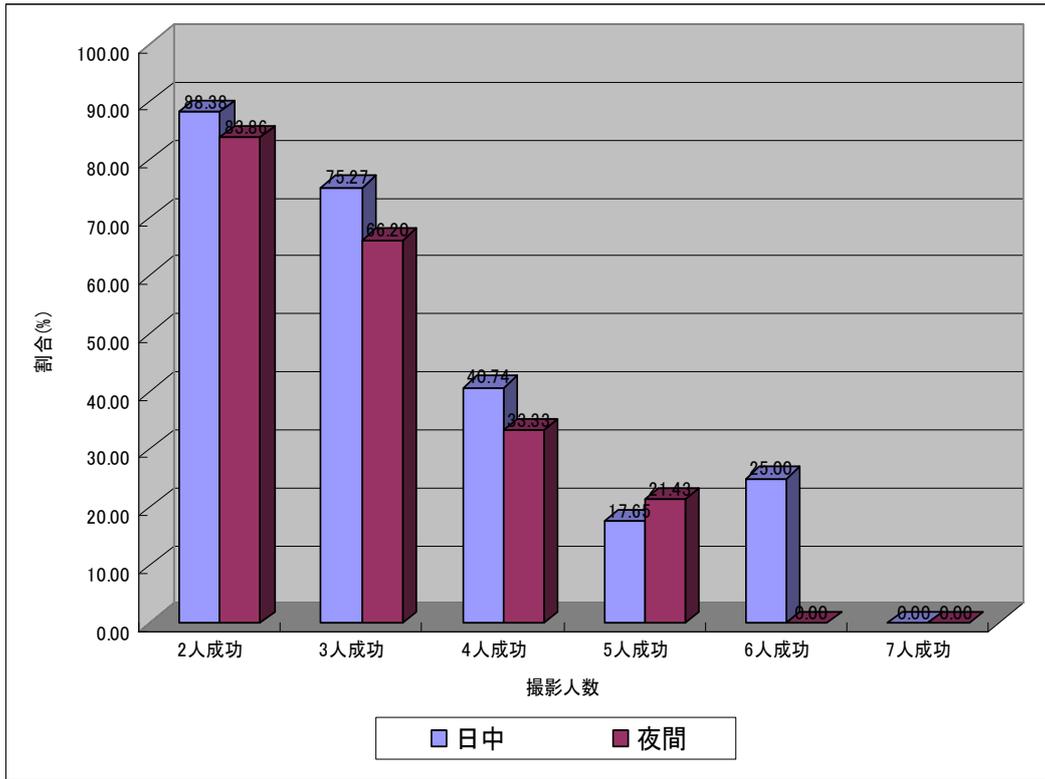


図 5.12 日中と夜間別の追跡成功率

5.3.2 考察

今回行った実験では、2,3 人の時と比べて、4 人以上から失敗率が成功率を上回る結果となった(図 5.10). 失敗した主な理由として、人数が増えていくほど HOG 特徴量を算出する 80×80 の探索領域が増えていき、処理速度が相対的に減少していくためだと考えられる。これは 2~3 人の時は 3fps 程度で動作していたのに対して、4 人以上になると 1,2fps 程度までに落ちるほどである。加えて、ナンバリングをする際、1 フレーム前にとった番号の座標との距離を計測しているため、fps が落ちるほど人物の移動距離が長くなり、精度が落ちると考えられる。成功率を上げるためには、人物の判別率は落ちるが HOG 特徴量の勾配強度を数値に閾値処理を行い 0 と 1 にするバイナリパターン化[7]などで次元数を減らし、全体の処理速度を上げる必要がある。

また実時間での動作を実現する上で前処理としてフレーム間差分を用い領域を限定しているため、6 人以上においては狭い画面内で正確に追跡することは難しかった。CCD カメラのレンズを広角レンズにして撮影領域を広げていけば、重なる領域も減り、成功率は増加していたと考えられる。

日中と夜間において実験を行ったが、あまり差は出ず、日中の方が全て 5%程度高くなる結果となった(図 5.12). これは屋根があり、電灯がある環境で実験を行ったこともあるが、夜間において日中とあまり変わらない結果がでたのは、5.2 検出実験でも述べたが照明や影の影響に頑健な HOG 特徴量を使用したためだと考えられる。5%の差についてはノイズなどの影響で前処理によるフレーム間差分での探索する領域が増え、処理速度が低下したためだと考えられる。

第6章 むすび

複数人物を対象とした単体追跡の実験では全体の成功率として 74.4%，試行人数の多かった 2,3 人の成功率で見ると 80.5%という結果となった。一部，人が乱雑した環境下を除き，実時間で安定したシステムを構築することができた。

今後のさらなる精度の向上として，人物検知精度の向上や処理速度の高速化が必要だと考えられる。一例を挙げると，5.1.2 で述べたさらなる多様な学習サンプルの収集による未検出率と誤検出率の最小化や 5.3.2 で述べた勾配強度の閾値処理を行うバイナリパターン化による処理速度の高速化などである。

また前処理にフレーム間差分を使用したのが，すれ違いにも対応できていたとはいえ，完全なものとは言えなかった。人物の歩行の軌跡も要素に入れてナンバリングを行えば，より精度の向上につながるのではないかと考えられる。

本研究により 2 台のカメラを使用することで従来手法ではできなかった複数人物を対象とした単体追跡を可能にすることができた。

謝辞

本研究を進めるにあたり，ご指導ならびに貴重なご意見を頂いた，法政大学理工学部応用情報工学科 赤松茂教授に心より感謝致します。

また，本研究の実験に協力していただいた，法政大学修士2年生の 中村亮太氏，菅原甫史氏，猪俣拓利氏，ならびに赤松研究室の皆様へ深く感謝申し上げます。

参考文献

- [1] 藤吉弘宣:”次世代デジタルカメラ/デジタルムービーを予測する:ねらった被写体を逃さない:物体追跡技術”、映像情報メディア学会誌 Vol.62,No.6
- [2] 山内悠嗣,吉弘亘,山下隆義 : 複数の特徴量間の関連性に着目した Boosting による物体検出, 映像情報メディア学会技術報告,32(35) ,pp85-88, 2008-08-01
- [3] N.Dalal and B.Triggs, “Histograms of Oriented gradients for human detection,”Proc.IEEE Conference on Computer Vision(ICCV) ,pp.1150-1157,1999
- [4] 都築勇司,藤吉弘亘,金出武雄 : SIFT 特徴量に基づく Mean-Shift 探索による特徴点追跡,Vol.49,No. SIG6(SVIM20) ,Page.35-45,2008-03-15
- [5] 林裕司, 藤吉弘亘, 岩堀祐之 : 照明変化を考慮した Mean-Shift によるカラートラッキング, 電子情報通信学会論文誌 Vol.J90-D No.8 ,pp.2039-2048
- [6] Kimisyugu Kon and Shigeru Akamatsu, ”An Automatic Tracking System Using A Single PTZ Camera,”Proc.of IWAIT2008(CD-ROM),jan2008
- [7] 松島千佳,山内悠嗣,山下隆義, 藤吉弘宣 : 人検出のための RealAdaBoost に基づく HOG 特徴量の効果的な削減法, 情報処理学会研究報告, 2009-CVIM-167(32), 1-8, 2009-06-02